



T. C.

SELÇUK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

SINIFLANDIRMA PROBLEMLERİNDE KURAL ÇIKARIMI İÇİN YENİ BİR YÖNTEM GELİŞTİRİLMESİ VE UYGULAMALARI

Murat KÖKLÜ

DOKTORA TEZİ

Bilgisayar Mühendisliği Anabilim Dalını

Eylül-2014
KONYA
Her Hakkı Saklıdır

ÖZET

DOKTORA TEZİ

SINIFLANDIRMA PROBLEMLERİNDE KURAL ÇIKARIMI İÇİN YENİ BİR YÖNTEM GELİŞTİRİLMESİ VE UYGULAMALARI

Murat KÖKLÜ

Selçuk Üniversitesi Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Novruz ALLAHVERDİ
2. Danışman: Yrd. Doç. Dr. Humar KAHRAMANLI
2014, 121 Sayfa
Jüri
Prof. Dr. Novruz ALLAHVERDİ
Prof. Dr. Şirzat KAHRAMANLI
Prof. Dr. Ahmet ARSLAN
Doç. Dr. Harun UĞUZ
Yrd. Doç. Dr. Halife KODAZ

Bilgisayar teknolojileri ve veri tabanı yazılımlarındaki gelişmeler sonucunda büyük miktarda veri birikmiş ve eldeki verilerden anlamlı bilgi çıkarma ihtiyacı ortaya çıkmıştır. Büyük miktarda depolanan bu veriler birçok gizli örüntü içermesine rağmen, toplanan veri miktarı büyüdükçe ve verilerdeki karmaşıklık arttıkça, geleneksel yöntemler ile veri bilgiye dönüştürülemez hale gelmektedir. Bu nedenle günümüzde büyük miktarda verileri çözümlemek amacıyla veri madenciliği yöntemleri yaygın bir şekilde kullanılmaktadır. Veri madenciliği yöntemlerinden sınıflandırma, kümeleme ve birliktelik kuralı keşfetme sıklıkla tercih edilmektedir.

Sınıflandırma, önceden kategorisi belli olan veriler kullanılarak bir model oluşturulup, yeni karşılaşılan verinin hangi sınıfa ait olduğunu belirleme işlemidir. Yeni bir verinin, belirli sınıflar içinde hangi sınıfa ait olduğunu tespit edecek bir sınıflayıcı oluşturmak amacıyla veri madenciliği yöntemleri sıklıkla kullanılmaktadır. Bu sınıflandırma yöntemleri genellikle başarılı olmasına rağmen, ortaya konulan değişik sınıflandırma ve kural çıkarma algoritmaları özellikle çok sınıflı gerçek dünya problemleri için henüz arzu edilen seviyeye ulaşamamıştır.

Bu tez çalışmasında çok sınıflı verilerden kural çıkarımı için yeni bir yöntem geliştirilmiştir. Geliştirilen bu yöntemde ayrık ve gerçel öznitelikler farklı şekilde kodlanmıştır. Ayrık öznitelikler ikili olarak, gerçek öznitelikler ise, iki gerçel değer kullanılarak kodlanmıştır. Gerçel değerler kuralları oluşturan özniteliklerin değer aralıklarının orta noktası ve genişlemesini ifade etmektedir. Kural çıkarım işlemi için sınıflandırma başarısı uygunluk fonksiyonu olarak kullanılmıştır. Uygunluk fonksiyonunun optimizasyonu amacıyla Yapay Bağışıklık Sistemi (YBS) yöntemlerinden olan CLONALG algoritması kullanılmıştır. Önerilen yöntem en uygun aralıkları keşfettiğinden dolayı bu yönteme isim olarak "Aralık Keşfi" anlamına gelen INDISC (INterval DISCovery) verilmiştir. INDISC yöntemi 8 farklı veri kümesi üzerinde test edilmiştir.

INDISC yöntemini uygulamak için Pima yerlileri diyabet hastalığı, Orijinal wisconsin göğüs kanseri, Teşhis wisconsin göğüs kanseri, Deniz kabuğu, Süsen çiçeği, Cam kimliklendirme, Şarap ve Tiroid hastalığı veri kümeleri kullanılmıştır. Veriler Irvine California Üniversitesi (UCI) makine öğrenmesi veri deposundan temin edilmiştir. INDISC yöntemi, Pima yerlileri diyabet hastalığı %80.34, Wisconsin göğüs kanseri (orijinal) %99.12, Wisconsin göğüs kanseri (teşhis) %96.31, Deniz kabuğu %62.59, Süsen çiçeği %100, Cam kimliklendirme %77.10, Şarap %99.44 ve Tiroid hastalığı %93.95 doğrulukla sınıflandırmıştır. Diğer yöntemlerle elde edilen başarı yüzdeleri ile geliştirdiğimiz INDISC yöntemi başarı yüzdeleri karşılaştırılmıştır. Tüm veri kümelerinde INDISC yöntemi ile elde edilen sonuçların diğer yöntemlerle elde edilen sonuçlardan daha başarılı olduğu görülmüştür.

Anahtar Kelimeler: Gerçel değer kodlaması, INDISC yöntemi, Kural çıkarma, Sınıflandırma, Yapay Bağışıklık, Yapay Zekâ Teknikleri.

ABSTRACT

Ph.D THESIS

DEVELOPMENT AND APPLICATIONS OF A NEW METHOD FOR RULE EXTRACTION IN CLASSIFICATION PROBLEMS

Murat KÖKLÜ

**THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE
OF SELÇUK UNIVERSITY
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN COMPUTER ENGINEERING**

**Advisor: Prof. Dr. Novruz ALLAHVERDI
2. Advisor: Assist. Prof. Dr. Humar KAHRAMANLI**

2014, 121 Pages
Jury
**Prof. Dr. Novruz ALLAHVERDI
Prof. Dr. Sirzat KAHRAMANLI
Prof. Dr. Ahmet ARSLAN
Assoc. Prof. Dr. Harun UĞUZ
Assist. Prof. Dr. Halife KODAZ**

It has been aroused the necessity of extracting meaningful information from huge amount of available data that is accumulated as result of development in computer technology and database software. Traditional methods can't cope with turning the data to the knowledge due to amount and complexity of accumulated data that has so many hidden patterns in it. Thus, nowadays the data mining techniques are commonly used for analyzing huge amount of information. Classification, clustering and associated rule extraction of data mining techniques are preferred widely.

Classification is the operation of determining class of the data by forming a model that makes use of data whose categories are previously determined. Data mining techniques are frequently used to form a classifier that determines belonging class of a new data among the predetermined classes. Although these classification methods including different classification and rule extraction algorithms are generally successful they don't reach the required success levels when it comes to multi-class real world problems.

In this dissertation thesis a new method for rule extraction was developed. Real and discrete attributes were coded differently. Discrete and real attributes were coded as binary and two real values, respectively. Real values represent middle points and extensions of value intervals of attributes that forms of rules. Classification success was used as fitness function for rule extraction operation. One of the methods of Artificial Immune System (AIS) called CLONALG algorithm was used for optimization fitness function. Since the proposed method invents the most appropriate intervals it is called as INDISC (Interval DISCovery) meaning "Interval Invention". INDISC method was tested on 8 different data sets.

In order to apply INDISC method the data sets of Pima Indian diabetic illness, Original wisconsin breast cancer, Diagnosis wisconsin breast cancer, Abolone, Iris, Glass Identification, Wine and Newthyroid were used. The data were obtained from Irvine California University (UCI) machine learning data bank. INDISC method has classified Pima indian diabetic illness, Original Wisconsin breast cancer, Diagnosis Wisconsin breast cancer, Abolone, Iris, Glass Identification, Wine and Newthyroid in the success ratios of %80.34, %99.12, %96.31, %62.59, %100, %77.10, %99.44 and %93.95, respectively. Other methods were compared with proposed INDISC method according to success rates of classification. It has been seen that the results obtained from proposed INDISC method are more successful than all other methods.

Keywords: Artificial Immune System, Artificial Intelligence Methods, Classification, INDISC method, Real value codding, Rule Extraction.

ÖNSÖZ

Tez çalışmamda bana yol gösteren, sürekli teşvik eden, her konuda desteğini esirgemeyen ve her türlü bilimsel katkıyı sağlayan değerli tez danışmanım Sayın Prof. Dr. Novruz ALLAHVERDİ'ye, ikinci tez danışmanım Sayın Yrd. Doç. Dr. Humar KAHRAMANLI'ya teşekkür eder şükranlarımı sunarım. Tezim süresince bana her türlü desteği sağlayan Tez İzleme Komitesi üyeleri Sayın Prof. Dr. Ahmet ARSLAN ve Prof. Dr. Şirzat KAHRAMANLI'ya teşekkürü bir borç bilirim. Çalışmam esnasında büyük destekleri olan hocalarım Doç. Dr. İsmail SARITAŞ, Yrd. Doç. Dr. Kemal TÜTÜNCÜ, Öğr. Gör. Tahir SAĞ ve Öğr. Gör.Mustafa BÜBER'e teşekkür ederim. Ayrıca, bu tez çalışması süresince bana destek olan eşim Niğmet KÖKLÜ'ye ve oğlum Kemal KÖKLÜ'ye de minnettarlığımı ve sevgilerimi sunarım.

Murat KÖKLÜ
Konya-2014

İÇİNDEKİLER

ÖZET ..... **iv**

ABSTRACT ..... **v**

ÖNSÖZ ..... **vi**

İÇİNDEKİLER ..... **vii**

SİMGELER VE KISALTMALAR ..... **xi**

1. GİRİŞ **1**

1.1. Çalışmanın Amacı ve Önemi 2
1.2. Literatür Araştırması 3
1.3. Tezin Organizasyonu 14

2. SINIFLANDIRMA **15**

2.1. Sınıflandırma Algoritmaları 15
2.1.1. Karar ağaçları 15
2.1.2. Bayes sınıflandırma 20
2.1.3. Geri Yayılım ile Sınıflandırma 21
2.1.4. Diğer Sınıflandırma Metotları 21
2.2. Bölüm Değerlendirmesi 22

3. KURAL ÇIKARIMI **23**

3.1. Kural Çıkarımının Önemi 23
3.2. Kural Çıkarımının Avantajları 23
3.3. Kuralların Gösterimi 24
3.4. Çıkarılan Kuralların Özellikleri 24
3.5. Kural Çıkarımındaki Önemli Noktalar 25
3.6. Taksonomi 25
3.7. Bölüm Değerlendirmesi 27

4. BAĞIŞIKLIK VE YAPAY BAĞIŞIKLIK SİSTEMLERİ **29**

4.1. Bağışıklık Sistemi 29
4.1.1. İnsan bağışıklık sisteminin yapısı 31
4.1.2. Bağışıklık hücreleri 33
4.1.3. Bağışıklık sisteminin vücudu koruma süreci 36
4.2. Yapay Bağışıklık Sistemleri 39
4.2.1. Şekil uzayı 40
4.2.2. Duyarlılık hesaplaması 42
4.2.3. Klonal seçme algoritması (CLONALG) 43
4.3. Bölüm Değerlendirmesi 49

5. INDISC ALGORİTMASI 50

5.1. INDISC Algoritma Yapısı 50
5.1.1. Veri kodlanması 50
5.1.2. Kullanılan duyarlılık fonksiyonu 52
5.1.3. Kural Çözme 53
5.1.4. Kuralların Budanması 53
5.2. INDISC Algoritmasının Program Yazılımı 54
5.2.1. INDISC yazılımının arayüzü 54
5.2.2 INDISC yazılımıyla kural çıkarımı 58
5.3. INDISC Yazılımı İçin Örnek Uygulama 60
5.4. Bölüm Değerlendirmesi 63

6. DENEYSEL SONUÇLAR VE ANALİZ 64

6.1. İki Sınıflı Veri Setleri Üzerindeki Uygulamalar 64
6.1.1. Pima yerlileri diyabet hastalığı veri kümesi 64
6.1.2. Wisconsin göğüs kanseri (orijinal) veri kümesi 67
6.1.3. Wisconsin göğüs kanseri (teşhis) veri kümesi 69
6.2. Çok Sınıflı Veri Setleri Üzerindeki Uygulamalar 73
6.2.1. Deniz kabuğu veri kümesi 73
6.2.2. Süsen çiçeği veri kümesi 77
6.2.3. Cam kimliklendirme veri kümesi 78
6.2.4. Şarap veri kümesi 83
6.2.5. Tiroid hastalığı veri kümesi 85
6.3. Bölüm Değerlendirmesi 87

7. SONUÇ VE ÖNERİLER 88

7.1. Sonuçlar 88
7.2. Öneriler 90

EKLER 92

EK-1 Süsen Çiçeği (İris) Veritabanı Dosyası 92
EK-2 Süsen Çiçeği (İris) Veri Kümesinin Öznitelik ve Aralık Değerler Dosyası 98

KAYNAKLAR 99

ÖZGEÇMİŞ 108

ŞEKİLLER

Şekil 2.1. Bilgisayar satın alma durumuna yönelik örnek bir karar ağacı yapısı 17
Şekil 2.2. Tenis örneği için oluşan karar ağacı 19
Şekil 3.1. Kural çıkarımı metotlarının sınıflandırma taksonomisi..... 26
Şekil 4.1. Bağışıklık sisteminin sınıflandırılması ve temel elemanları..... 30
Şekil 4.2. Bağışıklık sisteminin çok katmanlı mimarisi 31
Şekil 4.3. Antikor'un yapısı..... 33
Şekil 4.4. Bağışık yanıt sonuçları 37
Şekil 4.5. Bağışıklık sisteminin basit aktivasyon mekanizması 38
Şekil 4.6. YBS' nin katmanlı yapısı 40
Şekil 4.7. Şekil uzayı gösterimi 41
Şekil 4.8. Tanıma çemberi şekil uzayı gösterimi..... 42
Şekil 4.9. CLONALG algoritmasının genel çalışma şeması 46
Şekil 5.1. Önerilen algoritmanın temel adımları ve algoritmanın akış diyagramı..... 51
Şekil 5.2. Kural çözme işleminin kaba kodu 53
Şekil 5.3. INDISC yazılımının arayüz ekranı 54
Şekil 5.4. CLONALG parametreleri..... 55
Şekil 5.5. Parametrelerin antikor popülasyonunda gösterimi 55
Şekil 5.6. Problem parametreleri 57
Şekil 5.7. Programın çalışma esnasındaki oluşan yakınsama grafiği 57
Şekil 5.8. Kuralları oluşturacak sonuç ekranı 58
Şekil 5.9. Öznitelikler ve sınıfların sayısal değerleri..... 58
Şekil 5.10. Çıkan kuralların listesi..... 59
Şekil 5.11. Teşhis edilemeyen sınıf ve öznitelik değerleri 59
Şekil 5.12. Budama sonrası nihai kuralların oluşturulması 60
Şekil 5.13. Örnek uygulama için programın çalışma ekranı..... 61
Şekil 5.14. Örnek uygulama için programın sonuç çıktı ekranı 61
Şekil 5.15. Örnek uygulama için oluşan kurallar..... 62
Şekil 5.16. Budama sonrası kalan kurallar 62

ÇİZELGELER

Çizelge 1.1. Michie ve ark. (1994) çalışmalarındaki sınıflandırma başarıları 4
Çizelge 2.1. Tenis örneği için girdi özellikler 19
Çizelge 4.1. Duyarlılık ölçütü için kullanılabilecek uzaklıklar 43
Çizelge 6.1. Pima yerlileri diyabet hastalığı kümesinin öznitelik ve aralık değerleri 65
Çizelge 6.2. Pima yerlileri diyabet hastalığı veri kümesine ait kurallar 65
Çizelge 6.3. Pima yerlileri diyabet hastalığı veri kümesine ait deneysel başarılar 66
Çizelge 6.4. Wisconsin göğüs kanseri (orijinal) veri kümesinin öznitelik ve aralık değerleri 67
Çizelge 6.5. Wisconsin göğüs kanseri (orijinal) veri kümesine ait kurallar 68
Çizelge 6.6. Wisconsin göğüs kanseri (orijinal) veri kümesine ait deneysel başarılar... 68
Çizelge 6.7. Wisconsin göğüs kanseri (teşhis) veri kümesinin öznitelik ve aralık değerleri 70
Çizelge 6.8. Wisconsin göğüs kanseri (teşhis) veri kümesine ait kurallar 71
Çizelge 6.9. Wisconsin göğüs kanseri (teşhis) veri kümesine ait deneysel başarılar 72
Çizelge 6.10. Deniz kabuğu veri kümesinin öznitelik ve aralık değerleri 74
Çizelge 6.11. Deniz kabuğu veri kümesi kuralları 75
Çizelge 6.12. Süsen çiçeği veri kümesinin öznitelik ve aralık değerleri 77
Çizelge 6.13. Süsen çiçeği veri kümesi kuralları 78
Çizelge 6.14. Süsen çiçeği veri kümesine ait deneysel başarılar 78
Çizelge 6.15. Cam kimliklendirme veri kümesinin öznitelik ve aralık değerleri 79
Çizelge 6.16. Cam kimliklendirme veri kümesi kuralları 80
Çizelge 6.17. Cam kimliklendirme veri kümesine ait deneysel başarılar 83
Çizelge 6.18. Şarap veri kümesinin öznitelik ve aralık değerleri 83
Çizelge 6.19. Şarap veri kümesi kuralları 84
Çizelge 6.20. Şarap veri kümesine ait deneysel başarılar 85
Çizelge 6.21. Tiroid hastalığı veri kümesinin öznitelik ve aralık değerleri 86
Çizelge 6.22. Tiroid hastalığı veri kümesi kuralları 86
Çizelge 6.23. Tiroid hastalığı veri kümesine ait deneysel başarılar 86

SİMGELER VE KISALTMALAR

Simgeler

$\%$	: Yüzde
$<$	: Küçüktür
$<$	: Büyüktür
$=$	: Eşittir
ϵ	: Elemanıdır
L	: Vektör uzunluğu
E	: Duyarlılık eşiği
d	: Kalan yer değiştirme boyutu
β	: Klonal Faktör
G	: Jenerasyon sayısı

Kısaltmalar

ADAP : Adaptive Learning Routine That Generates and Executes Digital Analogs of Perceptron-Like Devices
AIS : Artificial Immune System
APC : Antigen Presenting Cells
BAT : Bağışıklık Ağı Teorisi
BFS : Breadth First Search
BMI : Body Mass Indeksi
BNND : Bayesian Network with Dependence
BNNF : Bayesian Network with Naive Dependence and Feature Selection
C : Constant
C4.5 : Karar Ağacı Oluşturma Algoritması
CAR : Class Association Rules
CBFS : Case-Based Forecasting System
CLONALG : Clonal Selection Algorithm
DNA : Deoxyribonucleic Acid
DTGA : Dextro-Transposition of the Great Arteries
EA : Evrimsel Algoritma
EKG : Elektrokardiyografi
EMA : Evrimsel Memetik Algoritma
FERNN : Fast Extraction of Rules from NN
FN : False Negativies
FP : False Positivies
GA : Genetik Algoritma
GLIM : Generalized Linear Models
GNG : Growing Neural Gas
GSVM : Granular Support Vector Machines
ID3 : Iterative Dichotomiser 3
INDISC : Interval Discovery
IPSO : İkili Parçacık Sürü Optimizasyonu
ITI : Incremental Tree Inducer
KNN : K-Nearest Neighbors
MARS : Multivariate Adaptive Splines

MDA : Mesocyclone Detection Algorithm
MHC : Major Histocompatibility Complex
MSDD : Multi-Stream Dependency Detection
NK : Natural Killer
PSA : Penalized Log Likelihood Smoothing Spline Analysis
RSVM : Reduced Support Vector Machine
SFAM : Simplified Fuzzy Art-Map
SLIQ : Supervised Learning In Quest
SPRINT : Scalable Parallelizable Inndution of Decision Trees
SSVM : Smooth Support Vector Machine
SVM : Support Vector Machine
TACO : Touring Ant Colony Optimization
TAKKO : Tur Atan Karınca Koloni Optimizasyon Algoritması
TCR : T Cell Receptor
TI : Thymus Independent
TN : True Negativies
TP : True Positives
UCI : Irvine California University
YBS : Yapay Bağışıklık Sistemi
YSA : Yapay Sinir Ağları
μGA : Micro-Genetic Algorithm-Based

1. GİRİŞ

Günümüzde bilgisayarların yüksek miktarda veri depolayabilme yeteneği sonucu biriken verilerden anlamlı bilgilerin çıkartılması problemi gündeme gelmiştir. Bu da veri madenciliğine ilgiyi arttırmıştır. Veri madenciliği yöntemleri veriyi bilgiye dönüştürmek ve bu bilgiden yararlanarak yeni bilgilere ulaşmak için kullanılır. Veri sınıflandırma, veri madenciliğinin önemli bir alanıdır. Bugüne kadar araştırmacılar sınıflandırma için değişik teori ve yöntemler üzerine farklı çalışmalar yapmış ve birçok önemli sonuca ulaşmışlardır. Sınıflandırma yöntemleri genellikle şeffaf olmayan yöntemler olduğundan birçok araştırmacı bu metotları kural çıkarımı olarak isimlendirilen şeffaf modellere çevirmeye çalışmışlardır.

Bu çalışmada verilerden kural çıkaran bir yöntem sunulmuştur. Önerilen yöntem kullanılarak iki sınıflı problemlerin yanında, çok sınıflı problemlerden de sınıflandırma kuralları çıkarılabilmektedir. Uygunluk fonksiyonunun optimizasyonu amacıyla Yapay Bağışıklık Sistemi (YBS) algoritmalarından literatürde olan CLONALG algoritması kullanılmıştır. Kural çıkarma çalışmalarında genellikle veriler ikili kodlama kullanılarak kodlanmaktadır. Bunun için gerçel değer alan öznitelikler belirli mantıkla kesikli hale getirilmektedir. Bu çalışmada ise öncekilerden farklı olarak veriler optimizasyon uygulanarak en uygun aralık elde edilerek kodlanmıştır. Yalnız birçok veri kümesi gerçel özniteliklerin yanında ayrık değerler alan özniteliklere de sahip olmaktadır. Örneğin cinsiyet özelliği erkek ve kadın olmak üzere 2 farklı değer alabilmektedir ki, bu özniteliği gerçel kodlamak imkânsızdır. Bu nedenle bu tip özelliklerin ikili olarak kodlanmasının daha uygun olacağı düşünülmüştür. Böylece bu çalışmada kullanılan yöntemin ana fikri gerçel ve ayrık değerlerin farklı kodlanması olmuştur. Bu amaçla ayrık değerler ikili sistemle kodlanmıştır. Gerçel değerler ise iki gerçel sayı kullanılarak kodlanmıştır. Bu sayılar kuralları oluşturan aralıkların genişlemesi ve orta noktasını ifade etmektedir. Önerilen yöntem en uygun aralıkları keşfettiğinden dolayı bu yönteme "Aralık Keşfi" anlamına gelen INDISC (*INterval DISCovery*) kelimesi isim olarak verilmiştir. INDISC yöntemi 8 farklı veri kümesi üzerinde uygulanmıştır. Tüm veri kümelerinde INDISC yöntemi ile elde edilen sonuçların diğer yöntemlerle elde edilen sonuçlardan daha başarılı olduğu görülmüştür.

1.1. Çalışmanın Amacı ve Önemi

Araştırmacılar genellikle verileri ikili veya gerçel olarak kodlamış ve uygulama yapmışlardır. Ama veritabanlarında öznitelikler genellikle iki tiptir: Ayrık ve gerçel. Örneğin bir hastalık dolayısıyla başvuran hastanın kalp ağrısı varsa, bu sadece tipik anjin (kalbe yetersiz kan gelmesi sonucu oluşan ağrı), atipik anjin, anjin olmayan ağrı ve asemptomatik ağrı olmak üzere dört şekilde olabilmektedir veya bir hastanın ailesinde belli bir hastalık ya vardır ya da yoktur. Bu tip veriler ikili kodlamaya çok müsaitken, hastanın tansiyonunun değeri, kolesterol miktarı geniş bir alanda değer alabilmekte ve doktorlar tarafından düşük, normal, yüksek olarak değerlendirilmektedir. Bu düşük, normal ve yüksek değerler göreceli olabilmekte, bir hasta için düşük sayılan 90 tansiyon değeri, örneğin böbrek rahatsızlığı olan birisi için normal bir değer olabilmektedir. Bundan dolayı bu tür verileri ikili değil ayrık ve gerçel öznitelikler farklı şekilde kodlanmıştır. Ayrık öznitelikler ikili olarak, gerçek öznitelikler ise, iki gerçel değer kullanılarak kodlanmıştır.

Bu çalışmadaki temel amaç etkin bir sınıflandırma kural çıkarma yönteminin üretilmesidir. Bu alanda çalışırken bir takım problemlerle karşı karşıya kalınmaktadır. Bunlar veri toplama araçlarındaki eksiklikler, yeterli miktarda verinin bulunmaması, ölçümlerin yanlış yapılmış, bazen yapılmamış veya kaydedilmemiş olmasıdır. Bu olumsuzluklara rağmen bu konuda birçok çalışma yapılmıştır. Ancak halen bu çalışmalarda elde edilen sonuçların tam başarılı olduğu söylenememektedir. Çalışmada önerilen yöntem yüksek hassasiyet ve doğruluk oranı ile sınıflandırma alanına önemli katkı sağlayacaktır.

Önerilen kural çıkarma yöntemi sağlık, fen ve sosyal bilimlerde karar alma, teşhis, analiz ve süreç takibi gibi alanlarda düşük hata payına sahip sistemler oluşturulabilecektir. Örneğin bir bankanın müşteri kredi notu oluşturması, bir sağlık kurumunda hastalığın teşhisinin sağlanmasına katkıda bulunmak ve bir işletmede verimlilik göstergelerinin değerlendirilmesinde kullanılabilir. Önerilen yöntemin geniş yelpazede uyarlanacak olan tüm veri tabanlarında kararlı sonuçlar üretecek olması çalışmanın önemini göstermektedir.

1.2. Literatür Araştırması

İlk kural çıkarma algoritmalarından biri Saito ve Nakano tarafından geliştirilmiştir (Saito ve Nakano, 1988). İkili verilere sahip problemlerde Önce Enine arama (Breadth-first - BFS) algoritması kullanarak bileşik kurallar çıkarmışlardır. Oluşabilecek kural kümelerinin sayısını azaltmak için iki kısıtlama kullanılmıştır. İlk kısıtlama olarak derinlik adı verdikleri bir $t<n$ sayısı belirlemiş ve kural kümelerinin eleman sayısını en fazla t ile sınırlamışlar. İkinci kısıtlama olarak ise verilerden yola çıkarak bazı kuralları elemişler. Örneğin, eğer veri kümesinde $x1$=doğru, $x2$=doğru şeklinde bir veri yoksa $y=x1x2$ şeklinde bir kural hiçbir zaman doğru olmayacağına göre, bu kural baştan elenmiştir. Çalışmada kuralların ağı ifade ettiğini test etmek için aşağıdaki yöntem kullanılmıştır: Önce pozitif girişlere 1, diğer bütün girişlere 0 atanmış ve ağın çıkışı hesaplanmıştır. Sonra değilli girişlere 1 atanmış, diğerleri aynı kalmış ve ağın çıkışı hesaplanmıştır. Eğer ilk denemede sınıflandırma işlemi doğru, ikincide yanlış yapılmışsa, bu kural doğru olarak kabul edilmiştir.

Smith ve ark. (1988) sınıflandırma yapmak için ADAP algoritması kullanmışlar ve yöntemi diyabet veri kümesine uygulamışlardır. Bu veri kümesi 768 hasta verisinden oluşmaktadır. Verilerin 576'sı eğitim, 192'si test için kullanılmıştır. Test verisinde %76 oranında sınıflandırma başarısı elde edilmiştir. Bu veri kümesi üzerinde birçok araştırma yapılmış ve çeşitli sonuçlar elde edilmiştir.

Wahba ve ark. (1992) aynı veritabanına PSA (Penalized log likelihood Smoothing Spline Analysis) ve GLIM (Generalized Linear Models) algoritmalarını uygulamışlar. Glikoz ve Body Mass Indeksi (BMI) sıfır olan hastalar elenerek, kalan 752 veriden 500'ü eğitim, 252'si test için kullanılmıştır. Test verisinde PSA %72, GLIM ise %74 sınıflandırma başarısına ulaşılmıştır.

Gallant (1993), Saito ve Nakano'nun geliştirdiği yönteme benzer bir algoritma geliştirmiştir. Gallant'ın çalışmasındaki fark, kuralların hangisinin ağı ifade ettiğini test etme yöntemidir. Bu yöntemde ağın girişine sadece kuralda belirtilen girişlerin değerleri atanmış ve ağın çıkışı sadece bu girişlere bağlı olarak hesaplanmıştır. Örneğin, üç girişli bir ağda kural denenecekse, ağın birinci girişine 1, üçüncü girişine 0 atanmış, ikinci girişine ise veri atanmamıştır. Gallant bu yöntemin sadece geçerli kuralları bulduğunu iddia etmiştir.

Oates (1994) yine aynı veri tabanına MSDD (Multi-Stream Dependency Detection) algoritmasını uygulamıştır. Verilerin 2/3'ü eğitim, 1/3'ü test için kullanılmıştır. Test verisinde sınıflandırma başarısı %71.33 olarak rapor edilmiştir.

Tchoumatchenko ve Ganascia (1994) sinir ağından çoğunluk-oyu (majority vote) kuralı çıkarmak için yeni bir yöntem önerdi. Çoğunluk oyu kuralı bir sınıfa ait verileri doğrulayan veya doğrulamayan değerler listesidir. Çoğunluk oyu kuralındaki değerlerin çoğu örnekle aynıysa, örnek kuralda belirtilen sınıfa ait kabul edilir. Tchoumatchenko ve Ganascia'nın bu çalışmasında özel bir aktivasyon fonksiyonu kullanarak sinir ağının ağırlıklarının {-1, 0, 1}'e çok yakın değerlere ulaşması sağlanmıştır. Daha sonra algoritma her giriş değerinin belli bir sınıfa ait olup olmadığına bakarak kural için değerler oluşturur. Girilen değerin bir sınıfa ait olup olmadığına ağdaki ağırlıkların işaretlerine bakarak karar verilir. Bu yöntemle her sınıf için bir çoğunluk oyu kuralı oluşturulmuştur. Bu kaba bir kural çıkartma algoritması olup, başarı oranı düşüktür.

Michie ve ark. (1994) bu veritabanına 22 farklı algoritma ve 12 kez çapraz doğrulama uygulamış ve Çizelge 1.1'de belirtilen sonuçları rapor etmişler. Bu algoritmalardan en başarılısının %77.7 doğru sınıflandırma oranıyla lojik diskriminant, en başarısızının %67.6 doğru sınıflandırma oranıyla k-en yakın komşu algoritması olduğu gözlemlenmiştir.

Çizelge 1.1. Michie ve ark. (1994) çalışmalarındaki sınıflandırma başarıları.

Yöntem	Başarı %
Discrim	77.5
Quaddisc	73.8
Logdisc	77.7
SMART	76.8
ALLOC80	69.9
k-NN	67.6
CASTLE	74.2
CART	74.5
IndCART	72.9
NewID	71.1
AC^2	72.4
Baytree	72.9
NaiveBay	73.8
CN2	71.1
C4.5	73
Itrule	75.5
Cal5	75
Kohonen	72.2
DIPOL92	77.6
Backprop	75.2
RBF	75.7
LVQ	72.8

Thrun (1995) Gallant'ın yönteminin genelleştirilmiş ve daha güçlü modeli olan VIA yöntemini sunmuştur. İki yöntem arasındaki temel fark ağın girişlerini tespit etmek için doğrusal programlama kullanmasıdır. Bu da aktivasyon aralığının geriye yayılarak girişlerin hesaplanmasını ve giriş için daha dar aralıkların bulunmasını sağlamaktadır. Bu yöntemde ilk aşamada Gallant gibi ağa sınırlı veriler girilmiş ve çıkış hesaplanmıştır. Sonra çıkışa göre girişleri hesaplamıştır. Bu yöntemle Thrun, Gallant'ın tespit edemediği kuralları bulmuştur.

Craven ve Shavlik (1996) TREPAN algoritmasını önerdiler. Bu algoritma YSA için belirli bir eğitim algoritması veya belirli bir ağ yapısına gerek duymadan, sınıflandırma ağaçları kullanarak kural çıkarmaktadır.

Lu ve ark. (1996) gizli katmandaki aktivasyon değerlerini kümeleyerek kural çıkarmak için bir yöntem önermişlerdir.

Bioch ve ark. (1996), Wahba ve ark. (1992) gibi glükoz ve BMI'si sıfır olan hastaları elemişler ve elde edilen verilere Bayes sinir ağını uygulamışlar. Standart sinir ağının sınıflandırma başarısı %75.4 iken, Bayes sinir ağının sınıflandırma başarısı %79.5 olarak rapor edilmiştir.

Weijters ve ark. (1997) BP-SOM adlı YSA'yı eğiten ve sınıflandırma kuralı çıkaran algoritma önermişlerdir. BP-SOM Kohonen ağını kullanarak kümeleme yapmaktadır.

Jo ve ark. (1997) MDA ve CBFS (Case-Based Forecasting System) ve yapay sinir ağlarını kullanarak Kore firmalarını finans durumuna göre sınıflandırmış ve en başarılı sistemin yapay sinir ağı olduğunu rapor etmişlerdir.

Olmeda ve Fernandez (1997) İspanya bankaları için iflas durumu araştırması yaparken MLFF-BP, lojistik regresyon, MARS (Multivariate Adaptive Splines), C4.5 ve MDA algoritmalarını tek tek ve değişik kombinasyonlu hibritler şeklinde kullanmışlar. Sinir ağının tek kullanıldığı zaman iflas durumundaki bankaları tahmin etmede diğer bütün ağlardan daha başarılı olduğunu ve sinir ağı, lojistik regresyon, C4.5 ve MDA kombinasyonunun aynı problem için diğer bütün kombinasyonlarıdan daha iyi olduğunu rapor etmişlerdir.

Carpenter ve Markuzon (1998) Pima yerlileri diabet veri kümesini kullanmış ve ARTMAP-IC, ARTMAP, lojistik regresyon ve k-en yakın komşu algoritmalarını uygulamışlar. Bu çalışmada ARTMAP-IC'nin başarısı %81, ARTMAP'in başarısı %66, lojistik regresyon ve KNN'nin başarısı %77 olarak rapor edilmiştir.

Eklund ve Hoang (1998) Pima yerlileri diabet veri kümesine 5 farklı algoritma uygulamışlardır. Verilerin %80'i eğitim, %20'si test için kullanılmıştır. Algoritmaların başarısı aşağıdaki gibi rapor edilmiştir: C4.5 algoritması %71.02, C4.5 kural %71.55, ITI %73.16, LMDT %73.51 ve CN2'nin başarısı %72.19 olarak rapor edilmiştir.

Das ve Mozer (1998) ikili kodlanmış verilerle eğitilmiş yinelenen (recurrent) sinir ağından kural çıkarmak için bir yöntem önermişlerdir. Yazarlar bu yöntemin diğer sinir ağı algoritmaları için de geçerli olduğunu belirtmişlerdir.

Duch ve ark. (1998) ağırlıkları {-1, 0, 1} kümesine ait olan bulanık sinir ağından kural çıkarmak için bir yöntem geliştirdiler. Bu çalışmada en az sayıda kural elde etmek için ağırlıkların çoğunun sıfır olmasını sağlayan bir yöntem önerdiler.

Ludermir (1998) Boole ağları için kural çıkarma yöntemi önermiştir.

Liu (1998) sınıflandırma ve birliktelik kuralları madenciliğini CAR (Class Association Rules) algoritmasında birleştirmiştir. CAR modeli %73.1, C4.5 kurallar algoritması %75.5 başarı elde etmiştir.

Duch ve ark. (1999) ÇKP'ye gizli nöronlar ekleyerek yeni mantıksal kurallar çıkarma yöntemi önermişlerdir.

Keedwell ve ark. (2000) sinir ağının giriş değerlerinden kural çıkarmak için genetik algoritma kullanmışlardır.

Setiono ve Leow (2000) FERNN adlı sinir ağından kural çıkarmak için hızlı yöntem geliştirmişlerdir. Önerilen yöntem gizli katman nöronlarının bilgi üzerine etkisinin incelenmesine dayanmaktadır.

Palade ve ark. (2000) sinir ağından kural çıkarmak için aralık yayılmasına dayalı bir yöntem geliştirmişlerdir.

Garcez ve ark. (2000) ayrık girişli sinir ağından monoton olmayan kurallar çıkarmak için yöntem geliştirmişlerdir.

Pham ve ark. (2000) aynı veritabanına RULES-4 algoritmasını uygulamış ve %55.90 sınıflandırma başarısı elde etmişlerdir.

Çiftçi (2001) çalışmasında, büyük boyutlu sınıflandırma problemlerin çözümü için temelinde çokyüzlü konik fonksiyonlar olan yeni bir matematiksel programlama yaklaşımı sunmuştur. Önerdikleri yaklaşımda, problemlerin çözümü için K-Ortalamalar ve Gürbüz doğrusal programlama yaklaşımları kullanılmıştır. Literatürde en sık karşılaşılan büyük boyutlu problemler, hem geliştirilen yeni yaklaşım ile hem de alanda en yaygın kullanılan başarıları kanıtlanmış yöntemler ile çözmeye çalışmışlardır.

Snyders ve Omlin (2001) adaptif ve sabit biaslı sinir ağından kural çıkarmış ve sonuçları karşılaştırmışlar. Test için moleküler biyoloji verileri kullanılmıştır.

Cheung (2001) Karaciğer Hastalıkları veri tabanında sınıflandırma yapmak için 4 algoritma kullanmıştır. Çalışmada C4.5 algoritması kullanılarak %65.5, Naive Bayes sınıflandırması kullanılarak %63.39, BNND (Bayesian Network with Dependence) kullanılarak %61.83 ve BNNF (Bayesian Network with Naive Dependence ve Feature Selection) kullanılarak %61.42 sınıflandırma başarısı elde edilmiştir.

Lee ve Mangasarian (2001a) aynı veritabanına SSVM (Smooth Support Vector Machine) algoritması uygulamış ve %70.33 sınıflandırma başarısı elde etmişlerdir.

Lee ve Mangasarian (2001b) aynı veritabanına RSVM (Reduced Support Vector Machine) algoritması uygulamış ve %74.86 sınıflandırma başarısı elde etmişlerdir.

Dorado ve ark. (2002), kural araştırma tekniği olarak Genetik Programlamayı kullanan ve her türdeki YSA'a uygulanabilen bir kural çıkarım sistemi önermişlerdir. Geliştirdikleri sistemin hiçbir yapay sinir ağı mimari ve eğitim ihtiyacı yoktur. Yaptıkları deneysel çalışmalarda yöntemlerinin doğru sonuçlar verdiğini ve her türdeki yapay sinir ağlarına uygulanabildiğini bildirmişlerdir.

Van Gestel ve ark. (2002) SVM kullanarak aynı veritabanında %69.2 sınıflandırma başarısı elde etmişlerdir.

Elalfi ve ark. (2004) genetik algoritmayı kullanarak eğitilmiş sinir ağları aracılığıyla veritabanlarından doğru ve anlaşılır kurallar çıkarmak için yeni bir algoritma sunmuşlardır. Ortaya koydukları algoritma, YSA eğitim algoritmalarına bağlı olmamakla birlikte, eğitim sonuçlarını da değiştirmemektedir. Genetik algoritma, k çıkış düğümünün Ψ_k çıktı fonksiyon değerini maksimum yapan, X_m giriş özelliklerinin (kromozom) en iyi değerlerini elde etmek için kullanılmıştır. Ψ_k fonksiyonu, sinir ağındaki sırasıyla giriş katmanı ile gizli katman arasındaki ve gizli katman ile çıkış katman arasındaki ağırlık değerlerine bağlı, doğrusal olmayan üstel bir fonksiyondur. En iyi kromozom, k sınıfına ait kuralı elde etmek için kullanılmıştır. Bu yaklaşımda bütün giriş özellikleri dikkate alındığı için doğru kurallar ortaya konmaktadır.

Tang ve ark. (2004) GSVM (Granular Support Vector Machines) algoritmasını önermiş ve bu algoritmayı Wisconsin Göğüs Kanseri, Cleveland Kalp Hastalıkları, BUPA Karaciğer Hastalıkları veritabanlarında denemişlerdir. Göğüs Kanseri veritabanında hastalık teşhisi için GSVM %97.72, SVM %96.4, Cleveland Veri Tabanında GSVM %84.04, SVM %83.84, BUPA veritabanında GSVM %49.66, SVM %48.97 ile başarılı olmuştur.

Loo (2005) Fuzzy ARTMAP (SFAM) algoritmasını önermiş ve bu yöntemi 5 tıbbi veri tabanına uygulamıştır. Dermotoloji veri tabanında doğru teşhis oranı %93.14, hepatit veritabanında %75.16, Wiskonsin Göğüs Hastalıkları veritabanında %93.14, Cleveland Kalp Hastalıkları veritabanında %75.16 ve Kalaazar hastalıklarında %89.21 olarak elde edilmiştir.

Tokinaga ve ark. (2005) çalışmalarında, genetik programlamayı temel alan sinir ağı kural çıkarımı tekniğini ortaya koymuşlardır. Basit ve uygun sınıflandırma kuralları elde etmek için sinir ağlarını ve ikili sınıflandırmayı kullanmışlardır. Kuralları oluşturan basit ancak sağlam ikili ifadeler nöronlar arasındaki ağırlıkların budanması ile elde edilmektedir. Yöntem iflas tahmini için kural çıkarımı üzerine uygulamışlardır.

Shin ve ark. (2005), şirketlerin iflas durumunu araştırmak için Destek Vektör Makinesi (Support Vector Machines – SVM) kullanmışlar ve şirketlerin mali durumuna göre sınıflandırılmasında SVM'nin MLFF-BP'den daha başarılı olduğunu açıklamışlardır. SVM'nin MLFF-BP'ye göre daha az veriyle eğitilebilmesinin bir avantaj teşkil ettiğini not etmişlerdir.

Markowska-Kaczmar (2005), çalışmasında yapay sinir ağlından kural çıkarmak için genetik algoritmalara dayanan GEX (Genetic Rule Extraction) algoritmasını geliştirmiştir ve metodun parametrelerinin son sonuçlardaki etkilerini deneysel çalışmalarla incelemiştir. İlk olarak genetik parametrelerin etkilerini incelemiş, daha sonra ise kural çıkarımının etkinliğini etkileyen parametreleri test etmiştir. İkili, sürekli ve kesikli değişkenler için farklı kromozom yapıları kullanmıştır. GEX algoritmasında her popülasyonu, bir sınıfın kurallarını içerecek şekilde özelleştirmiştir. Bu da sınıflandırma probleminde mevcut olan sınıf sayısı kadar popülasyon oluşturulmaktadır. Popülasyondaki her bir birey bir kuralı ifade etmektedir. Deneysel çalışmaları farklı tipteki örnekler için GEX algoritmasının kural çıkarmakta esnek bir metot olduğunu açıklamışlardır.

Tan ve ark. (2006), sınıflandırma metotlarını kural tabanlı ve kural tabanlı olmayan metotlar olmak üzere ikiye ayırmıştır. Kural tabanlı metotlara C4.5, karar tabloları, ID3 gibi algoritmalar örnek verilebilir. Yapay sinir ağları ve destek vektör makineleri ise kural tabanlı olmayan metotlara örnektir. Kural-tabanlı sınıflandırma metotları veriden gizli bilgiyi doğrudan çıkarırlar ve kullanıcılar bu bilgileri kolaylıkla anlayabilirler. Kural tabanlı olmayan sınıflandırma metotları ise genellikle kural tabanlı sınıflandırma metotlarına göre daha doğru sonuçlar verirler fakat kara-kutu gibi davrandıklarından dolayı anlaşılırlık yönünden rekabetçi değillerdir.

Hruschka ve Ebecken, (2006) çalışmalarında kümeleme genetik algoritmasını ortaya koymuşlardır. Bu kural çıkarım algoritması iki basamaktan oluşmaktadır. İlk olarak, gizli birim aktivasyon değerlerinin kümelerini belirlemek için kümeleme genetik algoritması uygulanmaktadır. Daha sonra, girişler ile ilgili olan, bu kümeleri ifade eden sınıflandırma kuralları ortaya konulmaktadır. Yaptıkları bu çalışmalarda dört farklı kıyaslama problemi kullanmışlar ve literatürdeki en iyi sonuçlarla kıyaslanacak ölçüde sonuçlar elde etmişlerdir.

Avşar (2007), çalışmasında, kaba küme teorisi kullanılarak eksik veri kümelerinden bilgi çıkarımı yapmıştır. Önemsiz ve kayıp olmak üzere, iki farklı tipte eksik veri içeren, bir veri kümesi ile çalışılmıştır. Nicel değerli eksik veri kümelerinden, kaba küme yaklaşımı ile bulanık kurallar çıkaran önerdikleri bir algoritma ile bilgi çıkarımı sağlanmışlardır. Geliştirdikleri algoritma için tiroit hastalığı verileri üzerinde çalışan bir uygulama yazılımı oluşturulmuştur. Eksik özniteliklerin sayıları ve türleri üzerinde yapılan değişikliklerle elde edilen altı farklı durum için yazılım test edilmiştir.

Coenen ve Leng (2007), birliktelik kuralı madenciliği parametrelerinin, sınıflandırma tahminleyici doğruluğu üzerindeki etkilerini incelemişlerdir. Parametrelerin uygun seçilmesiyle doğruluğun iyileşeceğini göstermiş ve en iyi parametre seçimi için bir tırmanış metodu sunmuşlardır.

De Falco ve ark. (2007), çok sınıflı veritabanlarında sınıflandırma problemini çözmekte parça sürü optimizasyonu algoritmasının kullanımını ele almışlardır. PSO'nun üç farklı versiyonuyla üç farklı uygunluk fonksiyonunu değerlendirmişlerdir ve en iyi için PSO'yu dokuz farklı teknik ile karşılaştırmışlardır. Elde ettikleri sonuçlar PSO algoritmasının rekabetçi olduğunu göstermiştir.

Patterson ve Zhang (2007), sınıf dengesizliği problemleri içeren ikili sınıflandırma için bir genetik programlama yaklaşımı geliştirmişlerdir. Elde ettikleri sonuçlar, uygunluk fonksiyonu olarak doğruluk kullanıldığında, Genetik Programlama sisteminin baskın sınıfa yöneldiğini göstermiştir. Sınıf dengesizliği probleminin üstesinden gelmek için yeni uygunluk fonksiyonları önermişlerdir. Deneysel çalışmaları önerdikleri her bir uygunluk fonksiyonunun baskın olmayan sınıf performansını artırdığını bildirmiştir.

Thabtah ve Cowling (2007), çoklu sınıflı kurallar çıkaran, sıralanmış çoklu sınıflı kural (RMR) ismini verdikleri yeni bir birliktelikçi sınıflandırma tekniği geliştirmişlerdir. Algoritmaları kuralların çakışmasını engellemektedir. Yirmi farklı veri kümesi

üzerindeki deneysel çalışmaları sistemlerinin, çoklu sınıflarla ilgili kurallar içeren sınıflandırıcılar üretebildiğini göstermektedir.

Thangavel ve Jaganathan (2007), sınıflandırma kuralı çıkarımına TACO-miner ismini verdikleri yeni bir karınca koloni optimizasyonu algoritması önermişlerdir. Algoritmalarının temel amacı, basit ve anlaşılır sınıflandırma kuralları çıkarmaktır. Geliştirdikleri algoritmayı, literatürde yer alan ve sınıflandırmada karınca koloni optimizasyonu algoritmasını kullanan algoritmalarla karşılaştırmışlar ve daha iyi sonuçlar elde ettiklerini belirtmişlerdir.

Zahiri ve Seyedin (2007), etkin bir sürü zekası tabanlı sınıflandırıcı geliştirmekte parça sürü optimizasyonu algoritmasından faydalanmışlar ve geliştirdikleri algoritmayı IPS- classifier olarak adlandırmışlardır. Üç farklı veri kümesi üzerindeki analizleri algoritmalarının çok katmanlı algılayıcılardan ve k-en yakın komşu algoritmasından daha iyi sonuçlar elde ettiğini belirtmişlerdir.

Aliev ve arkadaşları (2008), etkin ve hızlı batarya şarjının kontrol kurallarının çıkarımı için esnek hesaplama tabanlı bir veri madenciliği tekniği kullanmışlardır. Elde edilen kurallar NiCd batarya şarjı için kullanılmıştır.

Delice (2008), çalışmasında eğitilmiş YSA'dan sınıflandırma kuralları çıkarmak için İkili PSO (İPSO) algoritması geliştirilmiştir. En uygun YSA ağırlık değerlerini kullanarak parçacıkların arama uzayını etkin bir biçimde taraması sağlanmıştır. Geliştirilen algoritma ilk olarak eğitim verisini ele almakta ve bu veri kümesini ifade eden kurallar çıkarılmaktadır. Daha sonra eğitim veri kümesinden bağımsız olan test veri kümesine elde edilen kuralların uygulanmasıyla kural kümesinin doğruluk değeri belirlenmektedir. Önerilen algoritma on bir farklı referans veri kümesi üzerinde deneysel olarak değerlendirilmiş ve literatürde yer alan klasik makine öğrenme algoritmaları ile karşılaştırılmıştır.

Dehuri ve ark. (2008), büyük boyutlu veri tabanlarından sınıflandırma kuralları çıkarmak için birçok amaçlı genetik algoritma (EMOGA) geliştirmişlerdir. Kuralların doğruluğu ve anlaşılırlığı üzerinde durulmuştur. Sonuçlarını basit genetik algoritma ile karşılaştırmış ve kendi algoritmalarının basit GA üzerindeki üstünlüğünü ortaya koymaya çalışmışlardır.

Karabatak (2008), çalışmasında birliktelik kuralı yöntemi farklı alanlara uygulanmış ve gösterdiği başarımlar değerlendirilmiştir. Bu doğrultuda üç farklı uygulama yapılmıştır:

- Herhangi bir veri tabanına ait öznitelikler arasındaki ilişkiler tespit edilerek özellik seçimi uygulaması yapılmıştır. Birliktelik kuralına dayalı özellik seçimi yöntemi önerilerek elde edilen sonuçlar diğer özellik seçimi yöntemleri ile karşılaştırmışlardır.
- Birliktelik kuralı, veri tabanına ait öznitelikler arasındaki ilişkileri ortaya çıkarabilmektedir. Bu özelliği sayesinde, birliktelik kuralı yöntemi kullanılarak doku sınıflama işlemi gerçekleştirilmiştir. Bu işlemlerden hız ve başarım artışı sağlayabilmek için, kenar çıkarma ve dalgacık dönüşümü yöntemleri ile dokudan özellik çıkarımı yapılmıştır.
- Birliktelik kuralı, öğrenci verilerine uygulanarak öğrencilerin derslerden aldığı notlar analiz edilmiş ve geleceğe yönelik notlar ile ilgili öngörüler yapılmıştır. Ayrıca bu amaca uygun bir yazılım geliştirilmiştir.

Yapılan bu üç ayrı uygulama alanında birliktelik kuralı yönteminin önemli başarılar elde ettiği görülmüştür. Elde edilen sonuçların literatürde yer alan diğer yöntemlerle karşılaştırılması sonucu, birliktelik kuralı yönteminin, söz konusu özellik seçimi, doku sınıflama ve not öngörüsü uygulamalarındaki başarısından bahsedilmiştir.

Otero ve ark. (2008), sürekli değişkenlerle başa çıkmakta Ant-Miner algoritmasının bir versiyonu olan cAnt-Miner algoritmasını geliştirmişlerdir. Sürekli değişkenleri ele almak için kural kurulum sürecinde entropi tabanlı kesiklendirme metodunu kullanmışlardır. Algoritmalarını Ant-Miner algoritması ile 8 farklı veri kümesi üzerinde çalışmışlardır. Elde ettikleri sonuçlara göre geliştirdikleri algoritma daha doğru ve basit sınıflandırma kuralları çıkarttığını ifade etmektedir.

Kulluk (2009) çalışmasında sınıflandırma problemleri için, eğitilmiş yapay sinir ağlarından bilgi kazanımına yönelik bir algoritma geliştirmiştir. Geliştirilen algoritma, YSA yapısında bağlantı ağırlıkları formunda bulunan gizli bilgiyi keşfetmek için eğitilmiş yapay sinir ağları üzerinde çalışmaktadır. Önerilen algoritma, temelde Tur Atan Karınca Koloni Optimizasyon Algoritması (TAKKO) olarak bilinen bir meta-sezgisel yönteme dayanmaktadır ve iki-adımlı hiyerarşik bir yapıya sahiptir. İlk adımda çok katmanlı algılayıcı tipi sinir ağı eğitilmekte ve ağırlıkları çıkarılmaktadır. Ağırlıklar elde edildikten sonra, ikinci adımda TAKKO algoritması sınıflandırma kurallarının üretimi için kullanılmaktadır. Önerdikleri algoritma deneysel olarak on iki ikili ve çok-sınıflı referans veri kümesinde analiz edilip, değerlendirilmiştir. Bu deneysel çalışmalar diğer

klasik ve modern kural çıkarım algoritmalar ile karşılaştırmış ve sonuçları yorumlamışlardır.

Özbakır ve ark. (2009), çalışmalarında yapay sinir ağlarından kural çıkarmak için TACO-miner isimli yeni bir meta-heuristik algoritma önermişlerdir. Önerilen algoritma temelde gezici karınca kolonisi algoritması (touring ant colony optimization -TACO) diye bilinen meta-heuristik yöntemee dayanır ve iki basamaklı hiyerarşik yapıdan oluşmaktadır. Önerilmiş yöntem altı veri kümesi üzerinde test edilmiş ve başarıları diğer literatür çalışmlarıyla kıyaslamışlardır.

Kahramanlı H. ve Allahverdi N. (2009a) hibrit sinir ağından kural çıkarmak için yapay bağışıklık algoritması olan OptaiNet kullanmıştır. Yöntemi test etmek amacıyla iki veri kümesi kullanılmıştır. Kullanılan veriler Kaliforniya Üniversitesi Makine deposundan elde edilen göğüs kanseri ve EKG veri kümeleridir. Önerilen bu metot sırasıyla EKG için % 94.59, göğüs kanseri için %92.31 sınıflandırma doğruluğuna ulaşmıştır.

Kahramanlı H. ve Allahverdi N. (2009b), adaptif yapay sinir ağından kural çıkarmak için yine yapay bağışıklık algoritması OptaiNet kullanmıştır. Yöntemi test etmek amacıyla iki veri kümesi kullanılmıştır. Verileri Kaliforniya Üniversitesi Makine veri deposundan elde etmiştir. Veri kümeleri Cleveland kalp hastalığı ve hepatit verileridir. Önerilen metot Cleveland kalp hastalığı veri kümesi ve hepatit veri kümesinde sırasıyla %96.4 ve %96.8 doğruluk değerlerini elde etmiştir.

Özbakır ve ark. (2010), "DIFACONN-Miner" olarak adlandırılan yeni bir algoritma önermişlerdir. Bu algoritmada YSA'nın eğitimi için diferansiyel gelişim algoritması ve kural çıkarımı için ise karınca kolonisi algoritmasını kullanılmıştır. Sonuçları ve başarıları diğer literatür çalışmlarıyla kıyaslamışlardır.

Ang ve arkadaşları (2010), kural çıkarımı için evrimsel algoritmaların (EA) global araştırma kapasitesini tamamlamak için yerel arama yoğunluk şeması kullanan evrimsel memetik algoritma (EMA) önermişler. Onlar yerel arama için iki şema kullanmışlar: EMA-μGA olarak isimlendirilen yöntem Mikro genetik algoritma (GA) tekniği kullanırken, yapay bağışıklık siteminden (AIS) esinlenerek hücre üretimi için klonal seçme kullanan diğer yöntem EMA-AIS olarak adlandırılmıştır.

Papageorgiou (2011), birçok farklı bilgi çıkarım yöntemi kullanarak verilerden mevcut bilgiyi bulanık kurallar olarak çıkarmış ve bunu bulanık bilişsel haritalarında (FCM) kullanmıştır.

Rodriguez ve arkadaşları (2011) sınıflandırma kural çıkarımı için ada modeli tabanlı bir dağıtık iyileştirilmiş genetik algoritma sunmuştur.

Turna (2011), çalışmasında Türkiye'deki bir tramvay işletmesinin 4 yıllık tramvay arıza kayıtları veri kümesi üzerinde çalışma yapmıştır. Bu veri kümesi üzerinde veri madenciliği teknikleri uygulanarak, sefere engel arızaların meydana gelmesiyle ilgili kuralların çıkarılması amaçlanmıştır. Eldeki ham veri; veri temizleme, birleştirme ve dönüştürme işlemlerine tabi tutulmuş, kaba kümeler yaklaşımını kullanan Rosetta programı ile indirgenmiş ve indirgenmiş veriyi sınıflandırmak için Weka'nın karar ağaçları ve kural bulma algoritmaları kullanılmıştır. Sonuçta, anlamlı ve yararlı kurallar bulunmuş ve yorumlanmıştır.

Sarkar ve arkadaşları (2012), karar verme ağacı ve genetik algoritma (DTGA) olarak isimlendirdikleri bölge, büyüklük, boyutluluk ve sınıf dağılımından bağımsız herhangi bir sınıflandırma problemleri üzerinde kestirim doğruluğunu iyileştirmeyi amaçlayan bir doğruluk tabanlı eğitim sistemi kurmuşlardır.

Kınacı (2013) çalışmasında eğitilmiş bir yapay sinir ağının kapalı kutu özelliğini ortadan kaldıracak bir eğitimsel, bağlantıcı ve melez bir yaklaşım önermiştir. Önerilen yaklaşım YSA'yı kapalı kutu olarak kabul etmekte ve sadece girdi çıktı değerlerini ele almaktadır. Bağlantıcı bir algoritma olan Growing Neural Gas (GNG) kullanılarak, eğitilmiş ağ tarafından temsil edilen kural karar bölgelerinin topolojileri bulunmaktadır. GNG tarafından bulunan, topolojileri ifade eden çizge kullanılarak kural kümesi oluşturulmaktadır. GNG algoritmasının parametrelerinin problem veri kümesine göne en iyileştirilmesi için genetik algoritmalardan faydalanılmakta ve böylece melez bir zeki sistem oluşturulmaktadır. Geliştirilen yöntemin pratikteki başarısının ölçülmesi için çok katmanlı algılayıcı yapay sinir ağı modeli ile iris, Ecoli ve Wisconsin teşhis göğüs kanseri veri kümeleri üzerinde k-kat çapraz geçerleme yaklaşımı kullanılarak performans ölçümleri yapmıştır.

Görüldüğü üzere sınıflandırma ve kural çıkarma konusunda birçok çalışma yapılmıştır. Bu çalışmalar tahmin konusunda yer yer başarılı olmalarına rağmen özellikle tıp gibi gerçek dünya problemlerinde arzu edilen teşhis başarısına ulaşılamamıştır. Bu çalışmada bu amaçla bir yöntem önerilmiş ve sonuç başarısının yükseldiği gözlemlenmiştir.

1.3. Tezin Organizasyonu

Bu tez çalışmasının ana hatları aşağıdaki gibidir:

Birinci bölümde, tez çalışması hakkında genel bir bakış açısı verilmeye çalışılmıştır. Çalışmanın amacı ve çerçevesi ele alınmış, mevcut çalışmalara göre değerlendirilmiştir.

İkinci bölümde, sınıflandırma ile ilgili temel kavramlardan bahsedilmiş ve sınıflandırma algoritmaları üzerinde durulmuştur.

Üçüncü bölümde kural çıkarımı hakkında bilgiler verilmiştir. Kural çıkarımının önemi ve avantajları konusu anlatılmıştır. Ayrıca literatüre girmiş kural çıkarma yöntemleri hakkında açıklamalar yapılmıştır.

Dördüncü bölümde, insan bağışıklık sitemleri hakkında temel bilgiler verilmiştir. Ayrıca yapay bağışıklık sistemleri hakkında ve bu konudaki algoritmalardan bahsedilmiştir.

Beşinci bölümde önerilen kural çıkarma yöntemi INDISC anlatılmıştır. Bu yöntem için yazılmış olan INDISC programı tanıtılmış ve örnek uygulaması anlatılarak programın kullanımı açıklanmıştır.

Altıncı bölümde, önerilen algoritma ve yazılan program UCI veri deposundan indirilen çok bilinen 3'ü ikili sınıf ve 5'i, çok sınıflı veriler üzerinde uygulamalar yapılmıştır. Bu uygulamaların deneysel sonuç ve analizleri açıklanmıştır

Yedinci bölümde ise, çalışmanın sonuçları üzerine genel bir değerlendirme yapılmış ve çalışma ile ilgili öneriler verilmiştir.

2. SINIFLANDIRMA

Sınıflandırma ve tahmin, geleceğe ait veri tahmini yapmak veya önemli veri sınıflarını açıklayan modeller çıkarmak için kullanılan iki farklı veri analiz biçimidir. Sınıflandırma, kategorik tanımlamalar yaparken, tahmin, sürekli-değerli fonksiyonları modellemektedir (Han ve Kamber, 2001). Örneğin bir sınıflandırma modeli banka kredi uygulamalarının güvenli ya da riskli olduğunu kategorize etmek için inşa edilebilirken, tahmin modeli, gelir düzeyi ve meslek bilgilerinden yararlanarak, potansiyel müşterilerin bilgisayar malzemeleri tüketimini öngörmek için inşa edilebilir.

Sınıflandırma modeli iki adımdan oluşmaktadır. İlk adımda sınıflandırma algoritması kullanılarak gözlenmiş verilerden sınıflandırma kuralları oluşturulur. İkinci adımda ise oluşturulan sınıflandırma kuralları kullanılarak farklı ve yeni veriler sınıflandırılır. Araştırmacılar tarafından makine öğrenmesi, uzman sistemler, istatistik ve nöro-biyoloji için birçok sınıflandırma ve tahmin metodu önerilmiştir. Günümüzdeki veri madenciliği araştırmaları, ölçeklenebilir sınıflandırma geliştirme ve geniş disk kaplayan verileri işleme yeteneğine sahip tahmin teknikleri üzerine inşa edilmiştir. Genel olarak bu teknikler paralel ve dağıtılmış işlemleri göz önünde bulundurarak çalışmaktadır (Han ve Kamber, 2001).

Kaynak araştırmasından da görüldüğü gibi büyük veri tabanları üzerinde sınıflandırma yapmak için birçok yöntem önerilmiş ve değişik alanlarda uygulamalar yapılmıştır. Bazı yöntemler belirli alanlarda başarılı olsa da her alanda başarılı olan bir yöntem henüz bulunamamıştır.

2.1. Sınıflandırma Algoritmaları

2.1.1. Karar ağaçları

Karar Ağacı yaygın kullanılan tahmin yöntemlerinden bir tanesidir. Ağaçtaki her düğüm bir özellikteki testi gösterir. Düğüm dalları testin sonucunu belirtir. Ağaç yaprakları sınıf etiketlerini içerir. Karar ağacı çıkarımı iki aşamadan oluşur (Han ve Kamber, 2001).

1. Ağaç oluşturulması (*Tree Construction*).

Başlangıçta bütün öğrenme örnekleri kök düğümdedir.

Örnekler seçilmiş özelliklere göre tekrarlamalı olarak bölünür.

2. Ağaç Budama (*Tree pruning*).

Gürültü ve istisna kararları içeren dallar belirlenir ve budanır.

Karar ağaçları, her bir dahili düğümün özellik için testi belirttiği, her bir dalın testin sonucunu gösterdiği ve yaprak düğümlerin de sınıf yada sınıf dağıtımlarını gösterdiği, ağaç yapısındaki akış şemalarıdır. En üst seviyedeki düğüm ise kök düğüm adını alır.

Bilinmeyen bir örneği sınıflandırmak için, örneğin özellik değerleri karar ağacına karşı test edilir. İzlenen yol kökten, o örnek için sınıf tahminini tutan yaprak düğüme doğrudur.

Karar ağacı yöntemine ait temel algoritma, tepeden alta tekrarlamalı ve bölüp alma biçiminde karar ağacı inşa eden açgözlü bir algoritmadır (Han ve Kamber, 2001). Karar ağaçları kolaylıkla eğer o halde formunda sınıflandırma kurallarına dönüştürülebilir.

Karar Ağaçlan, verileri belirli özellik değerlerine göre sınıflandırmaya yarar. Bunun için algoritmaya belirli özellikler girdi olarak verilirken belli bir özellik de çıktı olarak verilir ve algoritma bu çıktı özelliğindeki değerlere ulaşmak için hangi girdi özelliklerin hangi değerlerde olması gerektiğini ağaç veri yapıları kullanarak keşfeder.

Karar ağacı, gürültülü veri kümesinde de etkili çalışır ve sınıfların ayırıcı özelliklerini keşfeder. Karar ağacı algoritmalarını bir probleme uygulayabilmek için aşağıdaki koşulların sağlanması gerekir (Aydoğan, 2003):

1. Olayların özelliklerle ifade edilebilmesi ve nesnelerin belli sayıda özellik değerleriyle ifade edilebilmesi gerekir. Örneğin; soğuk, sıcak vb.

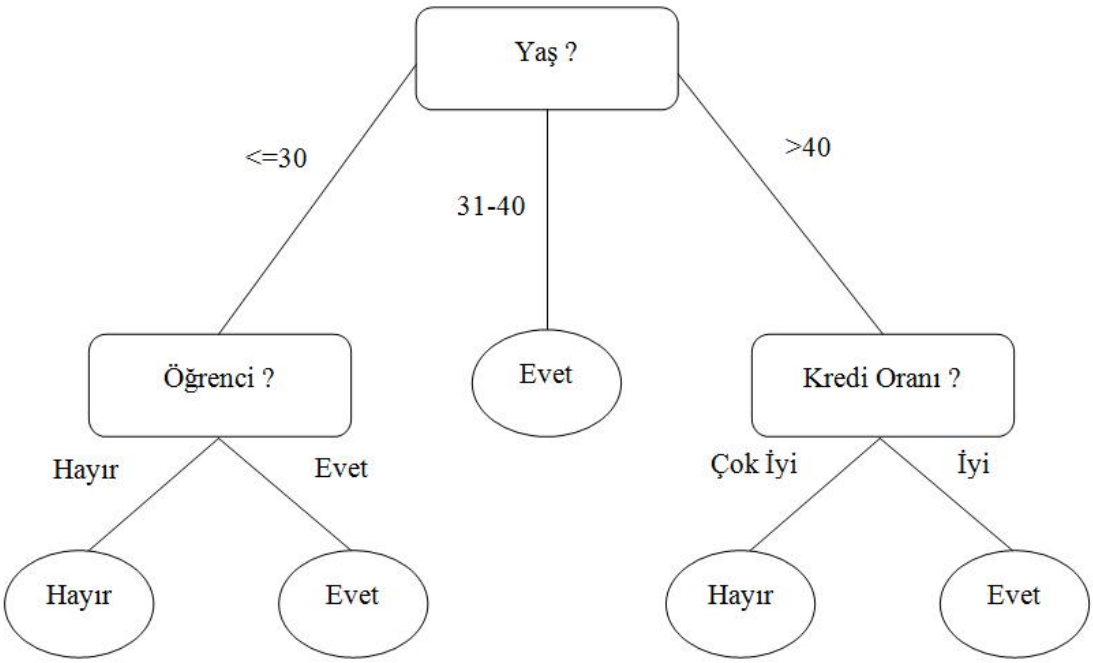
2. Sınıfları belirleyebilmek için gereken ayırıcı özelliklerin olması gerekir.

2.1.1.1. Karar ağaçlarından sınıflandırma kuralı çıkarımı

Karar ağaçlarında sunulan bilgi, Eğer ... O halde sınıflandırma kuralları yolu ile elde edilebilir ve sunulabilir. Kökten yaprak düğüme doğru olan her bir yol için bir kural

çıkarılır. Kuralın önde gelen kısmında (eğer kısmı), verilen yol boyunca her bir özellik-değer çifti birleşim oluşturur. Kuralın sonda gelen kısmında (o halde kısmı), yaprak düğüm sınıf tahminini gerçekleştirir. Özellikle verilen ağaç yapısı çok büyük ise, eğer - o halde kurallarını anlamak insanlar için daha kolaydır (Han ve Kamber, 2001).

Örnek:



Şekil 2.1. Bilgisayar satın alma durumuna yönelik örnek bir karar ağacı yapısı

Şekil 2.1' deki karar ağacı yapısı, ağaçtaki kök düğümden her bir yaprak düğüme doğru ilerleyerek eğer – o halde sınıflandırma kurallarına dönüştürülebilir. Elde edilecek kurallar şu şekilde olacaktır:

Eğer yaş = "<=30" ve öğrenci mi = "Hayır" O halde bilgisayar alımı = "Hayır"

Eğer yaş = "<=30" ve öğrenci mi = "Evet" O halde bilgisayar alımı = "Evet"

Eğer yaş = "31 - 40" O halde bilgisayar alımı = "Evet"

Eğer yaş = ">40" ve kredi oranı = "çok iyi" O halde bilgisayar alımı = "Hayır"

Eğer yaş = ">40" ve kredi oranı = "iyi" O halde bilgisayar alımı = "Evet"

2.1.1.2. Ölçeklenebilirlik ve karar ağacı sonuç çıkarımı

ID3 ve C4.5 gibi mevcut karar ağacı algoritmalarının etkinliği nispeten küçük veri kümeleri için kanıtlanmıştır (Han ve Kamber, 2001). Etkinlik ve ölçeklenebilirlik, bu

algoritmaların büyük miktardaki gerçek hayat veritabanına uygulandığında sorun olmaktadır. Çoğu karar ağacı algoritmasının eğitim örneklerinin hafızada yer kaplama bakımından önemli kısıtlamaları vardır. Veri madenciliği uygulamalarında genellikle milyonlarca örneğe ait çok geniş eğitim kümelerinin olması yaygın bir durumdur. Bu nedenle, eğitim örneklerinin ana bellek ve önbellek (cache bellek) arasında değiş tokuş edilmesi nedeniyle karar ağacı yapısının etkisiz duruma gelmesi şeklindeki kısıtlaması, bu gibi algoritmaların ölçeklenebilirliğini kısıtlamaktadır (Han ve Kamber, 2001).

Karar ağacı sonuç çıkarımı için önceki stratejiler, her bir düğümdeki örnek veri ve sürekli özelliklerin ayrıklaştırılmasını içermektedir. Bu durum, eğitim kümesinin hafızaya uygun olacağını varsaymaktadır. Bir alternatif metot öncelikle, veriyi her birisinin hafızaya sığdığı bireysel alt kümelere ayırmak ve daha sonra her bir alt kümeden karar ağaçları oluşturmaktır. Son çıktı sınıflandırıcı, alt kümelerden elde edilen her bir sınıflandırıcıyı birleştirir. Bu metot büyük veri kümelerinin sınıflandırılması için uygun olmasına rağmen, sahip olduğu sınıflandırma doğruluğu, tüm veriyi kullanarak inşa edilmiş tek bir sınıflandırıcı kadar yüksek bir doğruluğa sahip olamamaktadır (Han ve Kamber, 2001).

Son yıllarda ölçeklenebilirlik sorununa yönelik karar ağacı algoritmaları önerilmiştir. Çok geniş eğitim kümelerinden karar ağacı sonuç çıkarımı için gerekli SLIQ ve SPRINT isimli algoritmalar, kategorik ve sürekli değerli özellikleri ele alabilmektedir. Her iki algoritma hafızaya sığamayacak büyüklükteki ve diskte yer alan veriler üzerine ön sıralama teknikleri önermektedir. Her ikisi de ağaç yapısını kolaylaştıracak şekilde yeni veri yapısı kullanımını önermektedir (Han ve Kamber, 2001).

2.1.1.3. ID3 ve C4.5 karar ağaçlarının kurulması

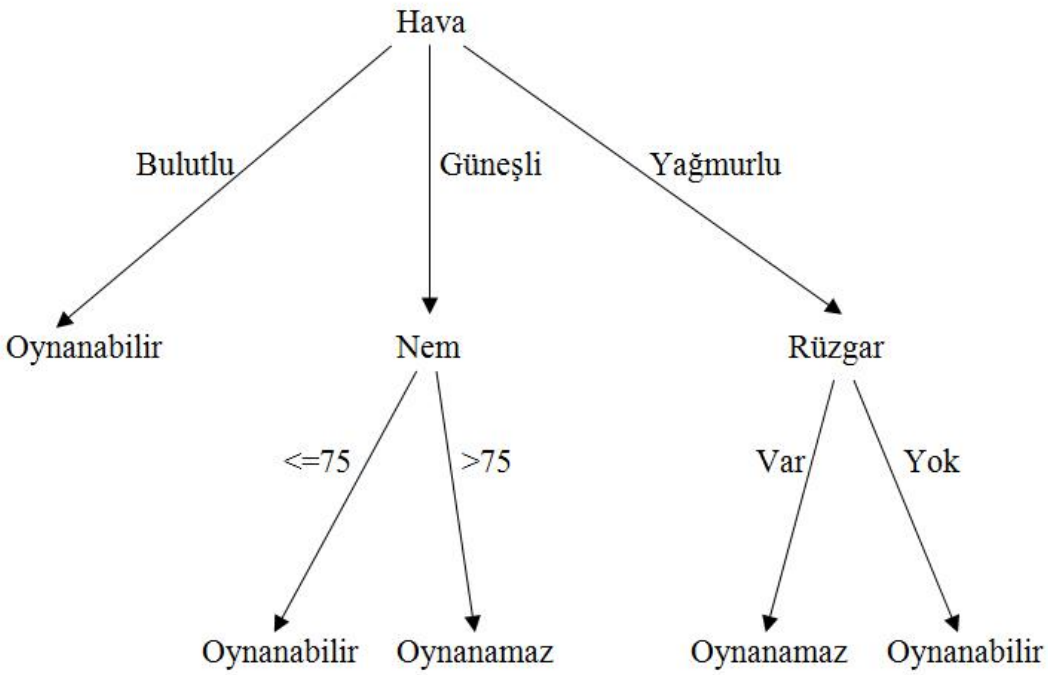
Karar ağaçları olarak da adlandırılan ID3 ve C4.5 algoritmaları, sınıflandırma modellerini oluşturmak için Quinlan tarafından geliştirilmiştir (Quinlan, 1983, Quinlan, 1993). C4.5, ID3'ün geliştirilmiş bir halidir. C4.5 eksik ve sürekli özellik değerlerini ele alabilmekte, karar ağacının budanması ve kural çıkarımı gibi işlemleri yapabilmektedir. Karar ağacının kurulması için girdi olarak kullanılacak bir dizi kayıt verilirse bu kayıtlardan her biri aynı yapıda olan birtakım özellik çiftlerinden oluşur. Bu özelliklerden biri kaydın hedefini belirtir. Problem, hedef-olmayan özellikler kullanılarak hedef özellik değerini doğru kestiren bir karar ağacı belirlemektir. Hedef özellik çoğunlukla sadece (evet, hayır), veya (başarılı, başarısız) gibi ikili değerler alır (Aydoğan, 2003).

Örnek: Tenis oynamak isteyen kişiler için hava koşulları kayıtlarıyla ilgilenilsin. Hedef özellik olan tenis oynama durumu iki değer alabilsin. Bunlar oynanabilir ve oynanamaz olsun. Çizelge 2.1'deki veriler bu örnek için verilmiş özellik ve bunların alabileceği olası değerler olsun.

Çizelge 2.1. Tenis örneği için girdi özellikler

ÖZELLİK	OLASI DEĞERLER
Hava durumu	Güneşli, Bulutlu, Yağmurlu
Nem	Sürekli
Sıcaklık	Sürekli
Rüzgar	Var, Yok

Şekil 2.2'de ise tenis örneği için oluşan karar ağacı gösterilmektedir.



Şekil 2.2. Tenis örneği için oluşan karar ağacı

Tenis örneğinde, sıcaklık ve nem özellikleri süreklidir. ID3 bu türden verilerle çalışamaz, bu nedenle ID3 algoritması geliştirilmiş ve C4.5 algoritması ortaya çıkartılmıştır (Quinlan, 1993). Bir karar ağacı ne bildiğimizi özetlediği için değil, yeni durumların sınıflandırılmasını doğru yaptığı için önemlidir (Aydoğan, 2003).

2.1.1.4. C4.5 algoritması

ID3'ün gelişmiş bir şekli olan C4.5 algoritmasının en büyük özelliği karar ağacının gereksiz dallarını tespit ederek budamasıdır. Bu algoritmanın bir diğer özelliği ise eksik veri ve sayısal değerlerle de çalışabilmesidir (Akgöbek ve Öztemel, 2006).

Bir karar ağacı kurulurken, kazanç hesaplamasıyla eğitim kümesindeki değerler bilinmeden sadece özellikler bilinerek işlem yapılır. Karar ağacı kullanımında bilinmeyen özellik değerlerine sahip olan kayıtlar, mümkün olan sonuçların olasılıklarını tahmin ederek sınıflanabilir (Aydoğan, 2003).

C4.5 algoritması ID3 algoritmasına çeşitli yenilikler getirmiştir. Bunlardan bazıları şunlardır:

1- Sürekli ve ayrık özellikleri ele alması: Sürekli değerli özellikleri ele almak için, C4.5 bir eşik değeri oluşturur ve özellik değeri eşik değerin altındaki örnekleri bir grup, eşik değere eşit ve büyük olan örnekleri de diğer bir grup olacak şekilde listeyi bölümlendirir.

2- Eksik özellik değerine sahip eğitim verilerini ele alması: C4.5 eksik özellik değerlerinin "?" şeklinde ele alınmasına izin verir ve bu eksik değerler kazanç ve entropi hesaplamalarında işleme dahil edilmez.

3- Farklı değerlerdeki özellikleri ele alabilmesi.

4- Oluşan ağaç modelinin budanması: Ağaç yapısı oluşturulduktan sonra C4.5 algoritması geriye doğru hareket ederek, faydalı olmayan kısımların yaprak düğümlerle yer değiştirilmesi şeklinde gereksiz dalları budar.

2.1.2. Bayes sınıflandırma

Bayes sınıflayıcıları istatistiksel sınıflayıcılardır. Verilen örneğin belirli bir sınıfa ait olma olasılığı şeklindeki, sınıf üyeliği olasılığını öngörebilmektedirler. Bayes teoremi üzerine inşa edilmiştir. Bunlar geniş veritabanlarına uygulandıklarında yüksek doğruluk ve hız elde etmektedirler (Han ve Kamber, 2001).

Bayes sınıflandırıcılar verinin tek bir kez taranmasını gerektiren basit sınıflandırıcılardır. Bu nedenle büyük veri yığınlarında yüksek doğruluk ve hız elde edebilirler. Performansları karar ağaçları ve sinir ağları ile rekabet edebilecek ölçüdedir (Mitra ve Acharya, 2003)

2.1.3. Geri Yayılım ile Sınıflandırma

Kabaca, sinir ağları, her birinin ağırlık değerleri olan ve birbiriyle bağlantılı giriş/çıkış birimleri olarak ifade edilebilir. Eğitim süresince ağ, giriş örneklerinin sınıf etiketlerini doğru bir şekilde tahmin edebilecek ağırlıkların belirlenmesi yolu ile eğitilmiş olur.

Sinir ağları uzun eğitim zamanı gerektirmektedir ve sinir ağlarını belirli uygulamalar için kullanmak daha uygundur. Deneysel yolla belirlenen ağ topolojisi ve yapı gibi belirli sayıda parametre gerektirmektedir. Sinir ağları, insanların eğitilmiş ağırlıkların arkasında yatan sembolik anlamları yorumlayamamaları şeklinde ortaya çıkan, zayıf yorumlanabilirlik özelliği nedeniyle eleştirilmektedir. Bu özellik nedeniyle önceleri sinir ağları veri madenciliğinde yaygın olarak kullanılamıyordu. Sinir ağlarının avantajları olarak ise, gürültülü veriye karşı yüksek tolerans gücüne sahip olması ve eğitmede kullanılmayan örüntüleri sınıflandırmadaki yüksek kabiliyeti sayılabilir. Ayrıca, YSA'nın dezavantajlarının üstesinden gelmekle eğitilmiş sinir ağlarından kurallar çıkarmak için çeşitli algoritmalar geliştirilmiştir. Bu etkenler veri madenciliğinde sınıflandırma için sinir ağlarının kullanımına katkıda bulunmaktadır (Han ve Kamber, 2001).

Geri yayılım algoritması sınıflandırma için kullanılan bir YSA algoritmalarından en bilinenidir. Veri örneğinin gerçek sınıf etiketi ile ağın sınıf tahmini arasındaki ortalama karesel hata (OKH)'yı minimum edecek veriyi modelleyebilecek uygun ağırlık kümesini aramaktadır. Eğitilmiş ağın yorumlanabilirliğini sağlamaya yardım etmek için, eğitilmiş sinir ağlarından kurallar çıkarılabilir.

2.1.4. Diğer Sınıflandırma Metotları

K-en yakın komşu sınıflayıcısı, durum temelli nedenleme, genetik algoritma, kaba küme yaklaşımı, bulanık küme yaklaşımı diğer sınıflandırma metotları olarak sayılabilir. En yakın komşu ile durum temelli nedenleme sınıflayıcıları, bütün eğitim örneklerinin örüntü uzayında saklandığı, örnek temelli sınıflandırma metotlardır. Bu nedenle her ikisi de etkin bir indeksleme tekniği gerektirmektedir. Genetik algoritmada, popülasyondaki bütün kuralların belirlenmiş olan eşik değerini bulana kadar, popülasyon, çaprazlama ve mutasyon işlemleri vasıtası ile evrim geçirmeye devam eder. Kaba küme yaklaşımı,

uygun özellikler üzerine tesis edilmiş ayırt edilemeyen sınıfları yaklaşık olarak belirlemek için kullanılabilir (Han ve Kamber, 2001).

2.2. Bölüm Değerlendirmesi

Bu bölümde sınıflandırma konusu üzerinde durulmuştur. Sınıflandırma algoritmaları, verilen eğitim kümesinden bu dağılım şeklini öğrenirler ve daha sonra sınıfının belirli olmadığı test verileri geldiğinde doğru şekilde sınıflandırmaya çalışırlar. Çok bilinen sınıflandırma algoritmaları incelenerek özetlenmiş ve sınıflandırma çalışmalarında bu algoritmalarının nasıl kullanılabileceği ortaya konmuştur.

3. KURAL ÇIKARIMI

Bilgisayar teknolojileri ve veri tabanı yazılımlarındaki gelişmeler sonucunda büyük miktarda veri birikmiş ve eldeki verilerden anlamlı bilgi çıkarma ihtiyacı ortaya çıkmıştır. Büyük miktarda depolanan bu veriler birçok gizli örüntü içermesine rağmen, toplanan veri miktarı büyüdükçe ve verilerdeki karmaşıklık arttıkça, geleneksel yöntemler ile veri bilgiye dönüştürülemez hale gelmiştir. Bu nedenle günümüzde büyük miktarda verileri çözümlemek amacıyla veri madenciliği yöntemleri yaygın bir şekilde kullanılmaktadır.

Kural çıkarımı doğrulama amacıyla ağın öğrendiği kuralları denetlemek, veri de mevcut olabilecek yeni ilişkileri keşfetmek ve çözümlerin genellenmesini iyileştirmekte kullanılabilir (Andrews ve ark., 1995).

3.1. Kural Çıkarımının Önemi

Kuttiyıl (2004), kural çıkarımının önemi şu şekilde sıralamaktadır:

1. Kullanıcıya yorumlama kapasitesi sağlaması,
2. Veri madenciliği ve bilgi keşfi sağlaması,
3. Uzman sistemler için bilgi kazancı sağlaması.

3.2. Kural Çıkarımının Avantajları

Kuttiyıl (2004), kural çıkarımı yönteminin sağladığı avantajlar şu şekilde belirtmişdir:

- Kurallar şeklinde ağın giriş çıkış haritalamasını yorumlayabilen bir mekanizma sağlaması,
- Orijinal eğitimdeki eksiklikleri tanımlama kabiliyeti sağlaması,
- Çıkarıldıklarında ağ performansını arttıracak gereksiz ağ parametrelerini belirlemesi,
- Verideki önceden bilinmeyen ilişkilerin analizini yapması,
- Muhakeme ve açıklama kapasitesi sağlaması,

- Çapraz ilgi desteği ve doğrulama kapasitesi sağlaması,
- Bilgi kazancı problemini azaltma ve giriş bilgilerini düzeltme.

3.3. Kuralların Gösterimi

Kurallar verilen bir durum için yapılması gereken işlerin kümesi olarak tanımlanan, keşif veya buluşları temsil etmek için kullanılır. Kuralın "Eğer" kısmı kuralın uygulanabilir olmasını sağlayan olayı yâda veriyi tanımlayan deyim dizileridir. Kuralın "O Halde" kısmı ise kural uygulanabilir olduğunda yürütülecek olan işlemleri tanımlanan kısımdır.

Genel olarak kullanılan kural yapısı örnek (1)'de örnek biçiminde sunulmuştur.

Eğer A Doğru ve B Doğru ve C Yanlış ise O halde X Doğrudur (1)

Kuralları ifade eden diğer bir yöntem olan 'M of N' metodu ise yalnızca bazı ağ yapılarından çıkarılan kuralları ifade etmek için kullanılan alternatif bir gösterimdir.

Bu metoda ait örnek örnek (2)'de verilmiştir (Kuttiyil, 2004).

Eğer N adet durumun M tanesi Doğru ise O halde X Doğrudur (2)

3.4. Çıkarılan Kuralların Özellikleri

Ağdan çıkarılan kuralların sahip olmaları gereken özellikler şu şekilde sıralanabilir (Kuttiyil, 2004):

Geçerlilik: Kuralda yer almayan değişkenlerin değerlerinin ne olduğuna bakılmaksızın kurallar tutulmalıdır.

Örneğin: Kural 1: Eğer a_1 Doğru ve a_3 Yanlış ise O halde a_6 Doğrudur

Eğer bu kural geçerli ise, diğer değerlere bakılmaksızın bu kural tutulmalıdır (a_2 a_4 a_5).

Maksimum Genellik: Durumlardan herhangi birisi silinirse, kuralların geçerliliği ortadan kalkacaktır.

Örneğin: Kural 2: Eğer a_1 Doğru ve a_3 Yanlış ve a_5 Yanlış ise O halde a_6 Doğrudur

Eğer ‘a_5 Yanlış’ durumu silinirse, bu örnekteki Kural 2’nin geçerliliği ortadan kalkacak, fakat önceki örnekteki Kural 1 hala geçerli kalacaktır. Buradan, Kural 1 maksimum genellikte iken Kural 2 değildir, sonucuna ulaşılmaktadır.

Oluşturulan kuralların eksiksiz olması gerekmektedir (Kuttiyil, 2004).

3.5. Kural Çıkarımındaki Önemli Noktalar

Kural çıkarımı ile ilgili önemli noktalar şu şekilde sıralanabilir (Kuttiyil, 2004):

Taneciklilik: Sistemin kendi çıkış kararlarının her biri ile sağlandığının kanıtı ve detaylı hipotezin seviyesidir.

Saydamlık: Karar veya yargıların ne derece iyi açıklanabildiğidir.

Genelleme Yeteneği: Kuralların mümkün olduğunca çok durumu sağlamasıdır.

Değiştirilebilirlik: Uygun eğitilmiş ağ yapısı güncellendiğinde ya da farklı veri kümesiyle yeniden eğitildiğinde, çıkarılmış kuralların güncellenebilme yeteneğidir.

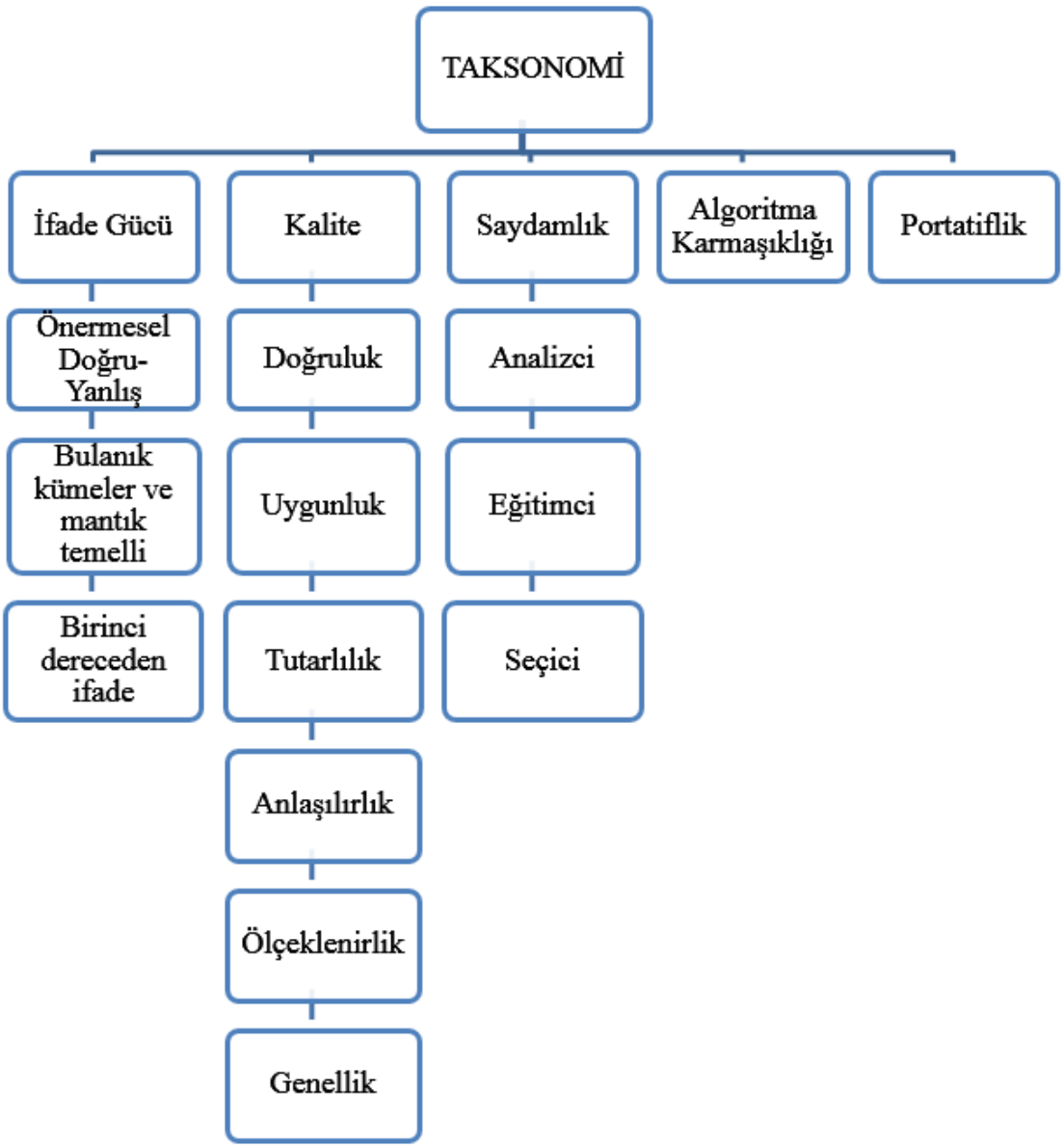
Sağlamlık: Eğitim verisindeki ya da giriş alanındaki bozukluklardan yöntemin etkilenmeme ölçüsüdür.

3.6. Taksonomi

Andrews ve arkadaşları (1995), eğitilmiş sinir ağlarından kurallar çıkarma tekniklerini kategorize etmek için geniş kapsamlı bir taksonomi geliştirmişlerdir. Bu taxonomi aşağıdaki 5 kıstası içermektedir (Kuttiyil, 2004).

- İfade Gücü
- Kalite
- Saydamlık
- Algoritma karmaşıklığı
- Portatiflik

Kural Çıkarımı Metotlarının Sınıflandırma Taksonomisi Şekil 3.1'de verilmiştir (Andrews ve ark., 1995).



Şekil 3.1. Kural çıkarımı metotlarının sınıflandırma taksonomisi

İfade gücü : Çıkarılan kurallar çeşitli şekillerde ifade edilebilmektedir (Kuttiyil, 2004).

- Geleneksel sembolik kurallara (boole mantığı - doğru yanlış değeri alan ve önerme şeklinde) Eğer ... O Halde ve M of N örnek olarak verilebilir. Önermeler mantığı, mantıksal ifadelerin doğruluğunu anlamaya yarayan matematiksel bir modeldir. Sadece tamamlanmış deyimlerin doğruluk değerleriyle uğraşma düşüncesi ve nesneler arasındaki genel ilişki ve bağımlılıkları dikkate almayarak sınırlandırılmıştır.

- Geleneksel olmayan (bulanık küme ve mantığını temel alan kurallar)
- Önerme mantığı, ölçülen önermeler için önerme, ölçüm (isteğe bağlı alanlar üzerinde bulunan değişkenler) ve sonuç çıkarımı kuralları ortaya koyarak önerme kuralları sunmaktadır.

Kalite : Kuttiyil'e (2004) göre kalite ölçütleri şu şekilde sıralanabilir:

Kural Doğruluğu: Kuralların eğitim sırasında görmediğimiz yeni örnekleri doğru bir şekilde sınıflandırmasıdır.

Kural Uygunluğu: Çıkarılan kuralların tam olarak bilgiyi sunma derecesidir.

Kural Tutarlılığı: YSA'nın farklı eğitim dönemlerinde görülmemiş örnekleri aynı şekilde sınıflandıran kural kümeleri üretme derecesidir.

Kural Anlaşılırlığı: Eğitilmiş ağdan çıkarılan her bir kuraldaki terim sayısı ve kural sayısı ile belirtilir.

Kural Ölçeklenebilirliği: Kural çıkarma algoritmasının çalışma süresinin ne kadar olduğu ve ağ, özellik kümesi ve eğitim kümesi boyutu gibi faktörlerin bir fonksiyonuyla değişen çıkarılmış modellerin anlaşılabilirliği ile ölçülür.

Kural Genelliği: Yöntemin ağ yapısı için özel olarak geliştirilmiş eğitim yöntemleri veya sınırlandırmalar gerektirme derecesidir.

3.7. Bölüm Değerlendirmesi

Literatürde birçok kural çıkarım çalışması vardır. Bu kural çıkarım çalışmalarının çoğunluğu YSA veya SVM tabanlı çalışmalardır. Bu çalışmalarda önce bir yöntem kullanılarak sınıflandırma yapılmış, sonra oluşan model optimize edilerek kurallar üretilmiştir.

Bu bölümde kural çıkarımı hakkında bilgiler verilmiştir. Kural çıkarımının önemi, kural çıkarımının avantajları, kuralların gösterimi, çıkarılan kuralların özellikleri ve kural çıkarımındaki önemli noktalardan bahsedilmiştir.

4. BAĞIŞIKLIK VE YAPAY BAĞIŞIKLIK SİSTEMLERİ

4.1. Bağışıklık Sistemi

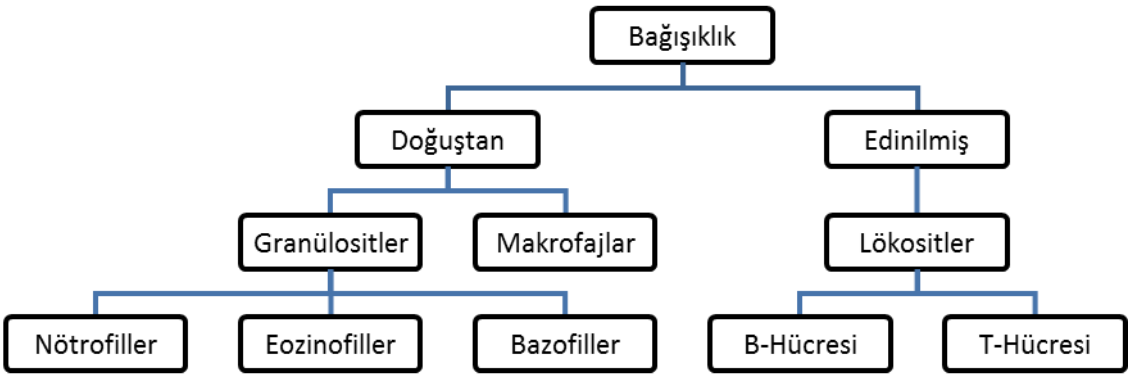
İnsanoğlu çok eskilerden beri gerek doğada gerçekleşen olayları anlamak, gerekse bu işleyişleri taklit ederek yaşamını kolaylaştırmak üzere çoğu problemin çözümünü hep doğada aramıştır. Teknolojinin ilerlemesi, beraberinde çözümü karmaşık problemleri getirmiştir ve bilinen yöntemler, karmaşıklaşan bu problemlerin çözümünde yetersiz kalmaya başlayınca, farklı çözüm yolları araştırılmaya başlamıştır.

İnsan bağışıklık sistemi virüs, bakteri gibi neredeyse sınırsız sayıda olan yabancı saldırganlara karşı doğal, hızlı ve etkili cevap verme yeteneğine sahip karmaşık bir ağ yapısı olarak tanımlanabilir. Bağışıklık sistemi örüntü tanıma, öğrenme, hafıza edinme, farklılıkların genelleştirilmesi, gürültü toleransı, genelleme ve optimizasyon gibi pek çok işlemi yerine getirme kabiliyetine sahip hücrelerden meydana gelir. Bağışıklık sistemini daha iyi anlamak ve mühendislik problemlerinin çözümünde kullanabilmek için bu biyolojik prensiplere dayanan yeni hesaplama teknikleri geliştirilmektedir. Bu amaçla insan bağışıklık sisteminden esinlenerek yapay bağışıklık sistemleri tanımlanmıştır (Kahramanlı, 2008).

Bağışıklık ağı teorisi ilk olarak 1974 yılında Jerne tarafından ortaya atılmıştır (Jerne, 1973). Bu teoride bağışıklık sisteminin, antijenlerin yokluğunda bile birbirleri ile etkileşen hücre ve moleküllerden oluştuğu varsayımı yapılmıştır. Jerne'den sonra bu alanda yapılan çalışmalar başlamış fakat 1990'lı yıllara kadar oldukça yavaş bir ilerleme kaydedilmiştir. 1990'lı yıllar, yapay bağışıklık sistemlerinin gelişmeye başladığı yıllar olarak tanımlanabilir. Yapay Bağışıklık Sistemleri (YBS) alanında yapılan çalışmalar şimdiye değin birkaç çalışma dışında hep sınırlı kalmıştır ve çok fazla ilgi görmemiştir. Çünkü yapılan çalışmaların büyük çoğunluğu genel bir çözüm üretmekten çok probleme bağımlı algoritmalar ile problemi ortadan kaldırmaya yönelik olmuştur. Fakat son zamanlarda oluşturulan sistemlerin birçoğu, ortak öznitelikteki problemlere genel çözümler oluşturma yolunda işlerlik kazanmaktadır ve YBS'ne olan ilgi de buna paralel olarak ivme kazanmaktadır (Özşen, 2008).

İnsan bağışıklık sistemi, dışarıdan gelebilecek saldırılara karşı vücudu koruma yeteneğine sahip kuvvetli bir savunma mekanizmasıdır. Bu mekanizmanın esas öğeleri beyaz kan hücreleri yani akyuvarlardır. Bağışıklık sistemi doğal ve edinilmiş olmak üzere

ikiye ayrılır. Şekil 4.1'de görüldüğü gibi doğal bağışıklık sisteminin elemanları granükloitler ve mikrofajlar iken adaptif bağışıklık sisteminin elemanları lenfositlerdir.



Şekil 4.1. Bağışıklık sisteminin sınıflandırılması ve temel elemanları

Bağışıklık bilimi genel anlamda kendi kalıtsal yapısına yabancılık özelliği taşıyan özel maddeleri ayırt edebilme yeteneğindeki canlıların, bu özel maddelere karşı gösterdiği reaksiyonların tümünü inceleyen bir bilim dalıdır.

En basit yapılı canlılarda bile kendinden olanı tanıma, yabancıyı ayırt edebilme özelliği mevcuttur. Evrimsel açıdan en gelişmiş canlı olan insan kendi yapısına yabancı olan maddeleri tanıyabilme ve onlarla baş edebilme özelliklerine sahiptir. Bu sayede mikroplar başta olmak üzere yabancı ve zararlı olabilecek maddelere karşı kendisini savunur. Genelde antijen olarak tanımladığımız bu maddelerin organizmaya girmesi ile başlayan ve birbiri ile ilişkili birçok biyolojik reaksiyonun meydana geldiği bağışık yanıt olayında, birçok sistem, organ ve hücre görev alır. Bağışıklık olaylarında rolü olan organ ve hücrelere bağışıklık sistemi adı verilir (Kahramanlı, 2008).

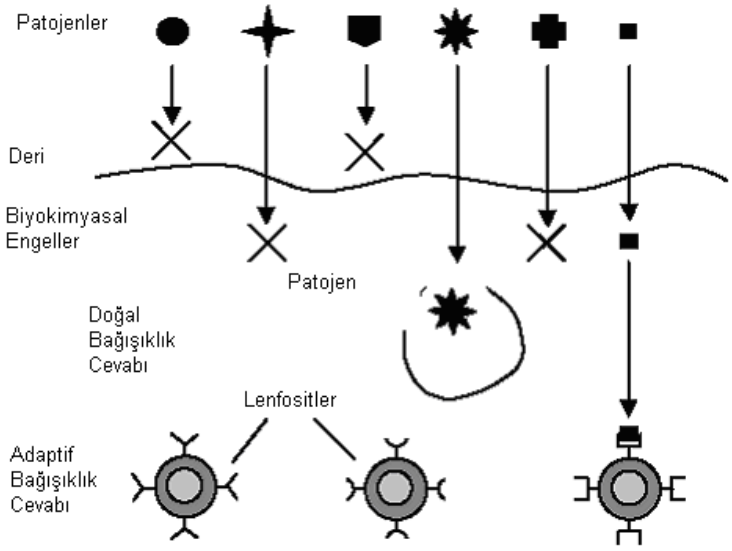
Bağışıklık bilimi ile ilgili ilk bilgilerimiz çok eskilere dayanmaktadır. Eski çağlardan beri kişilerin farklı dirençte oldukları fark edilerek, salgınlarda bazı kişilerin ölmesine karşın, bazılarının hiç hastalanmadığı veya hastalığı çok hafif geçirdiği, kızamık, çiçek gibi hastalıkların ikinci kez geçirilmediği (ömür boyu korunma) bu konudaki ilk gözlemler olmaktadır (De Castro, ve Timmis, 2002).

Bağışıklık bilimi tarihi ile ilgili olayların başında Türklerin de yer aldığını görülmektedir. 1721 yılında İstanbul'da İngiliz Büyükelçisi'nin eşi Leydi Montagu, İngiltere'ye yazdığı mektuplarda, Türkiye'de hafif çiçekli insandan sağlıklı insana aktarılarak uygulanan çiçek aşısından söz etmektedir. Ayrıca İstanbul'da çalışan İtalyan hekim E. Timoni de 1713'te bir toplantıda Türkiye'de 40 yıldan beri bu aşının uygulandığını belirtmektedir.

Bundan sonraki olay yine çiçek aşısı ile ilgilidir. İnek sağan kadınların ellerinde çıkan inek çiçeği yaraları sayesinde asıl çiçek hastalığından korunduklarını fark eden İngiliz hekim Edward Jenner, 1798'de bu olaydan yararlanarak çiçek aşısı uygulamasını başlatmıştır. İlk kez aşılama terimini kullanmıştır. Daha sonra yüzyıllık bir sessiz dönem geçmiş ve 1880'li yıllarda Fransız hekim Louis Pasteur hayvanlar için şarbon ve kolera, insan ve hayvanlar için ise kuduz aşısını geliştirerek bu konuda önemli hizmetler vermiştir. Bunu izleyen yıllarda da pek çok enfeksiyon hastalığı için başarılı aşı çalışmaları yapılmıştır (Akşit ve ark., 1996).

4.1.1. İnsan bağışıklık sisteminin yapısı

Bağışıklık sisteminin düzenleme ve koruma mekanizmalarının çeşitli seviyelerde yer aldığı çok katmanlı bir mimarisi olduğu söylenebilir. Bu katmanlar Şekil 4.2'de gösterilmiştir (De Castro, ve Timmis, 2002). Sırası ile bu katmanların açıklamaları aşağıdaki gibidir:



Şekil 4.2. Bağışıklık sisteminin çok katmanlı mimarisi

4.1.1.1. Fiziksel engeller

Derimiz zararlı olsun ya da olmasın yabancı maddelere karşı vücudumuzu korumak için kalkan gibi çalışır. Solunum sistemimiz de dış yüzey ile temasta olan organların

mukoza kaplı kısımları sayesinde dışardan gelecek muhtemel tehlikelere karşı vücut savunmasına yardımcı olur.

4.1.1.2. Biyokimyasal engeller

Vücudun salgıladığı tükürük, ter, gözyaşı gibi sıvılar yıkıcı enzimler içerirler. Mide asitleri ise yeme ve içme ile vücuda giren mikroorganizmaları öldürürler. Vücudun sıcaklığı ve pH değeri bazı saldırganlar için elverişsiz hayat şartları sunar.

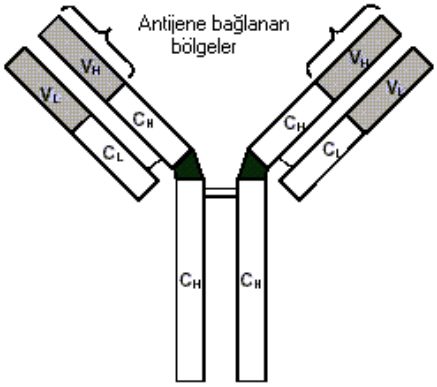
4.1.1.3. Doğal ve edinilmiş bağışıklık sistemi

Bağışıklık sistemini oluşturan iki önemli kavramdır. Doğal bağışıklık sistemi her insanın doğduğunda sahip olduğu tanımlama ve savunma bilgisidir. Bu bilgiler uzun bir evrimleşmenin sonucudurlar ve kalıtsal olarak yeni doğana aktarılırlar. Bu sayede bağışıklık sistemimiz ilk karşılaşmada bile pek çok patojeni yok edebilecek güçtedir. Doğal bağışıklık sisteminin önemli parçalarından biri, özel bir protein sınıfı olan "tamamlayıcı sistem"dir. Tamamlayıcı sisteme ait moleküller, antikor moleküllerinin (patojen türü tanımlama ve işaretlemede kullanılırlar) görevlerine yardımcı olurlar. Doğal bağışıklık sistemi, *B*-lenfositlerin olgunlaştığı lenf düğümü ve dalakta bulunan özel bölgeler olan jerminal merkezlerde kodlanan bir grup örüntü tanıma almaçı üzerine kuruludur. Bu almaçlar, mikrobik patojenlerin, üzerlerinde bulundurdukları özel moleküler örüntülerin tanımlanmasını sağlarlar. Bu örüntüler sadece patojenlere özeldir ve asla taşıyıcı beden dahilinde üretilmezler. Bu patojenlerin örüntü tanımlayıcı almaçlarca belirlenmesi bir dizi zincirleme kimyasal tepkimeye sebep olur. Bu tepkimeler patojenin varlığının tespit edildiğine dair sinyalleri taşır. Bağışıklık sisteminin tanımlamaya ait parçalarının kesinlikle bedendeki diğer hücrelerden ve moleküllerden ayrık bir yapıda olması gerekir. Aksi takdirde bedenin kendine ait hücreleriyle, patojenleri birbirinden ayırt edemez. Bu yapının sonucu olarak, doğal bağışıklık, kedine has olan ile olmayanı ayırt edebilir. Başka bir deyişle savunduğu bedene ait olanla olmayanı ayırt edebilir. Doğuştan bağışıklığın en önemli işlevlerinden biri de yardımcı-uyarı işaretleri üretmesidir. Bu işaretler, edinilmiş bağışıklık sistemine ait *T*-lenfositlerin etkinleşmesinde önemli rol oynar. Edinilmiş bağışıklık, beden dahilinde, ikizlemeyle özel olarak iki tür hücre üzerine dağıtılmış antijen almaçlarını kullanırlar. Söz edilen hücreler, *B*-lenfosit ve *T*-lenfosit hücreleridir. Bu hücreler üzerinde bulunan antijen almaçları

rastgele süreçlerle üretilirler ve bunun sonucu olarak, “edinilmiş bağışıklık” tepkisinin tasarımı, seçimsel ikizleme yoluyla çoğaltılmış, küçük farklılıklarla kendilerine özgü özellikler gösteren almaçlara sahip lenfositlere dayanır. Bu bağışıklık mekanizması sayesinde beden daha önce hiç karşılaşmamış olduğu patojenlere karşı da kendini savunabilme yeteneğini kazanır (De Castro, ve Timmis, 2002).

4.1.2. Bağışıklık hücreleri

Bağışıklık sistemi, kemik iliğinde üretilen ve çoğunlukla orada olgunlaşan, çok çeşitli hücrelerin bileşiminden oluşur. Kemik iliğinden yola çıkıp, kan dolaşımına, lenflere ve dokulara dağılırlar. Bir kısmı genel savunma amaçlıyken, diğer kısmı çok özel bir patojenle savaşmak üzere özelleşmektedir. Etkin bir savunma için bu hücrelerin sürekli birlikte çalışmaları gerekmektedir. Şekil 4.3’de, bağışıklık sistemini oluşturan hücreler ve salgılar yapısal olarak gruplanmaktadır (De Castro, ve Timmis, 2002).



Şekil 4.3. Antikor’un yapısı

Organizmaya girdiğinde, kendisine karşı bir bağışık yanıt oluşmasına yol açan ve bu cevap sonucunda oluşan ürün ile özgül olarak birleşebilen maddelere antijen denir. Antijenler "Ag" şeklinde gösterilirler. İmmünojen sözcüğü de antijen ile eş anlamlı olarak kullanılır. Oluşan ürüne ise antikor denir. Bağışıklık sisteminin temel hücresi lenfositlerdir. Lenfositlerde kalıtsal olarak bağışık yanıt ile ilgili ürünleri sentezleme yeteneği vardır. Nasıl çeşitli hormon ve enzimleri yapan ve salgılayan özel hücreler varsa, lenfositler de antijenle uyarıldıklarında immünglobülin (antikor) ve diğer bağışık yanıt maddelerini sentezlerler. Immünglobülinler kısaltma olarak "Ig" şeklinde gösterilirler (De Castro, ve Timmis, 2002).

Lenfositler bağışıklık sisteminin çok önemli hücreleridir. Kemik iliğindeki kök hücrelerde gelişirler ve merkezi lenfoid organlarda (Timus veya kemik iliği) olgunlaşırlar. Kan yoluyla periferik lenfoid doku ve organlara giderek özel bölgelerine yerleşirler. Erişkin bir insanda yaklaşık 10^{12} (~ 1 trilyon) lenfosit bulunur ve her gün 10^{9} (~ 1 milyar) lenfosit yapılarak kana verilir. Kandaki lökositlerin %20-30'unu lenfositler oluşturur. Lenfositler 8-12 mikron çapında, iri çekirdekli, dar sitoplazmalı hücrelerdir.
Görünümleri birbirine çok benzeyen ve ışık mikroskobunda ayırt edilemeyen farklı iki lenfosit tipi vardır. *B* ve *T* lenfositleri olarak tanımlanan bu hücrelerin olgunlaşmaları, işlevleri ve antijen yapıları birbirinden farklıdır. Ancak aralarında yakın bir işbirliği bulunmaktadır (De Castro, ve Timmis, 2002).

B-lenfositler antikora dayalı (hümoral) bağışıklıktan sorumlu hücrelerdir. Kandaki lenfositlerin %25'i, dalaktakilerin %50'si *B*-lenfosittir. *B*-lenfositler sentezledikleri immünglobulin moleküllerini hücre yüzeylerinde zarda taşırlar.
Bir *B*-lenfosit sadece tek bir çeşit antijene bağlanabilen yüzey immünglobulin reseptör taşır. Bu nedenle bağışıklık sisteminde, zaman içinde karşılaşma ihtimali olan on binlerce çeşit antijene karşı özgül reseptör taşıyan on binlerce *B*-lenfosit çeşidi hazır durumda bulunmaktadır (De Castro, ve Timmis, 2002).

Organizmaya antijen girdiğinde, yüzeyinde bu antijene özgül reseptör taşıyan B-lenfositleri bulur ve uyarır. Uyarılan *B*-lenfositler başkalaşıma uğrar ve plazma hücresine dönüşürler. Plazma hücresi de uyaran antijene özgül olan çok miktarda antikor (immunglobulin) sentezler. Plazma hücresinin çoğalma yeteneği yoktur ve ömrü kısadır (~ 2-3 gün). Ancak bir dakikada yaklaşık 20 bin antikor molekülü sentezleyebildiği gösterilmiştir.Uyarılan *B*-lenfositlerinden bir kısmı ise bellek hücresi haline gelir. Bellek *B*-lenfositleri uzun ömürlüdür (bazen bir ömür boyu) ve aynı antijenle tekrar karşılaştıklarında hızla çoğalarak daha hızlı ve güçlü antikor yanıtı oluştururlar. *T*-lenfositler hücresel tipte bağışık yanıttan sorumludur. Kemik iliğinde yapılan *T*-öncü hücreler timusta olgun *T*-lenfosit haline gelirler. Bu olgunlaşma sırasında *T*-lenfosit yüzeyinde pek çok reseptör yerleşir. *T*-hücre yüzeyinde yüzey immünglobulin bulunmaz. Bunun yerine antijenleri özgül olarak tanıyan "*T*-hücre reseptörü = TCR (T Cell Receptor)" bulunur. Bir *T*-lenfositi sadece tek bir çeşit antijen için TCR taşır ve *B*-lenfositlerinde olduğu gibi bağışıklık sisteminde zaman içinde karşılaşma ihtimali olan on binlerce çeşit antijene yanıt verebilecek on binlerce çeşit *T*-lenfositi bulunur (De Castro, ve Timmis, 2002).

Organizmaya antijen girdiğinde yüzeyinde bu antijene özgül reseptör taşıyan *T*-lenfositleri bulur ve uyarır. Uyarılan *T*-lenfositler başkalaşıma uğrar ve sonuçta o antijene duyarlı *T*-lenfositler oluşur. *T*-lenfositler bağışıklık sisteminin en önemli hücreleridir ve doğrudan antikora bağımlı olmayan ve hücrelerin yönettiği ve katıldığı özgül bağışıklığı oluştururlar. Bağışık yanıttaki rolleri açısından *T*-hücre topluluğunun homojen (tek tip) olmadığı ve yapı ve işlev özelliği farklı olan alt grupların bulunduğu bilinmektedir. *T* lenfositleri uzun ömürlüdür. Bellek *T* lenfositleri ise antijeni tanımış olan çok uzun ömürlü ve antijenle tekrar karşılaştığında çok hızlı ve güçlü bağışık yanıt oluşturan hücrelerdir (De Castro, ve Timmis, 2002).

4.1.2.1. Yan Zincirler Teorisi

1897'de P.Ehrlich tarafından ileri sürülmüştür. Ancak geçerliliği kalmamış olup, sadece tarihi değeri vardır. Bu teoriye göre hücrelerde besin maddelerini yakalayabilmek için her besin maddesine uyan almaçlar (reseptörler) vardır. Antijenler, hücrelerdeki bu reseptörlere rastlantı olarak benzer yapıda ise onlarla birleşirler. Ancak bu birleşme reseptörün görevini bozar. Hücrenin beslenmesi bozulur ve hücre yaşamını sürdürebilmek için çok sayıda reseptör yapmaya başlar, bunların fazlası dolaşıma geçer. Bunlara da antikor molekülleri denir (De Castro, ve Timmis, 2002).

4.1.2.2. Kalıp Teorisi (Direktif Teori)

1930'da F. Haurowitz tarafından ileri sürülmüştür. Bu teoriye göre, organizmaya giren antijenler Ig sentezlemekte olan hücrelere göç ederler ve Ig molekülleri antijeni bir kalıp gibi kullanarak ona uyacak şekilde aminoasid dizilişlerini şekillendirirler. Antijene uygun Ig sentezinin sürekli olması için de antijenin kısımları hücrede kalır ve böylece özgül antikorlar sentezlenmiş olur. Ancak son yıllarda yapılan çok duyarlı deneylere karşın Ig sentezleyen hücrelerde antijene rastlanmamıştır. Ayrıca immünglobülinler protein yapısında olduklarına göre bir kez sentezlendikten sonra yapılarının değişmesinin mümkün olmadığı bugün bilinmektedir. Bu nedenlerle teori desteğini kaybetmiştir (De Castro, ve Timmis, 2002).

4.1.2.3. Klonal Teori

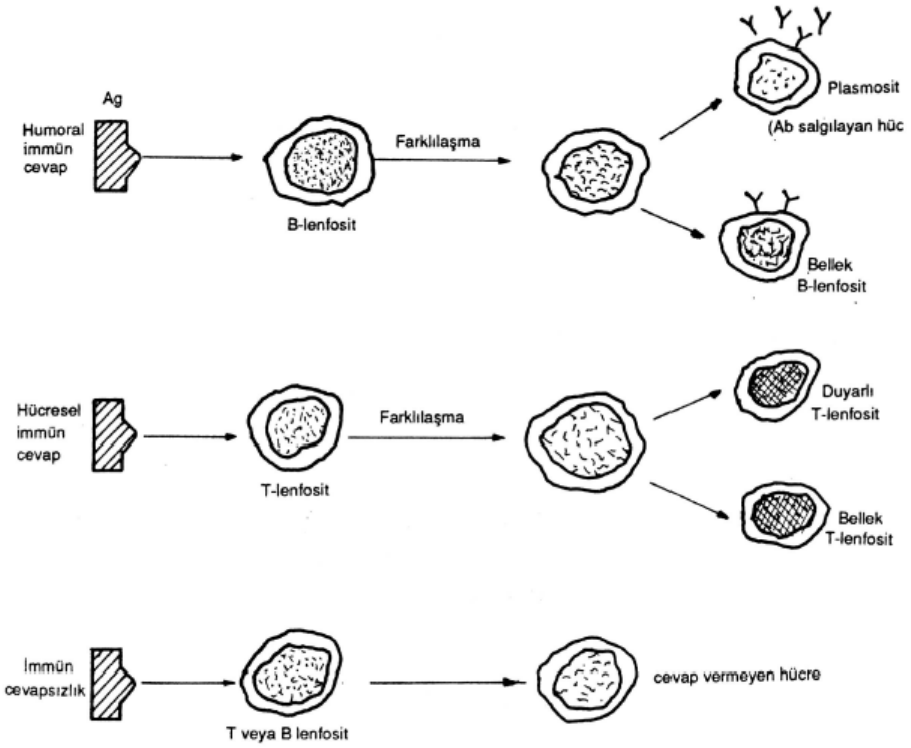
Klonal Teori (Selektif Teori, Doğal Klon Seçimi) 1955'te Jerne, 1957'de Burnet tarafından ileri sürülmüştür. Günümüzde de destek görmektedir. Bu teoriye göre bağışık yanıt verme yeteneğindeki hücrelerden biri veya bir grubu, tek bir antijen veya benzer yapıdaki bir grup antijen determinantı ile birleşebilme ve cevap verebilme yeteneğindedir. Böylece canlılarda, her biri bir antijene yanıt verebilecek yetenekte, birbirinden farklı, çok sayıda lenfosit bulunduğu kabul edilmektedir. Organizmaya giren antijen kendisine uyan hücreyi (*B*-lenfositi) bulur, bağlanır ve onu uyarır. Sadece onu seçmiş olur. Uyarım sonucu lenfositler hızla çoğalmaya başlar. Böylece belli bir antijene karşı özgül antikor sentezleyebilen bir hücre kolonisi meydana gelmiş olmaktadır.

4.1.3. Bağışıklık sisteminin vücudu koruma süreci

Bağışıklık sistemi vücudu dışardan gelen saldırılara karşı koruyan hücre ve moleküllerin karışık bir kümesinden oluşur. Vücudumuz bütün bağışıklık tepkilerinin nihai hedefi olan antijenin (Ag) tehdidine karşı edinilmiş bağışıklık sistemi tarafından korunmaktadır. Antijenik tehdide karşı organizmada üç çeşit bağışık yanıt oluşabilir. Bunlar;

- Hümoral (=Sıvısal) bağışık yanıt
- Hücresel bağışık yanıt
- Bağışık yanıtsızlık (=İmmünolojik tolerans)

Hümoral bağışık yanıttan *B* lenfositler sorumludur. Hümoral bağışık yanıt sonucu oluşan özgül antikorlar kan ve doku sıvılarında vücut savunmasında rol alırlar. Antikorlar, kan serumu ile kişiden kişiye aktarılabilirler. Bağışık yanıt sonuçları Şekil 4.4'te görülmektedir (Akşit ve ark., 1996).

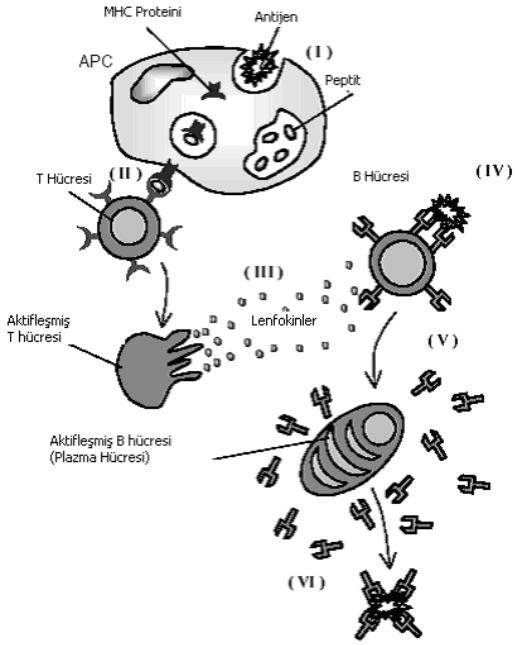


Şekil 4.4. Bağışık yanıt sonuçları

Hücresel bağışık yanıttan *T*-lenfositler sorumludur. Ancak organizmada, yabancı antijenlerin ortadan kaldırılması veya zararsız hale getirilmesi için *T*-lenfositler yanında fagositik hücreler (makrofaj ve granülositler), doğal öldürücü hücreler (NK hücreleri) ve diğer bazı hücrelerin işbirliği, hücresel bağışık yanıt içinde değerlendirilir. Ayrıca antikorlar da zaman zaman hücresel bağışıklığa katkıda bulunurlar. (Örn. opsonizasyon, antikora bağımlı hücresel sitotoksisite gibi) (Akşit ve ark., 1996).

İmmünolojik olgunluğa erişmiş canlıların normalde bağışık yanıt vermek durumunda oldukları, yabancı, belli bir Ag'e karşı bağışık yanıt oluşturmamalarına bağışık yanıtsızlık veya immünolojik hoşgörü (tolerans) denir (Akşit ve ark., 1996).

Şekil 4.5'de bu savunma mekanizmasının basitleştirilmiş gösterimi bulunmaktadır. Şekilde görülen bu adımların kısaca açıklaması aşağıda verildiği gibidir (De Castro, ve Timmis, 2002).



Şekil 4.5. Bağışıklık sisteminin basit aktivasyon mekanizması

Makrofajlar gibi antijen sunma yeteneğine sahip APC hücrelerinin yiyebilecekleri ve sindirebilecekleri antijenleri bulmak için vücutta dolaştığını ve bulduklarında ise onları antijenik peptitlerine ayırarak parçaladıklarını göstermektedir.

1. Bu peptitler MHC (major histocompatibility complex) moleküllerine bağlanır, APC hücre yüzeyi ile birleşerek MHC/Peptit kombinasyonunu oluştururlar.
2. *T*-hücreleri, farklı bir MHC/Peptit kombinasyonlarını tanımak için her biri aktif olan reseptör moleküllerine sahiptirler.
3. *T*-hücresi MHC/peptit kombinasyonu tanıdıktan sonra aktive olur ve parçalanarak lenfokinler ve ya kimyasal sinyaller salgılar. Bunlar diğer bağışıklık sistemi elemanlarını uyarır ve aktivasyona katılmasını sağlar.
4. *T*-hücrelerinin tersine *B*-lenfositleri MHC moleküllerinin yardımı olmaksızın serbest durumdaki antijenleri tanıma yeteneğine sahip reseptörlere sahiptir.
5. *B*-hücrelerinin reseptörleri tarafından uyarı sinyali alındığında *B*-hücreleri plazma hücrelerine dönüşerek çok sayıda antikor molekülü salgılarlar.
6. Serbest kalan bu antikorlar patojenleri etkisiz hale getirmede kullanılır.
7. Daha sonra bunlar enzimler veya temizleme hücreleri kullanılarak yok edilirler. Aktive olan bu *B* ve *T*-hücrelerinin bir kısmı değişerek hafıza

hücrelerini oluştururlar. Bu hücreler organizma içinde dolaşmaya devam edecek ve gelecekte olabilecek aynı veya benzer antijene daha iyi bir bağışıklık cevabı üretmeyi garantileyeceklerdir. *B*-lenfositlerde antikor üretimi için kullanılan genler sürekli mutasyona ve düzenlemeye uğradığından, tekrar eden saldırılar karşısında antikor cevabı eskisine göre daha etkin ve net bir hal almaya başlar. Bu olaya duyarlılık olgunlaşması denir (De Castro ve Timmis 2002).

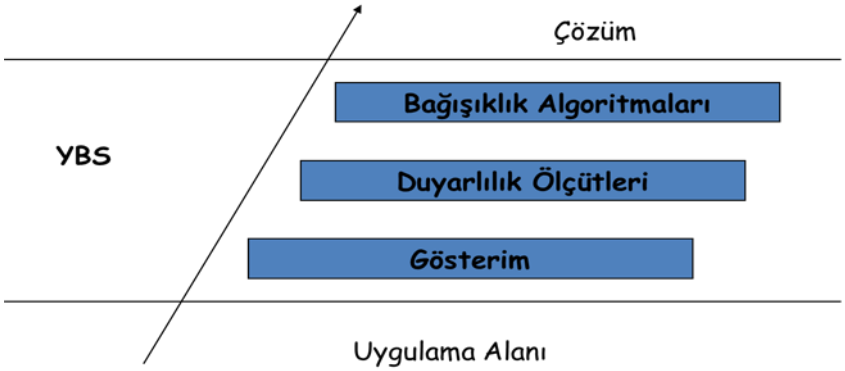
4.2. Yapay Bağışıklık Sistemleri

Son yıllarda biyolojiden esinlenerek geliştirilmiş sistemler hakkında yapılan çalışmalarda büyük bir miktarda artış gözlemlenmiştir. Bunların arasında yapay sinir ağları (YSA), evrimsel hesaplama, DNA hesaplama ve yapay bağışıklık sistemleri (YBS) sayılabilir. YBS, bağışıklık fonksiyonları, prensipleri ve modelleri gözlemlenerek, teorik bağışıklıktan esinlenilerek geliştirilen adaptif sistemlerdir (De Castro ve Timmis 2002). Bağışıklık sistemi uygulamalarının başlıca kullanım alanları örüntü tanıma, hata ve anormallik tespiti, veri analizleri (veri madenciliği, sınıflandırma vb.), etmen tabanlı sistemler, makine öğrenmesi, arama ve optimizasyon problemleri ve güvenlik ve bilgi sistemleridir.

Biyolojiden esinlenen diğer algoritmalarda olduğu gibi YBS'de de tasarım aşamasında aşağıdaki temel soruların cevaplarının verilmesi gerekmektedir.

- Sistemdeki elemanlar nasıl temsil edilecektir?
- Bireylerin çevre ve birbirleri ile olan ilişkilerini değerlendirmek için hangi değerlendirme mekanizmaları kullanılacaktır?
- Sistem dinamiğini yönetmek için adaptasyon prosedürlerinin belirlenmesi. Yani sistem davranışı zamanla neye göre değişecektir?

Yukarıdaki soruların cevaplanma süreçleri YBS oluşturulurken izlenmesi gereken katmanlı bir yapının aşamaları olarak da düşünülebilinir. Şekil 4.6'da YBS için katmanlı yapı yaklaşımı görülmektedir (De Castro, ve Timmis, 2002).



Şekil 4.6. YBS' nin katmanlı yapısı

Bu yapıya göre tasarım, bağışıklık hücre ve moleküllerin gösteriminde kullanılacak soyut modelin belirlenmesi ile başlar. 1979 yılında Perelson ve Oster'in sunduğu Şekil Uzayı en çok kabul gören ve en sık kullanılan yaklaşımdır. İkinci adım ise YBS elemanları arasındaki ilişkileri değerlendirmek için literatürde sunulan çeşitli ölçütlerden uygun olanını seçmektir. Örneğin iki eleman da ikili kodlanmış ise aralarındaki ilişki Hamming uzaklığı kullanılarak değerlendirilebilir. En son adımda ise sistemdeki dinamiği yönetmek için adaptasyon prosedürü olarak amaca uygun bir YBS algoritması seçilir (De Castro, ve Timmis, 2002).

4.2.1. Şekil uzayı

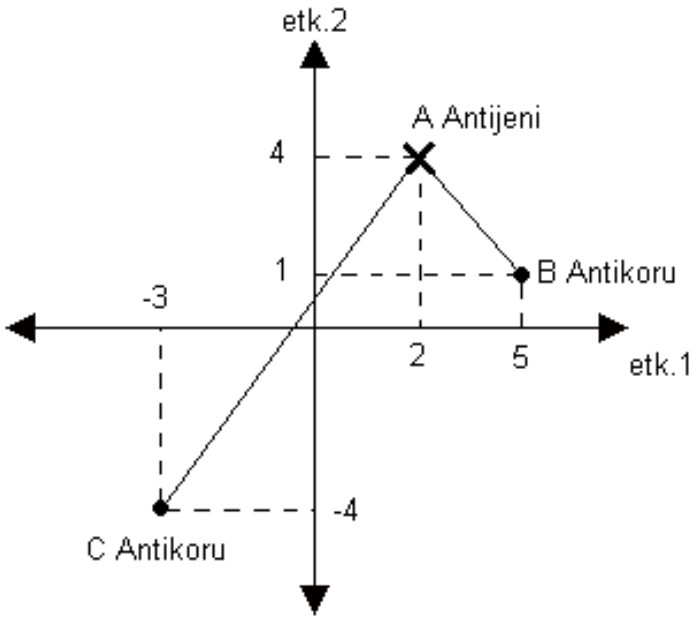
Bağışıklık elemanları arasındaki etkileşim pek çok fizikokimyasal olayı içerir. Mesela etkileşimdeki elemanlar arasında kovelent olmayan bağ, hidrojen bağı vb. bağların oluşması gibi. Bir antijenin bir antikor tarafından tanınması için, yüzeylerinin birbirlerini tamamlayacak şekilde bağlanmaları gerekir. Bu yüzden moleküller üzerinde geniş tamamlayıcı bölgelere ihtiyaç duyulur. Antijen ve antikorların birbirlerini tamamlamaları için gerekli olan şekil, elektrik dağılımları, uygun tamamlayıcı pozisyonları gibi özelliklerinin kümesine molekülün genelleştirilmiş şekli adı verilir.

Bir antikorun genelleştirilmiş şeklinin L tane parametre ile gösterilebildiğini varsayalım. Bu durumda L-boyutlu bir uzayda bir nokta, antijenin bağlanma özelliklerine göre bağışıklık hücrelerinin üzerindeki moleküler reseptörlerin antijene bağlandıkları alanların genelleştiriliş şeklini göstermiş olur ve şekil uzayı S olarak adlandırılır.

Antijen tanıma için gerekli olan tüm hücrelerin oluşturduğu sete, bağışıklık repertuarı denir. Vücudumuzun tüm antijenlerde korunabilmesi için bu repertuarın tam olması gereklidir. Vücut, herhangi bir antijeni tanıyabilecek reseptör moleküllerini

oluşturacak protein genlerine sahiptir ve bu genlerin farklı kombinasyonlarda dizilimi ile doğada bulunan hemen hemen tüm antijenleri tanıyabilir. Yani vücudumuzun bağışıklık repertuarı tamdır.

Bunu bir örnek ile açıklayalım. Mesela bir canlının *N*-boyutlu bir şekil uzayı olduğunu varsayalım. İlk başta da söylendiği gibi iki hücre birbirine bağlanırken aralarında pek çok fizikokimyasal olay meydana gelir. İşte varsayılan bu *N*-boyutlu şekil uzayında *N* tane eksen vardır ve her bir eksen, iki hücre arasında gerçekleşen fizikokimyasal etkileşimlerden birini temsil eder. Örneğin, bir *A* antijeni ile bir *B* Antikoru arasında iki çeşit fizikokimyasal etkileşim gerçekleşebilir kabulü yapılsın. Bu durumda her bir fizikokimyasal etkileşimi temsilen 2 eksen kullanılacağı için iki boyutlu bir şekil uzayı, *A* antijeni ile *B* ve *C* antikorunu göstermek için kullanılabilir. Şekil 4.7'de bu şekil uzayı gösterilmiştir (De Castro ve Von Zuben, 1999).



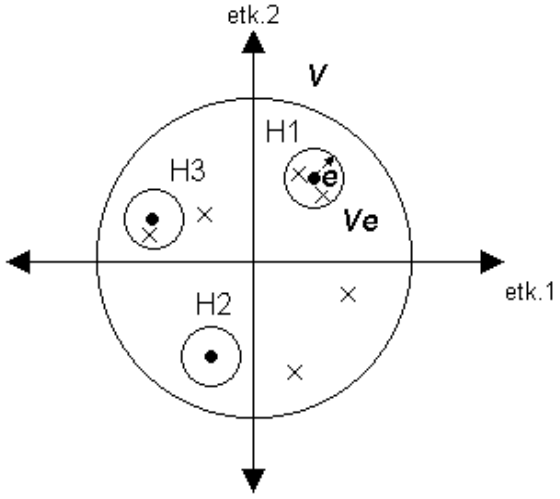
Şekil 4.7. Şekil uzayı gösterimi

Şekil 4.7'de *A* antijeni, *B* ve *C* antikoru ve sahip oldukları fizikokimyasal özellikler gösterilmektedir. *A* antijeni ile *C* antikoru arasındaki uzaklık, *B* antikorundan fazla olduğu için, *A* antijeni ile *C* antikoru arasındaki etkileşimin şiddeti, *A* antijeni ile *B* antikoru arasındaki etkileşimin şiddetinden daha fazladır diye gösterim yorumlanır.

Bağışıklık sisteminde antikorların yabancı maddeleri tanıyabilmeleri için bu antijenik yapılardan gelen uyarıların belirli bir eşik seviyesini aşması gerekmektedir. Bu özellik ise

şekil uzayında tanıma çemberi adı verilen ve Şekil 4.8'de görülen başka bir modelleme ile ifade edilmiştir.

Şekil 4.8'de üç tane antikor ve antijenler iki boyutlu bir şekil uzayında gösterilmişlerdir. Her bir hücrenin tanıma çemberi **e** eşik seviyesi ile belirlenen *Ve* çemberidir. Şekilde *H1* tanıma çemberi içinde kalan iki antijeni, *H3* ise sadece bir antijeni tanır. *H2*'nin tanıma çemberinde herhangi bir antijen olmadığı için *H3* antijen tanımamaktadır. Doğada mevcut olabilecek antijen türleri, protein dizilimleri ile belirlendiği için belirli bir çeşitlilikte bulunabilirler (De Castro ve Von Zuben, 1999).



Şekil 4.8. Tanıma çemberi şekil uzayı gösterimi

Şekil 4.8'deki iki boyutlu şekil uzayında bulunabilecek tüm antijen türleri bir *V* hacmi ile ifade edilmiştir. Başka bir deyişle bulunabilecek tüm antijenler mutlaka bu hacim içerisinde gösterilirler. Yine aynı şekil uzayında bulunan her bir antikorun bir *Ve* tanıma hacmi olduğuna göre, şekil uzayında *V* hacmi içerisinde tüm noktaları kapsayacak biçimde bir antikor popülasyonu olduğunda bu antikor popülasyonu tüm antijenleri tanıyabilir yani repertuar tamdır denebilir.

4.2.2. Duyarlılık hesaplaması

Matematiksel olarak *S* şekil uzayındaki herhangi bir *m* molekülün genelleştirilmiş şekli *L* uzunluğunda bir öznitelik katarı olarak gösterilebilir. Yani $m = <m_1, m_2, \ldots m_L>$ katarı *L* boyutlu bir şekil uzayında bir noktadır. Bu katar tam sayılar, gerçel sayılar, bitler veya sembollerden oluşabilir. Bu özniteliklerin türü YBS problemine göre seçilir. Bu

tanımlamada önemli olan şey hangi ölçüt veya ölçütlerin özniteliklerin etkileşim kalitesini yani duyarlılığını belirlemede kullanılacağıdır.

Duyarlılık genellikle bağışıklık birimlerinin birbirini tanıma kalitesini ifade eder. Çözülmesi beklenen problemin türüne göre duyarlılık hesabında kullanılacak yöntem belirlenebilir. Genellikle antijen ile antikor arasındaki uzaklığın hesaplanması kullanılır. Yukarıda da bahsedildiği gibi iki eleman birbirinden ne kadar uzak ise birbirlerini o kadar tamamlayacağı düşünülür. Elemanların gösterimine göre kullanılacak uzaklık ölçütü de değişir. Genellikle Öklid veya Manhattan uzaklığı kullanılır. Eğer antijenin ve antikorun ikili gösterimi söz konusu ise Hamming Uzaklığı kullanılır. Bu uzaklıkların formülize edilmiş halleri Çizelge 4.1'de gösterilmektedir. Çizelge 4.1'de antikorlar $Ab=<Ab_1, Ab_2, \ldots Ab_L>$, antijenler $Ag=<Ag_1, Ag_2, \ldots\ldots, Ag_L>$ olarak kabul edilmiştir.

Çizelge 4.1. Duyarlılık ölçütü için kullanılabilecek uzaklıklar

Uzaklık Ölçütü	**Formülü**
Öklid Uzaklığı	$D = \sqrt{\sum_{i=1}^{L} (Ab_i - Ag_i)^2}$
Manhattan Uzaklığı	$D = \sqrt{\sum_{i=1}^{L} \lvert Ab_i - Ag_i \rvert}$
Hamming Uzaklığı	$D = \sum_{i=1}^{L} \delta \qquad \delta = \begin{cases} 1 & , \quad Ab_i \neq Ag_i \\ 0 & , \quad Ab_i = Ag_i \end{cases}$

4.2.3. Klonal seçme algoritması (CLONALG)

Klonal Seçim Teorisi edinilmiş bağışıklığın çalışmasını açıklamak için kullanılan bir varsayımdır. Özellikle organizmayı istiladan korumak için kullanılan antikorların çeşitliliğini açıklayan bir teoridir. Bir antikor, tek bir antijeni nötralize edebilen B lenfosit hücreleri tarafından üretilen bir moleküldür. Her bir B lenfositi, benzersiz veya belirli bir türe özelleştirilmiş antikorlar oluşturur. Bu teori, organizmanın belirli bir özgünlük ile tüm antijenleri tanıyabilen önceden mevcut olan heterojen bir antikor havuzuna sahip olduğunu gösterir. Bir antijen bir antikor ile eşleştiğinde, bu durum hücrenin kimyasal olarak antijene bağlanmasına, replikasyonuna ve aynı reseptörlü daha fazla hücrenin üretilmesine neden olur. Bu hücre çoğalma aşaması boyunca klonlarda genetik mutasyonlar meydana gelir. Bu işlem hücrenin antijenle uygunluğunu veya benzerliğini

artırır. Antijenler tarafından yapılan bu replikasyon işlemi ile hayatta kalacak en uygun hücrelerin seçilmesi sağlanır ve genetik mutasyon ile hücre çeşitliliği korunur (De Castro ve Von Zuben 1999; Brownlee, 2005).

Bağışıklık sisteminde T ve B Lenfositlerinin olgunlaşması sırasında, Germinal Center'larda gerçekleşen *duyarlılık olgunlaşması* sırasında birçok seçme mekanizması devreye girmektedir. De Castro ve Von Zuben duyarlılık olgunlaşmasındaki işlemleri baz alarak Klonal Seçme Algoritmasını ortaya atmışlardır. Klonal Seçme Algoritması iki esas üzerine temellendirilmiştir:

1. Sadece antijeni tanıyan hücreler çoğalma için seçilirler,

2. Seçilen ve çoğalan hücreler duyarlılık olgunlaşması işlemine tabi tutularak, antijene olan duyarlılıkları artırılır.

CLONALG (CLONal selection ALGorithm), edinilmiş bağışıklıktaki klonal seçim teorisinin çalışmasından esinlenen bir yapay bağışıklık sistemi tekniğidir. Bu bölümde algoritmanın genel çalışma yapısı ve sistemde kullanıcı tanımlı parametreler hakkında detaylı bilgi verilmektedir (Brownlee, 2005).

CLONALG klonal seçim teorisinin aşağıdaki elementlerini kullanır (Brownlee, 2005):

- Özel bir hafıza kümesinin bakımı ve saklanması,
- En uyarıcı antikorların seçimi ve klonlanması,
- Uyarılmamış antikorların ortadan kaldırılması,
- Mutasyon,
- Klonların antijenlerin benzerliğiyle orantılı olarak tekrar seçimi,
- Farklılıkların üretimi ve korunması.

Algoritmanın amacı bir mühendislik problemine çözüm getirecek olan hafıza antikor havuzunu geliştirmektir. Algoritma tasarlanacak son hafıza antikor havuzunu araştırmak için iki mekanizma sağlar. Bunlardan birincisi klonlanan antikorların duyarlılık olgunlaşması aracılığıyla sağlanan bir yerel aramadır. Yerel aramanın kapsamı, seçilen antikor oranlarıyla ters orantılı olmasına rağmen; daha iyi eşleşme yapılan antikorlar için daha fazla klon üretilir. Bu sayede bu antikorların, olgunlaştırılacak daha geniş bir etki alanı içinde düşük özgünlüklü antijeni tanımasını sağlanır. İkinci arama mekanizması global bir kapsam sağlar ve ilerde çeşitliliği sağlamak için rasgele üretilen

antikorların popülasyona eklenmesini içerir. Ayrıca potansiyel olarak yerel optimadan kurtaracak bir yol temin eder (Brownlee, 2005).

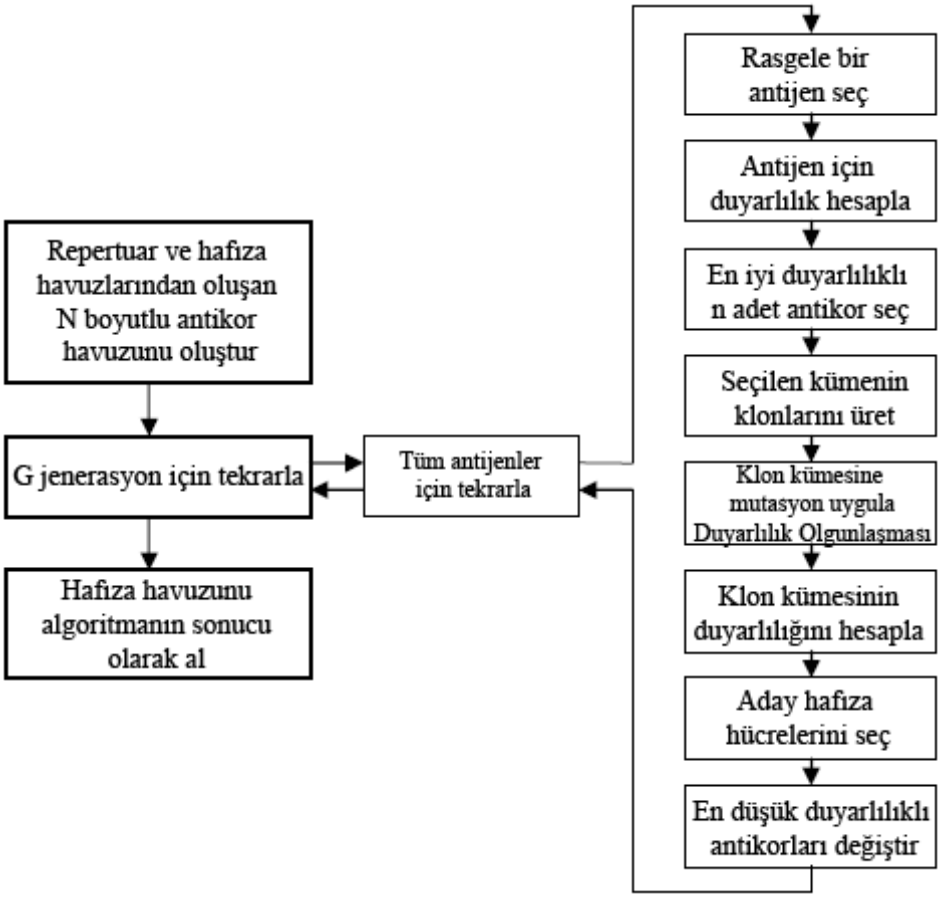
Aşağıdaki adımlar CLONALG algoritmasını genel hatlarıyla göstermektedir (Brownlee, 2005):

1. **Başla** – Algoritmanın ilk adımıdır. Bu adımda sabit bir *N* boyutlu antikor havuzu oluşturulur. Bu havuz daha sonra iki parçaya bölünür. Bunlar hafıza antikor havuzu *m* ve kalan antikor havuzu *r*. Hafıza antikorları en sonunda algoritmanın çözümünü gösterir. Kalan antikor havuzu ise sisteme çeşitlilik eklemek için kullanılır.
2. **Döngü** – Algoritma daha sonra bilinen tüm antijenler için sistemi belirli bir sayıda iterasyona maruz bırakarak çalışmasını devam ettirir. Tek bir tur çalıştırma veya bir iterasyon işlemi, bir jenerasyon olarak ifade edilir. Sistemde probleme özel bir sonlanma koşulu kullanılabilmesine rağmen, sistem jenerasyon sayısı *G*, kullanıcı tanımlı olarak çalıştırılabilir.
    a. **Antijen Seçimi** –Antijen havuzundan rasgele bir antijen seçilir.
    b. **Maruz Bırakma** – Sistem seçilen antijen için çalıştırılır. Duyarlılık değerleri, antijene karşı tüm antikorlar için hesaplanır. Duyarlılık bir benzerlik ölçüsüdür ve probleme bağlıdır. Yaygın olarak
    c. **Seçim** – *n* tanelik antikor kümesi, antijen ile en yüksek duyarlılığa sahip tüm antikor havuzundan seçilir.
    d. **Klonlama** – Seçilen antikorların kümesi duyarlılıklarıyla orantılı olarak klonlanır.
    e. **Duyarlılık Olgunlaşması (Mutasyon)** – Klonlar (kopya antijenlerin kümesi) söz konusu *m* antijeniyle daha iyi bir eşleşme sağlamak için bir duyarlılık olgunlaştırma sürecine tabi tutulur. Burada olgunlaşma derecesi ebeveynlerinin duyarlılığıyla ters orantılıdır. Yani daha yüksek duyarlılık daha az mutasyon demektir.
    f. **Klon Çalıştırma** – Daha sonra klonların duyarlılıkları hesaplanır.
    g. **Adaylık** – Klonlar içinde en yüksek uygunluklu antikor veya antikorlar aday hafıza antikorları olarak seçilir. Eğer bir aday hafıza hücresinin duyarlılığı, hafıza hücresindeki en yüksek uyarılmış antijenin duyarlılığından yüksek ise bunlar yer değiştirilir. Grup yer değiştirmeleri de benzer şekilde meydana gelir.

h. Yer Değiştirme – Son olarak en düşük duyarlılığa sahip *d* sayıda birey yeni rasgele antikorlar ile değiştirilir.

3. **Bitir** –Antijen havuzunun elemanlarından biri olan *m* sayıda hafıza elemanı çözüm olarak alınır. Probleme bağlı olarak, tek bir çözüm olabileceği gibi, havuzdaki tüm antijenler de çözüm olabilir.

CLONALG algoritmasının genel çalışma yapısını Şekil 4.9.'da verilmiştir (Brownlee, 2005)



Şekil 4.9. CLONALG algoritmasının genel çalışma şeması

YBS' de de sonuca etki eden parametreler şunlardır (Duman ve Akın, 2006):

- Populasyonun boyutu,
- Fitness fonksiyonunun uygunluğu ve çözünürlülüğü,
- *n* değerinin secimi,
- Mutasyon oranının seçimi (genelde genetik algoritmaya göre çok daha düşük bir oran),
- Yenileme oranı,

- Sonlandırma koşullarının belirlenmesi (Maksimum iterasyon sayısı veya amaçlanan doğruluk oranı).

CLONALG algoritmasının daha iyi anlaşılabilmesi açısından bir örnek uygulama üzerinde adım adım anlatılmıştır.

Problem: Verilen *x1* ve *x2* aralıklarına göre uygunluk fonksiyonunun maksimizasyonu istenmektedir.

Uygunluk Fonksiyonu: $f(x) = 21.5 + x1 \cdot \sin(4 \cdot \Pi \cdot x1) + x2 \cdot \sin(20 \cdot \Pi \cdot x2)$

Değişkenlerin Sınırlıkları: $-3 \le x1 \le 12.1, \qquad 4.1 \le x2 \le 5.8$

1. ADIM: Giriş kümesinin *(Ab)* oluşturulması.

Giriş kümesindeki antikor sayısı opsiyoneldir. Bu sayı kadar *x1* ve *x2* değişkenleri ilgili aralıklara uyularak rastgele üretilmiştir.

2. ADIM: Uygunluk fonksiyonlarının hesaplanması *(f)*.

Verilen sayı kadar rastgele üretilen *x1* ve *x2* değişkenleri kullanılarak her bir antikor için uygunluk fonksiyonu değeri *(f)* hesaplanmıştır.

3. ADIM: $Ab_{\{n\}}$ kümesinin oluşturulması.

Problemdeki amaç maksimizasyon olduğundan, *Ab* kümesindeki tüm antikorlar hesaplanan uygunluk fonksiyonu *(f)* değerlerine göre büyükten küçüğe doğru sıralanmıştır. Sıralanan *Ab* kümesinin elemanları içinden *n* tane antikor, algoritmanın devamında kullanılmak üzere seçilmektedir. Böylelikle $Ab_{\{n\}}$ kümesi oluşturulur.

4. ADIM: Klonlama

Bu adımda oluşturulan $Ab_{\{n\}}$ kümesindeki antikorlar klonlanacaktır. Klonlama işlemi denklem (4.1)'e göre gerçekleştirilmiştir:

$$C = Round\left(\frac{\beta \times s(Ab)}{i}\right) \qquad (4.1)$$

Formülü her bir antikor için hesaplanmıştır. Böylece her bir antikordan kaç adet klonlanacağı belirlenir. Burada;

β : dışarıdan girilen bir değişken,

$s(Ab)$: Ab giriş kümesinin eleman sayısı,

i: Klonlanacak antikorun Ab kümesi içindeki sıra numarasıdır.

5. ADIM: Hiper Mutasyon

Bu adımda, klonlanan antikorlar hiper mutasyona uğratılmıştır. İşlem denklem 4.2'ye göre gerçekleştirilir:

$$Round[(e^{\frac{-10\cdot(f(i))}{\max(f)}})\times 2] \qquad (4.2)$$

Formülü klonlanan her bir antikor için hesaplanmıştır. Burada *f(i)*; üzerinde çalışılan antikorun uygunluğu olmaktadır. 10 sabit bir sayı olup 2, problemdeki değişken sayısıdır (*x1* ve *x2*). *max(f)* ise hesaplanan uygunluk fonksiyonu değerlerinin en büyüğüdür. Formül sonucu bulunan değer:

a) 0 ise, hiper mutasyon gerçekleşmez.
b) 1 ise, değişkenlerden sadece birisi hipermutasyona uğratılacaktır. *x1* ve *x2* değişkenlerinden hangisinin hiper mutasyona uğratılacağını belirleyebilmek için rastgele olarak 0 ya da 1 sayısı üretilmiştir. 0 üretildiğinde *x1*, 1 üretildiğinde *x2* hiper mutasyona uğratılmıştır. Hiper mutasyon işlemi, rastgele sayı üretimi sonucunda ilgili değişkenin (*x1* yada *x2*) verilen sınırlar dahilinde yeniden rastgele üretilmesiyle gerçekleştirilmiştir.
c) 2 ise, *x1* ve *x2* değişkenlerinin her ikisi de hiper mutasyona uğratılacaktır. Hiper mutasyon işlemi, değişkenlerin verilen sınırlar dahilinde yeniden rastgele üretilmesiyle gerçekleştirilmiştir.

6. ADIM: Hiper mutasyona uğratılan antikorların uygunluk fonksiyonu değerleri hesaplanır.

7. ADIM: Seçim

Bu adımda klonlanmış, hiper mutasyona uğramış ve uygunluk fonksiyonları hesaplanmış antikorlardan oluşan büyük boyutlu bir küme vardır. Bu kümedeki antikorlar hesaplanan uygunluk fonksiyonlarına göre tekrar büyükten küçüğe doğru sıralanmıştır ve

içlerinden tekrar *n* tanesi seçilmiştir. Algoritmaya, klonlanmış ve hiper mutasyona uğratılmış antikorların içinden seçilen *n* tanesi ile devam edilecektir.

8. ADIM: *Ab* kümesinin yeniden oluşturulması.

Sonraki iterasyona başlayabilmek için *Ab* kümesinin tekrar oluşturulması gereklidir. Burada, bir önceki adımda seçilen n tane antikor *Ab* kümesine dahil edilmiştir. Kümenin geri kalanı ise ilgili antikorların problemde verilen aralıklarda yeniden rastgele üretilmesiyle tamamlanmaktadır.

Böylece oluşturulan Ab kümesi ile sonraki iterasyon tekrar ederek devam etmektedir.

4.3. Bölüm Değerlendirmesi

Bu bölümde, insan bağışıklık sitemleri hakkında temel bilgiler verilmiştir. Ayrıca yapay bağışıklık sistemlerinden bahsedilmiştir. Edinilmiş bağışıklıktaki klonal seçim teorisinin çalışmasından esinlenen bir yapay bağışıklık sistemi tekniği olan CLONALG algoritmasının genel çalışma yapısı ve sistemde kullanıcı tanımlı parametreleri hakkında bilgiler verilmektedir. CLONALG algoritmasının daha iyi anlaşılabilmesi açısından da bir örnek uygulama üzerinde adım adım anlatılmaktadır.

5. INDISC ALGORİTMASI

Şimdiye kadar geliştirilen kural çıkarma yöntemleri arasında bu veya diğer yöntemlerle %100 başarıya ulaşan çok az yöntem mevcuttur. Oysa başarı oranının artırılması gerektiği konuyla ilgili çalışmaların çoğunda bahsedilmektedir. Bu bölümde çok sınıflı verilerden kural çıkarımının başarı oranını artırmak için yeni bir yöntem açıklanmaktadır. Geliştirilen bu yöntemde ayrık ve gerçel öznitelikler farklı şekilde kodlanması öngörülmüştür. Ayrık öznitelikler ikili olarak, gerçek öznitelikler ise, iki gerçel değer kullanılarak kodlanmıştır. Gerçel değerler kuralları oluşturan özniteliklerin değer aralıklarının orta noktası ve genişlemesini ifade etmektedir. Kural çıkarım işlemi için sınıflandırma başarısı uygunluk fonksiyonu olarak kullanılmıştır. Uygunluk fonksiyonunun optimizasyonu amacıyla Yapay Bağışıklık Sistemi (YBS) yöntemlerinden olan CLONALG algoritması kullanılmıştır. Önerilen yöntem en uygun aralıkları keşfettiğinden dolayı bu yönteme isim olarak "Aralık Keşfi" anlamına gelen INDISC (INterval DISCovery) verilmiştir. Bu yöntemin temel adımları ve çalışma prensibi bu bölümde açıklanacaktır.

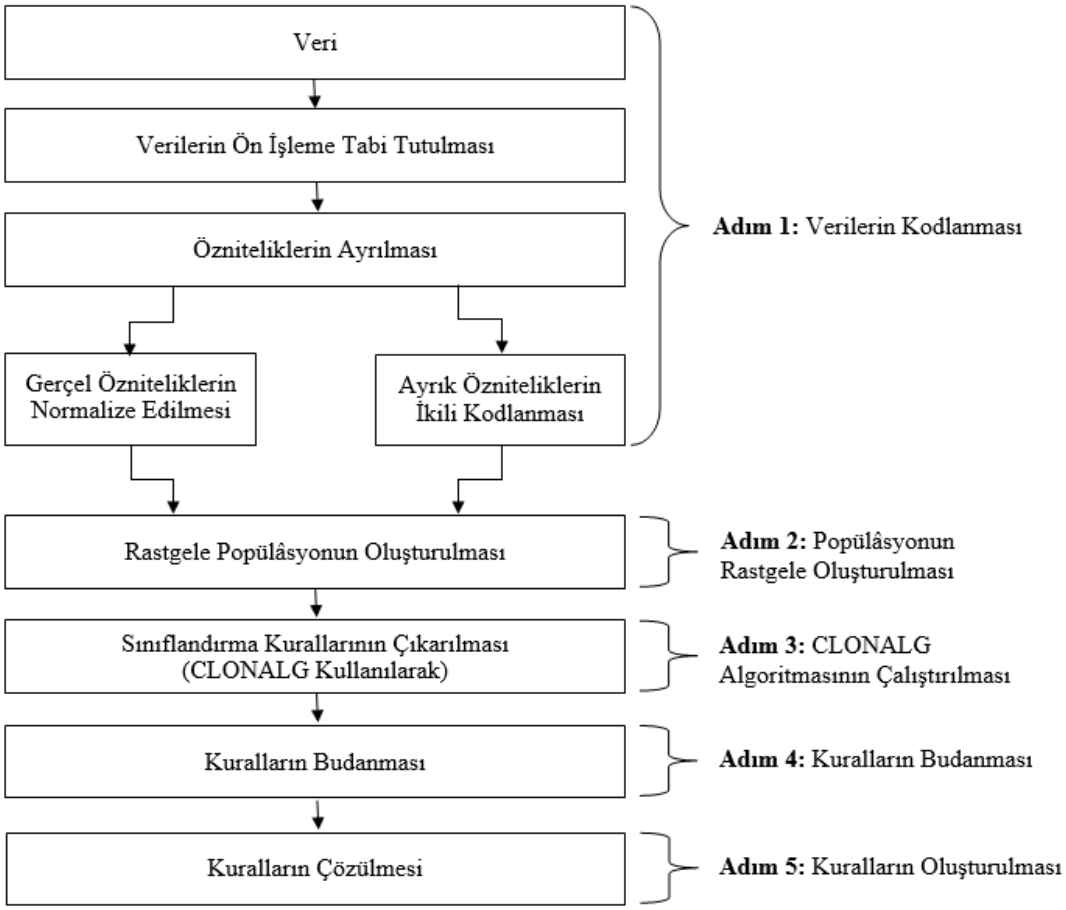
5.1. INDISC Algoritma Yapısı

Verilerden kural çıkarmak için önerilen yaklaşımda sınıflandırma doğruluk fonksiyonu CLONALG kullanılarak optimize edilmektedir. Önerilen algoritmanın temel adımları ve algoritmanın akış diyagramı Şekil 5.1'de gösterilmektedir.

5.1.1. Veri kodlanması

Veri kümesindeki öznitelikler ayrık ve gerçel olarak iki gruba ayrılmıştır ve farklı yöntemlerle kodlanmıştır. Ayrık öznitelikler ikili olarak kodlanmıştır. Bunun için aşağıdaki yöntem kullanılmıştır. Veri N sayıda kesin niteliğe sahip olsun. Her A_n $\{n=1,2,...,N\}$ özniteliği m_n sayıda $\{a_1, a_2, ..., a_{m_n}\}$ biçimli alt katara (string) bölünür ve $\{b_{n1}, b_{n2}, ..., b_{nm_n}\}$ olarak kodlanır.

Eğer A_n özniteliği a_i $(i=1,2,....m_n)$ alt katarına aitse, b_{nj} denklem (5.1)'deki, ayrık özniteliklerin kodlanmış hali denklem (5.2)'de, m vektörünün uzunluğu denklem (5.3)'deki gibidir.



Şekil 5.1. Önerilen algoritmanın temel adımları ve algoritmanın akış diyagramı

$$b_{nj} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \qquad j=1,2,\ldots,m_n \tag{5.1}$$

$$X = \bigcup_{n=1}^{N} \bigcup_{i=1}^{m_n} b_{ni} \tag{5.2}$$

$$m = \sum_{n=1}^{N} m_n \tag{5.3}$$

Gerçel özniteliklerin kodlanması:

Veri *K* adet gerçel niteliğe sahip olsun. Her bir L_k *{k = 1,2, ..., K}* özniteliği iki sayı ile temsil edilir. İlk *K* adet $\{l_{11}, l_{21}, \ldots, l_{K1}\}$ sayıları aralıkların orta noktalarını, ikinci *K* adet $\{l_{12}, l_{22}, \ldots, l_{K2}\}$ sayıları genişlemelerini denklem (5.4) belirtmektedir.

$$l_i \in (l_{i1} - l_{i2}, l_{i1} + l_{i2}) \quad i=1,2,...,K \tag{5.4}$$

Gerçel özelliklerin kodlanması denklem (5.5) ile hesaplanır. *XR* vektörünün uzunluğu *t=2*K'*dır.

$$XR = \bigcup_{n=1}^{K} \bigcup_{i=1}^{2} l_{ki} \tag{5.5}$$

Sonuç olarak veri kodlandıktan sonra oluşan vektör denklem (5.6)'deki gibi elde edilir.

$$V = X \cup XR = \left(\bigcup_{n=1}^{N} \bigcup_{i=1}^{m_n} b_{ni} \right) \cup \left(\bigcup_{n=1}^{K} \bigcup_{i=1}^{2} l_{ki} \right) \tag{5.6}$$

Vektörlerin toplam uzunluğu ise denklem (5.7) ile hesaplanır.

$$uz = m + t = \sum_{n=1}^{N} m_n + t \tag{5.7}$$

5.1.2. Kullanılan duyarlılık fonksiyonu

Bu çalışmada sınıflandırma kural çıkarımı için CLONALG algoritması kullanılmıştır. Duyarlılık için sınıflandırma başarısı kullanılmış ve sınıflandırma başarısı için denklem (5.8) kullanılmıştır (Parpinelli ve ark., 2001):

$$F = \frac{TP * TN}{(TP + FN) * (FP + TN)} \tag{5.8}$$

Burada *N* örneklerin toplam sayısıdır. ***TP*** (True Positives / Doğru Pozitifler) kural tarafından sınıfları doğru tahmin edilmiş örneklerin toplam sayısıdır. ***FP*** (False Positivies / Yanlış Pozitifler) kural tarafından sınıfları yanlış tahmin edilmiş örneklerin toplam sayısıdır. ***FN*** (False Negativies / Yanlış Negatifler) kuralın ait olduğu sınıfa ait ama bu kural tarafından belirlenememiş örneklerin toplam sayısıdır. ***TN*** (True

Negativies / Doğru Negatifler) kural tarafından belirlenememiş ve başka sınıfa ait olan örneklerin toplam sayısıdır.

5.1.3. Kural Çözme

Son aşama olarak bulunan en iyi antikorun kodu çözülür. Kural çözme işlem adımları kaba kodu Şekil 5.2'de gösterilmektedir.

Adım 1. Optimizasyon sonucu elde edilmiş antikorların kodu çözülür.
Adım 2. Her bir antikor iki alt kümeye ayrılır; Her antikorun ilk *m* elemanı ayrık öznitelikleri belirtir ve *X* ile gösterilir. Diğer *t* eleman gerçel öznitelikleri belirtir ve XR ile gösterilir.
    Adım 2.1. Ayrık öznitelikleri belirten *X*'in kodu çözülür.
        Adım 2.1.1. *X* kümesi *N* parçaya bölünür. Her parça A_n özniteliğini belirtir ve uzunluğu m_n'dir.
        Adım 2.1.2. Denklem (5.1)'e göre kodu çözülür.
        Adım 2.1.3. Aynı özniteliğin farklı değerlerini birleştirmek için "VEYA" kullan. Farklı öznitelikleri birleştirmek için "VE" kullanılır.
    Adım 2.2. Gerçek öznitelikleri belirten *XR*'nin kodu çözülür.
        Adım 2.2.1. XR' yi iki parçaya böl. İlk parça aralıkların orta noktalarını, ikinci parça genişlemelerini temsil eder.
        Adım 2.2.2. Denklem (5.4)'e göre kodu çözülür.
        Adım 2.2.3.Farklı özelliklerini birleştirmek için "VE" kullanılır.
Adım 3. *Adım 2.1.3* ve *Adım 2.2.3*'te bulunmuş alt kuralları "VE" ile birleştirerek kural tabanını oluşturulur.

Şekil 5.2. Kural çözme işleminin kaba kodu

5.1.4. Kuralların Budanması

Kurallar oluşturulduktan sonra budama aşamasına geçilmektedir. Bu amaçla Kim ve ark. (2005) çalışmalarında teklif edilen yöntemin benzeri kullanılmıştır. Kim ve ark., (2005) çalışmalarında öznitelik seçmek için tüm öznitelikleri sırayla kaldırmış, başarıya olan etkiyi incelemişlerdir. Bu çalışmamızda benzer işlemler kurallar için yapılmıştır. Tüm kurallar tek tek kaldırılarak bu durumda başarı aynı kalıyor ya da düşüyorsa ilgili

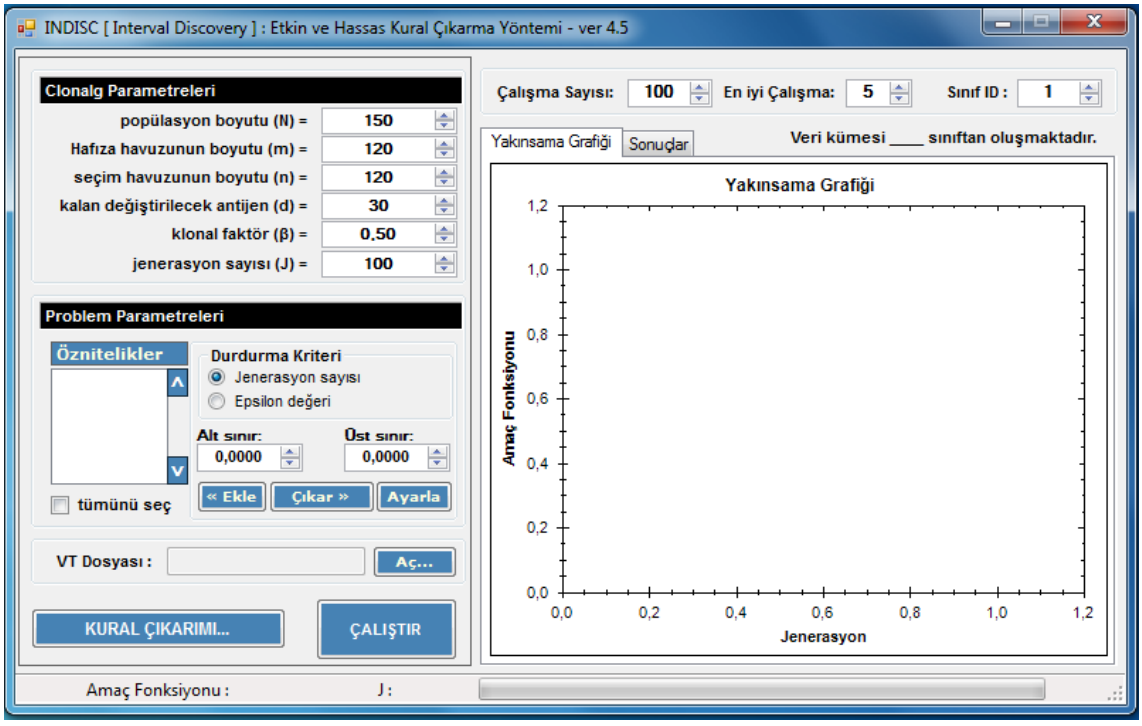
kuralın budanması sağlanmıştır. Eğer başarı yükseliyorsa ise ilgili kural kalmakta ve diğer kurallar incelenmeye devam edilme prensibiyle çalışmaktadır.

5.2. INDISC Algoritmasının Program Yazılımı

Bölüm 5.1'de teorik bilgileri ve detayları verilen INDISC yöntemi yazılım haline getirilmiştir. Bu bölümde INDISC yönteminin yazılım detayları anlatılacaktır.

5.2.1. INDISC yazılımının arayüzü

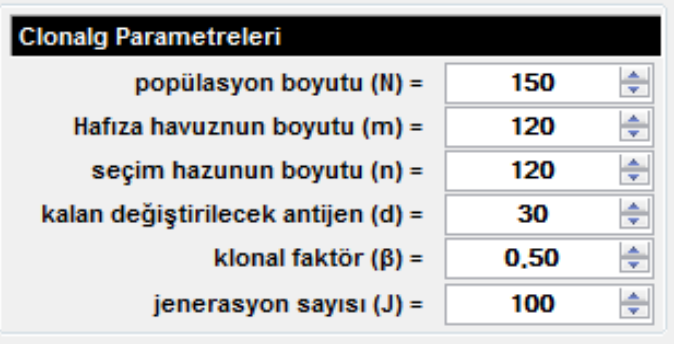
Bu çalışmada, INDISC yöntemiyle kural çıkarımı için C# programlama dili ile nesne yönelimli programlamaya uygun olarak geliştirilmiştir. Geliştirilen INDISC yazılımının genel arayüzü Şekil 5.3'de gösterilmektedir. Bu arayüz CLONALG parametrelerinin ve problem parametrelerinin ayarlanabildiği; aynı zamanda programın çalışma esnasında gerçek zamanlı elde edilen verilerin yakınsama grafiğinin ön izlemesinin yapılabildiği yazılımın ana arayüzüdür. Ayrıca kural çıkarma ve programın çalıştırılması gibi fonksiyonlar bu arayüzden kontrol edilmektedir.



Şekil 5.3. INDISC yazılımının arayüz ekranı

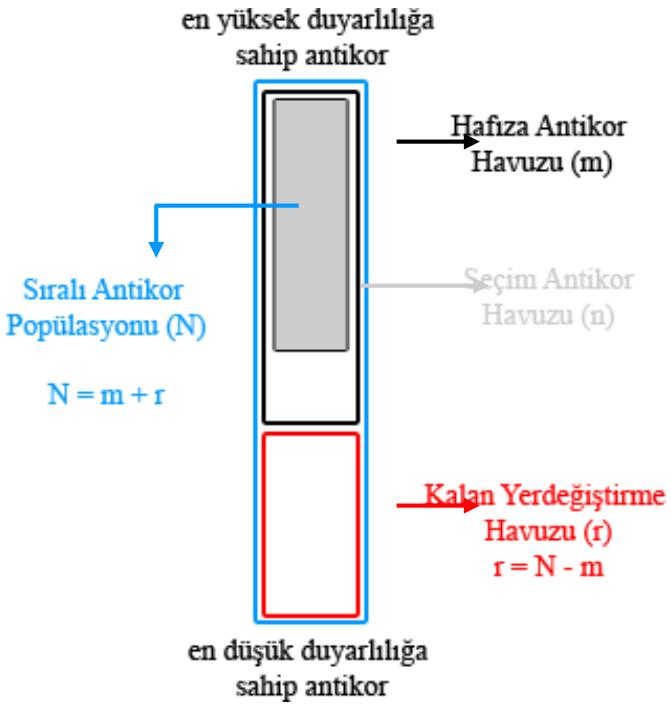
5.2.1.1. CLONALG parametreleri

CLONALG tekniği nispeten düşük bir karmaşıklığa sahiptir ve az sayıda kullanıcı tanımlı parametresi vardır. Bunlar; antikor popülasyonun boyutu (*N*), hafıza popülasyonunun boyutu (*m*), seçim havuzunun boyutu (*n*), kalan yer değiştirme boyutu (*d*), klonal faktör (*β*) ve jenerasyon sayısı (*J*) olarak sıralanabilir. CLONALG parametrelerinin ayarlanabildiği bölüm Şekil 5.4'de gösterilmektedir.



Şekil 5.4. CLONALG parametreleri

Ayrıca parametrelerin antikor popülasyonunda gösterimi ise Şekil 5.5'de gösterilmektedir.



Şekil 5.5. Parametrelerin antikor popülasyonunda gösterimi

Antikor popülasyonun boyutu (*N*): Sistem tarafından kullanılacak toplam antikor sayısını belirtir. Hafıza ve kalan antikor havuzlarını kapsar. Antikor havuzunun ne şekilde bölündüğü belirli değildir. Basit bir yaklaşım olarak seçilecek hafıza havuzunun boyutu $m \leq N$ olarak tanımlanmalıdır. Böylece kalan antikor havuzunun boyutu $r = N - m$ şeklinde hesaplanır.

Seçim havuzunun boyutu (*n*): Her bir antijenin gösteriminde klonlamak için antikor popülasyonundan alınan en yüksek duyarlılıklı antikorların toplam sayısını belirtir. $n \in [1, N]$. n değeri, yüksek duyarlılığa ulaşacak popülasyondaki seçim baskısının miktarını tanımlar. Daha küçük değerler baskıyı artırır ve sadece en iyi duyarlılığı koruyarak popülasyonun çeşitliliğini azaltabilirler.

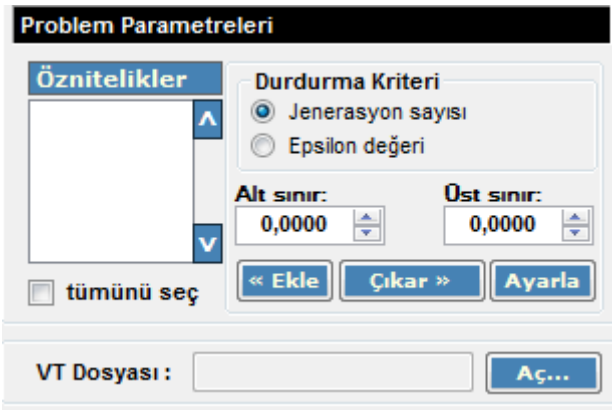
Kalan yer değiştirme boyutu (*d*): Her bir antijen uygulamasında rasgele antikorlar ile yer değiştirecek en düşük duyarlılıklı antikorların toplam sayısını belirtir. $d \in [0, r]$. Burada r kalan antikor havuzunun boyutudur. Sistemin çalışmasına ek bir çeşitlilik getirmek için kullanılır. Değeri sıfıra çekilerek iptal edilebilir.

Klonal Faktör (β): Seçilen antikorlarda oluşturulan klon sayısı için bir ölçekleme faktörünü gösterir. Yaygın olarak β (0,1] aralığında alınır. Bu değer ne kadar düşerse, algoritma tarafından yerel alanda o kadar daha fazla arama işlemi gerçekleştirilir.

Jenerasyon sayısı (*J*): Yürütülecek algoritma iterasyonlarının toplam sayısını gösterir. Tek bir iterasyon işleminde sistem bilinen tüm antijenlere maruz kalır. Bu parametre öğrenme miktarını kontrol eder.

5.1.1.2. Problem parametreleri

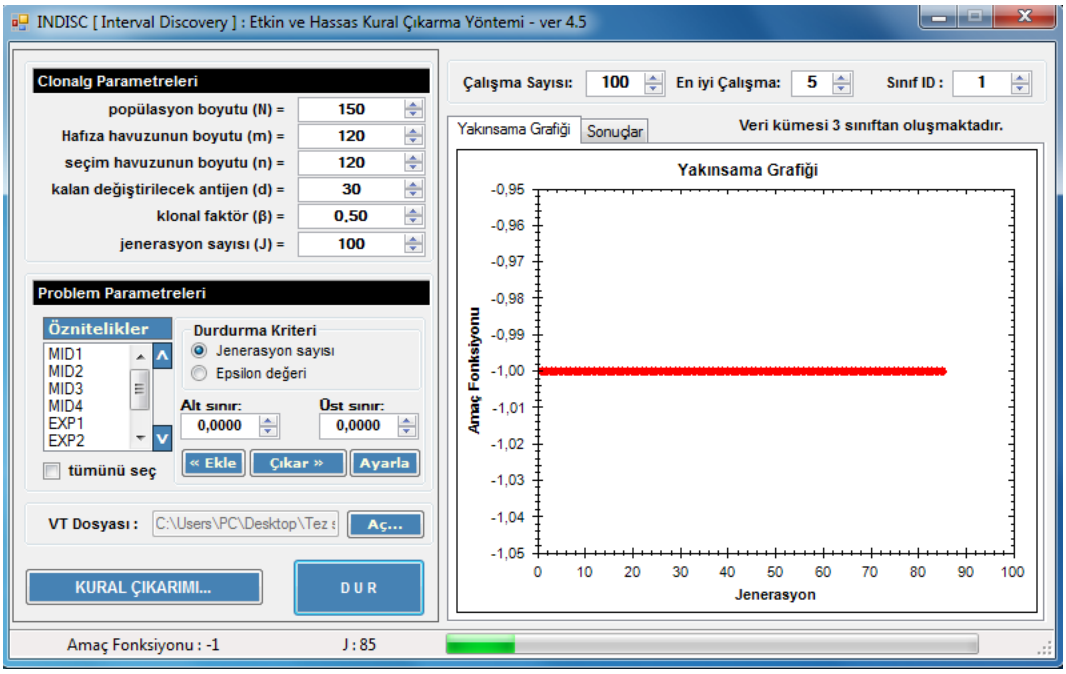
Veri kümesinin özniteliklerinin eklenip çıkarıldığı, durdurma kriterinin seçildiği ve veri tabanına ekleme işlevlerinin yerine getirildiği arayüz Şekil 5.6'da gösterilmektedir.



Şekil 5.6. Problem parametreleri

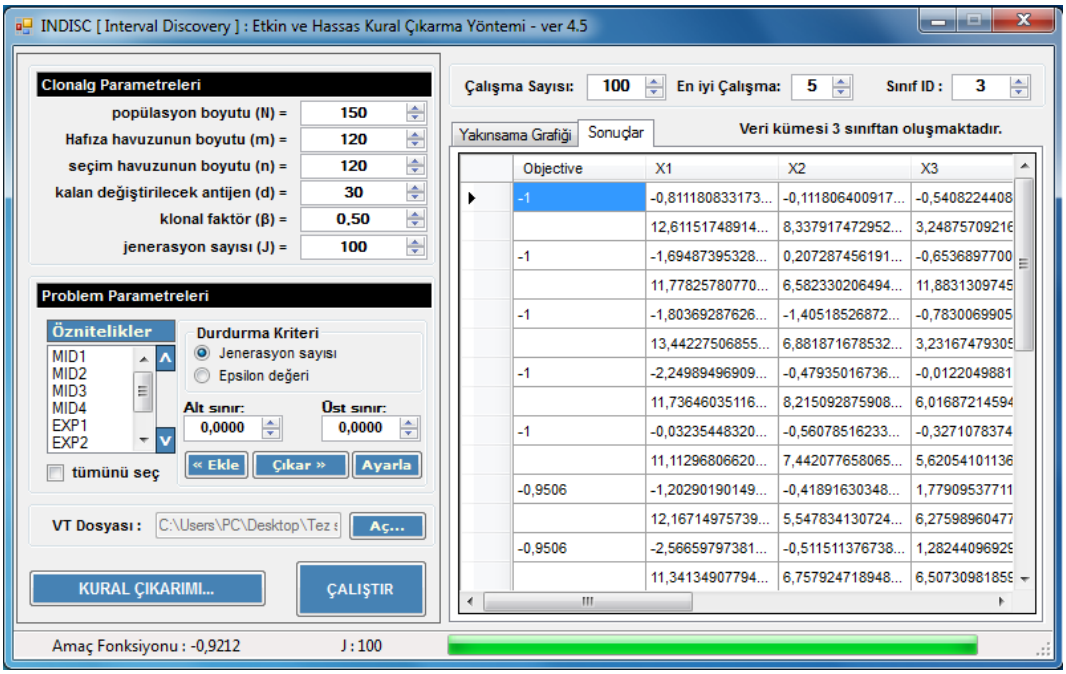
5.1.1.3. Grafik ve sonuçlar

INDISC yazılımında uygun CLONALG parametrelerinin girilmesi ve Veritabanının (öznitelikler ve hangi sınıf olduğunu belirten verilerinin yazıldığı "txt" uzantılı veri dosyası) sisteme yüklenmesinden sonra program çalıştırılmaktadır. Örnek veritabanı dosyası ekler bölümünde EK-1'de verilmektedir. Programın gerçek zamanlı elde edilen sonuçların yakınsama grafiği Şekil 5.7'deki görülmektedir.



Şekil 5.7. Programın çalışma esnasındaki oluşan yakınsama grafiği

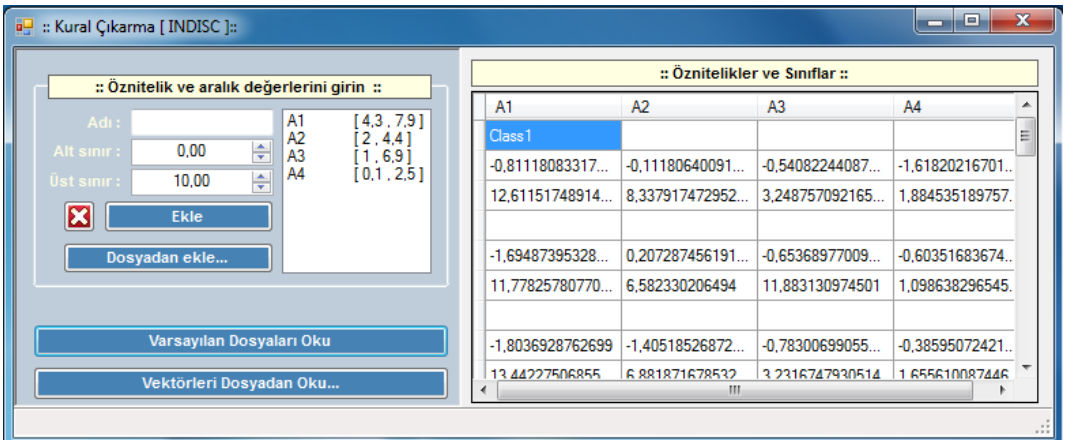
INDISC yazılımının çalıştırıldıktan sonra ekran kuralları oluşturacak vektörleri gösteren sonuçlar bölümü gelmektedir. Bu durum Şekil 5.8'de görülmektedir.



Şekil 5.8. Kuralları oluşturacak sonuç ekranı

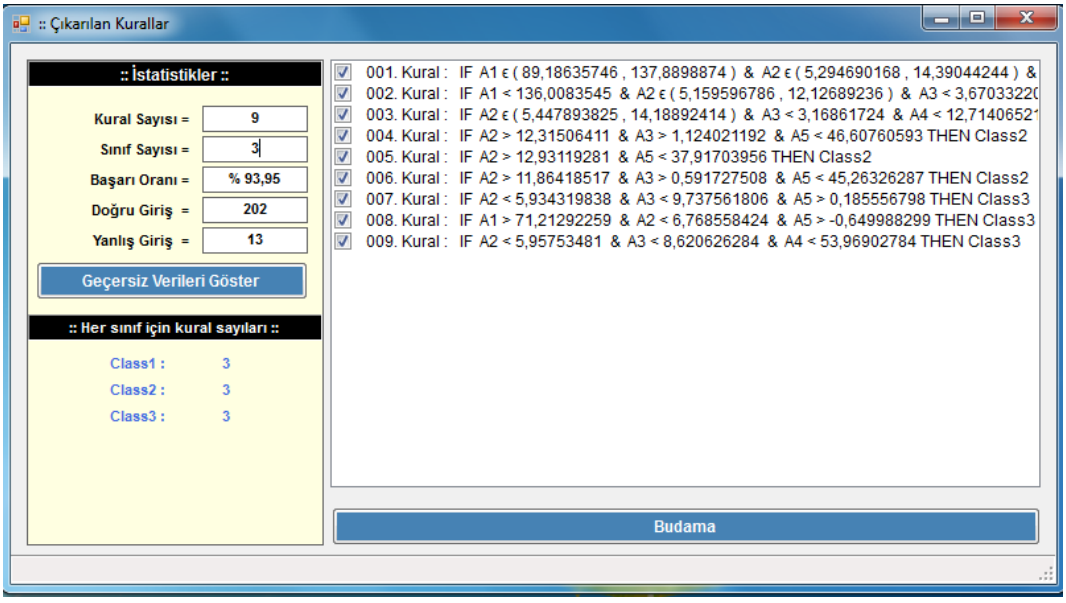
5.2.2 INDISC yazılımıyla kural çıkarımı

INDISC yazılımının diğer bir özelliği ise özniteliklerin tek tek elle girilebilmesinin yanı sıra uygun formatlı bir text dosyadan yüklenerek de gerçekleştirilebilmesidir. Bu sayede daha önce elde edilmiş sonuçların kuralları, dosyadan okunarak vakit kaybetmeden yeniden incelenebilmektedir. Örnek öznitelik dosyası ekler bölümünde EK-2'de verilmektedir. Kuralları ve sınıfları oluşturacak sonuçları görebileceğimiz kısım Şekil 5.9'da gösterilmektedir.



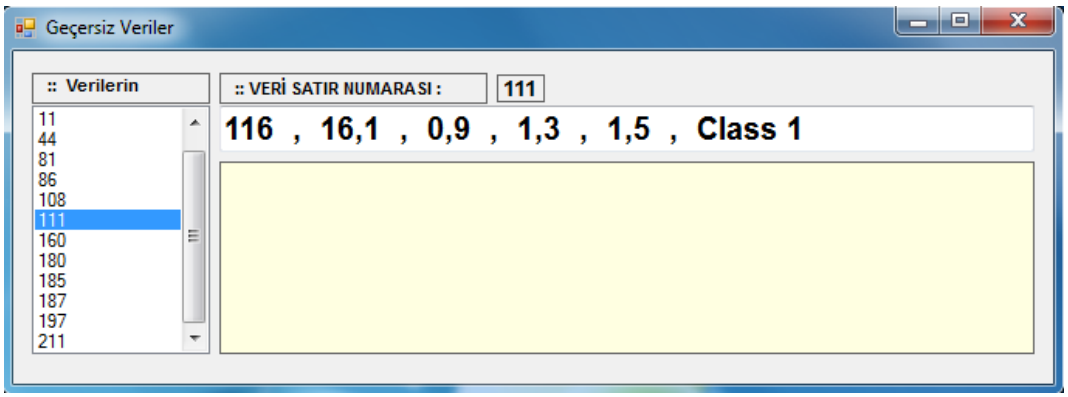
Şekil 5.9. Öznitelikler ve sınıfların sayısal değerleri

INDISC yazılımında kuralların bulunduğu arayüz kısmı Şekil 5.10 da görülmektedir. Burada görülen kurallar CLONALG algoritması ile bulunan vektörlerin çözümlenmesinden sonra ortaya çıkan kurallardır. Bu bölüm ayrıca; çıkan kurallar hakkındaki istatistiklerin yer aldığı bölümdür. Burada çıkan kural sayısı, veri kümesinin kaç sınıftan oluştuğu, üretilen kuralların doğru sınıfları teşhis edip edemediği ekranda gözükmektedir. Ayrıca her sınıf için çıkan kural sayıları ve doğruluk başarı oranı ekrandan listelenmektedir.



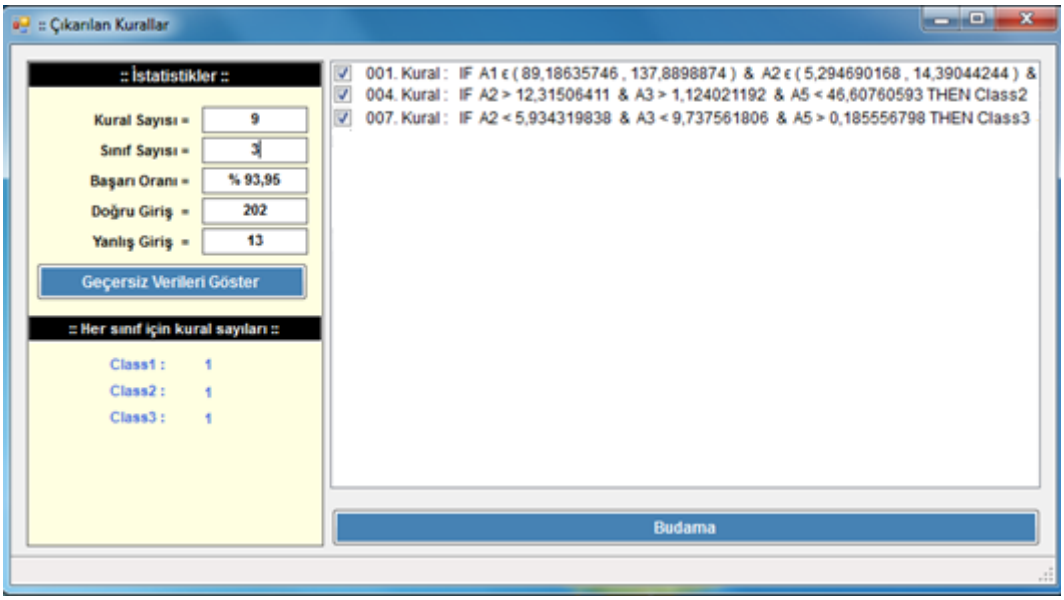
Şekil 5.10. Çıkan kuralların listesi

Programın bilgi amaçlı kullanılan "Geçersiz Verileri Göster" butonu basıldığında varsa teşhis edilemeyen sınıf ve öznitelik değerleri Şekil 5.11'deki gibi ekrana gelmektedir.



Şekil 5.11. Teşhis edilemeyen sınıf ve öznitelik değerleri

Kurallar oluşturulduktan sonra budama aşamasına geçilmektedir. Programda tüm kurallar tek tek kaldırılarak sonuçlar değerlendirilir. "Başarı aynı kalıyor yâda yükseliyorsa ilgili kural buda; başarı düşüyorsa ilgili kurala dokunma ve diğer kuralları incelenmeye devam et" prensibiyle çalışmaktadır. "Budama" butonuna basılarak budama işlemi tamamlanmakta ve en son nihai kurallar belirlenmiş olmaktadır. Örnek olarak budanmış kural listesi arayüzü Şekil 5.12'de gösterilmektedir.



Şekil 5.12. Budama sonrası nihai kuralların oluşturulması

5.3. INDISC Yazılımı İçin Örnek Uygulama

Bu uygulama için çok sınıflı veri kümelerinden süsen çiçeği (iris) veri kümesini kullanılacaktır. Bu veri kümesi 3 sınıflıdır ve 4 özniteliği bulunmaktadır. Bu veri kümesi her sınıftan 50 örnek olmak üzere toplam 150 örneğe sahiptir. Sınıf 1 (İris Setosa), Sınıf 2 (İris Versicolor) ve Sınıf 3 (İris Virginica) olarak sınıflandırılmıştır (Frank ve Asuncion, 2011).

Yapılan bu uygulama aşağıda adım adım açıklanmaktadır.

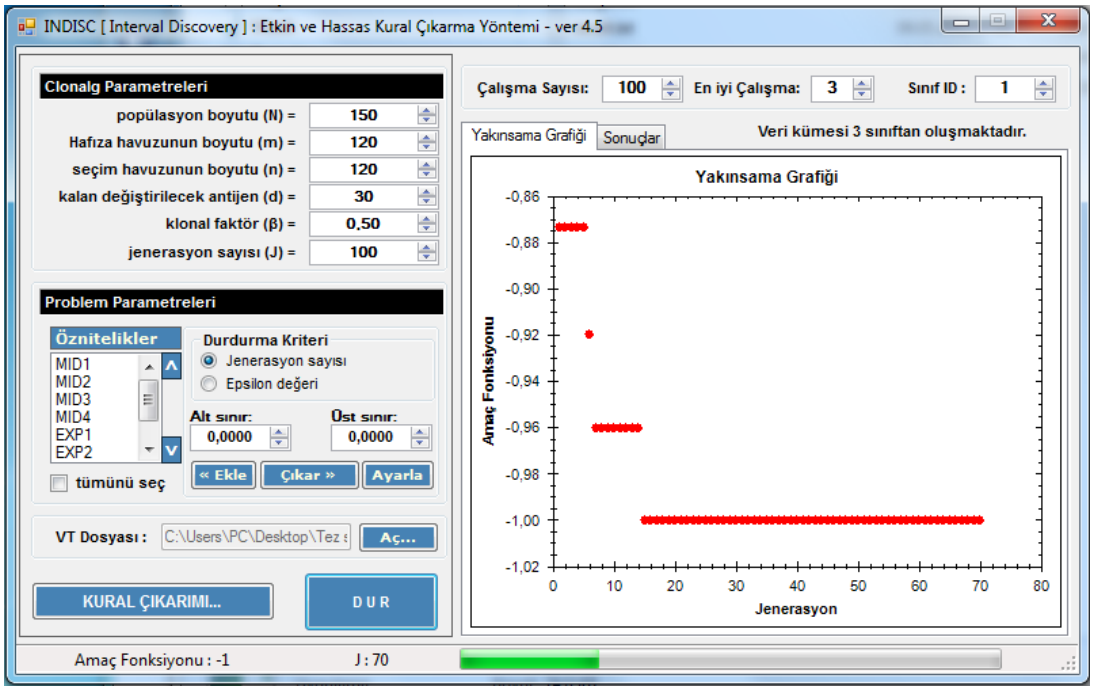
Adım 1: Programı açtıktan sonra öncelikli olarak "VT Dosyası:" kısmından "Aç…" butonuna basılarak süsen çiçeği veri kümesi veritabanı programa yüklenir.

Adım 2: CLONALG parametreleri ayarlanır.

Adım 3: Öznitelikler dosyadan yüklenir.

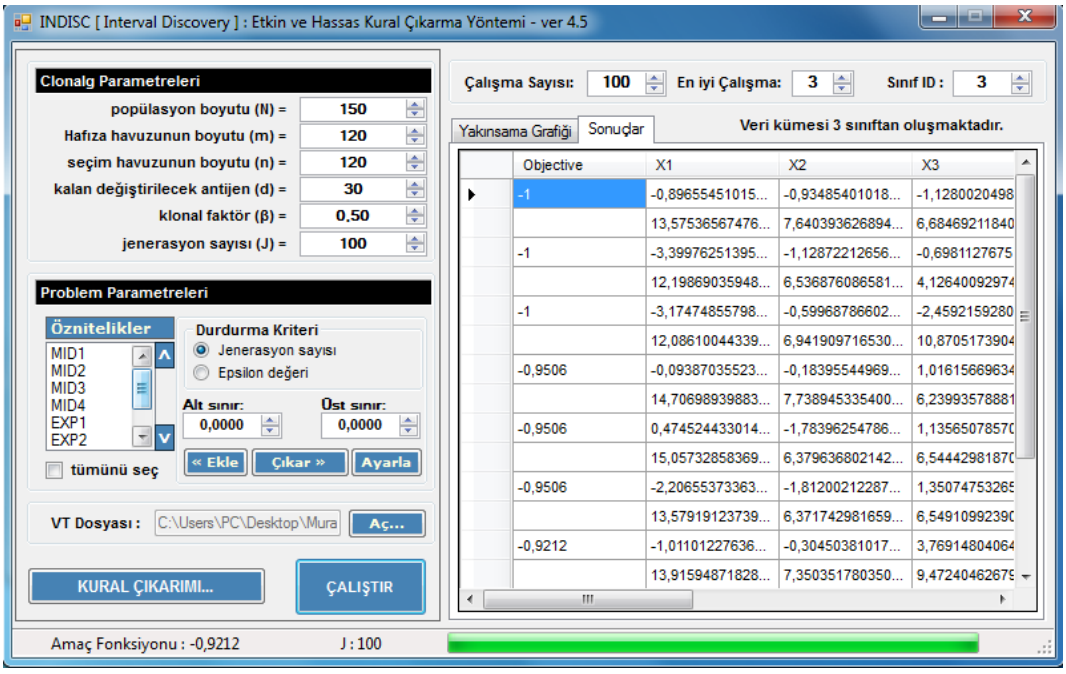
Adım 4: Çalışma sayısı belirlenir.

Adım 5: Çalıştır butonuna basılarak program çalıştırılır. (Şekil 5.13).



Şekil 5.13. Örnek uygulama için programın çalışma ekranı

Adım 6: Programın çalışmasını bitirdikten sonra ekrana sonuçlar gelmektedir (Şekil 5.14)

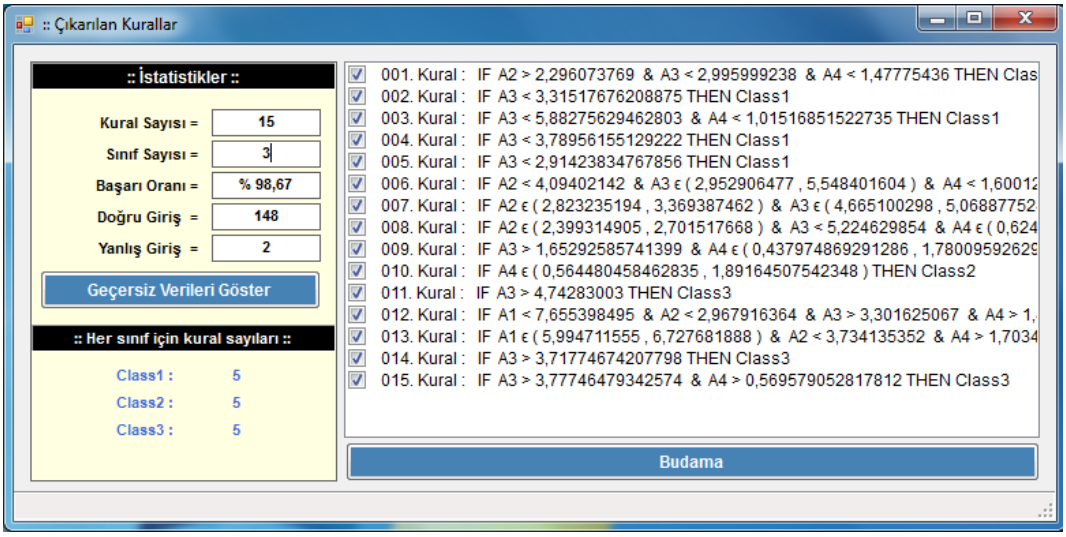


Şekil 5.14. Örnek uygulama için programın sonuç çıktı ekranı

Adım 7: "Kural Çıkarımı…" butonuna basılarak kural çıkarma arayüzüne gidilir.

Adım 8: "Dosyadan ekle" butonuna basılarak öznitelikler eklenir.

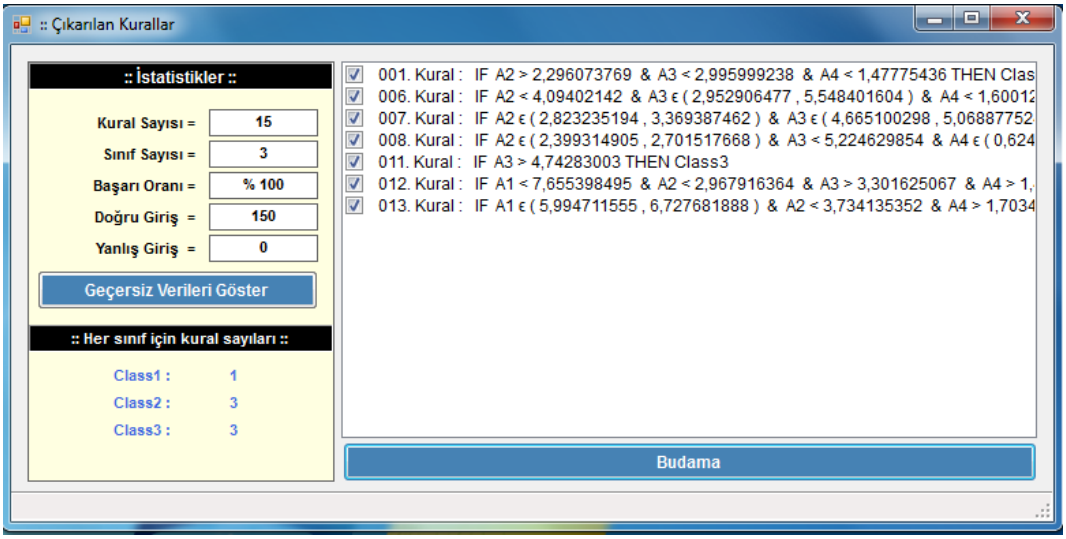
Adım 9: "Vektörleri Dosyadan oku" butonuna basılarak daha önce oluşan sonuç dosyasından çözümlenen vektörlerden kurallar çıkartılır (Şekil 5.15).



Şekil 5.15. Örnek uygulama için oluşan kurallar

Adım 10: Kurallar çıktıktan sonra budama aşamasına geçilmektedir. Programda tüm kurallar tek tek kaldırılarak bu durumda başarı aynı kalıyor yâda yükseliyorsa ilgili kuralın "Budama" butonuna basılarak budanması sağlanır. Eğer başarı düşüyor ise ilgili kural kalmakta ve diğer kurallar incelenmeye devam edilmektedir.

Adım 11: Budama işlemlerinin bitiminden sonra başarı oranı en yüksek olan kurallar kalmakta veri kümemizin kuralları oluşturmaktadır (Şekil 5.16).



Şekil 5.16. Budama sonrası kalan kurallar

Geliştirilen bu uygulamanın sonucu olarak, 7 kural (1 kural sınıf 1, 3 kural sınıf 2 ve 3 kural sınıf 3 için) oluşturulmuştur. Bu veri kümesi için örneklerin %100'ü doğru olarak sınıflandırılmıştır.

5.4. Bölüm Değerlendirmesi

Bu bölümde önerilen kural çıkarma yöntemi INDISC'in algoritma yapısı, bu yöntemde verilerin nasıl kodlandığı, kullanılan duyarlılık fonksiyonu, oluşan kuralların nasıl çözüldüğü ve çıkan kuralların budanarak en son nihai kuralların oluşturulduğu anlatılmaya çalışılmıştır.

INDISC program yazılımının arayüzü ve bu arayüzde kullanılan CLONALG parametreleri, problem parametrelerinin seçimi hakkında bilgiler verilmiştr. Özetlenecek olursa INDISC'in hem teorik hem de pratik çalışma detayları anlatılmıştır. INDISC programının kullanımı örnek uygulama yardımıyla anlatılarak yöntemin ve programın daha iyi anlaşılması sağlatılmaya çalışılmıştır.

6. DENEYSEL SONUÇLAR VE ANALİZ

Bu bölümde önerilen INDISC yöntemi için yazılan programı test etmek amacıyla UCI veri deposundan indirilen çok bilinen 3'ü iki sınıflı ve 5'i çok sınıflı olmak üzere 8 veri kümesi üzerinde uygulamalar yapılmıştır. Bu uygulamaların deneysel sonuç ve analizleri açıklanmıştır.

6.1. İki Sınıflı Veri Setleri Üzerindeki Uygulamalar

INDISC yöntemini uygulamak için iki sınıflı veriler Irvine California Üniversitesi (UCI) makine öğrenmesi veri tabanından elde edilmiştir. Bu veri kümeleri aşağıda verilmektedir.

1. Pima yerlileri diyabet hastalığı (Pima indians diabets) veri kümesi,
2. Wisconsin göğüs kanseri (orijinal) (Wisconsin breast cancer (original)) veri kümesi,
3. Wisconsin göğüs kanseri (teşhis) (Wisconsin breast cancer (diagnostic) veri kümesi.

6.1.1. Pima yerlileri diyabet hastalığı veri kümesi

Birinci deneysel uygulama çalışması için seçilen Machine Learning Repository (UCI)'den alınan Pima yerlileri diyabet (Pima indians diabets) veri kümesidir. (Frank ve Asuncion, 2011). Veri kümesi 2 sınıflı ve sekiz öznitelik olmak üzere toplam 768 örnekleme sahiptir. Bunlar:

1. Gebelik sayısı,
2. Oral glikoz tolerans testinde iki saat sonraki plazma glikoz konsantrasyonu,
3. Diyastolik kan basıncı (mm Hg),
4. Triceps deri kat kalınlığı (mm),
5. 2-Saatlik serum insülin (mu U / ml),
6. Vücut kitle indeksi [ölçüt: kg/boy (m)^2,
7. Diyabet pedigri fonksiyonu,
8. Yaş (yıl).

Sınıflar 0 ve 1 olarak verilmiştir. Sınıf 0 (500 örnek) ve sınıf 1 (268 örnek) sırasıyla test negatif (sınıf 0) ve diyabet için pozitif (sınıf 0) olan hastaları ifade etmektedir. Öznitelikler veri kümesinde farklı aralık değerlerine sahiptir ve bu veri aralıkları Çizelge 6.1'de verilmiştir.

Çizelge 6.1. Pima yerlileri diyabet hastalığı veri kümesinin öznitelik ve aralık değerleri

Öznitelik	Aralık Değerleri
Gebelik Sayısı - (GS)	0 – 17
Oral glikoz tolerans testinde iki saat sonraki plazma glikoz konsantrasyonu - (OGT)	0 – 199
Diyastolik kan basıncı mm Hg - (DKB)	0 – 122
Triceps deri kat kalınlığı mm - (TDKK)	0 – 99
2-Saatlik serum insülin (mu U/ml) - (SSI)	0 – 846
Vücut kitle indeksi [ölçüt: kg/boy(m)^2] - (VKI)	0 – 67.1
Diyabet pedigri fonksiyonu - (DPF)	0.078 – 2.42
Yaş - (YAS)	21 – 81

INDISC yöntemi Pima yerlileri diyabet hastalığı veri kümesine uygulanmıştır. Sonuç olarak, 16 kural (Sınıf 0 için 8 kural ve Sınıf 1 için 8 kural) oluşturulmuştur. Kurallarını görmek için, oluşan vektörler çözümlenmiştir. Çözümlemeden sonra çıkan kurallar Çizelge 6,2'de sunulmuştur.

Çizelge 6.2. Pima yerlileri diyabet hastalığı veri kümesine ait kurallar

Kural Numarası	Kurallar
1.Kural:	**Eğer** GS ϵ (0.1353, 0.2379) & OGT ϵ (0.2005, 0.5068) & DKB ϵ (0.6943, 0.7707) & TDKK ϵ (0.1063, 0.5224) & SSI ϵ (0.457, 0.6587) & VKI ϵ (0.4296, 0.9534) & DPF ϵ (0.6773, 0.9596) & YAS < 0.1762 **O Halde Sınıf 0**
2.Kural:	**Eğer** GS ϵ (0.1055, 0.5757) & OGT ϵ (0.9061, 0.3273) & DKB ϵ (0.509, 0.9433) & TDKK ϵ (0.1954, 0.5186) & SSI ϵ (0.4922, 0.9409) & VKI ϵ (0.2728, 0.2663) & DPF < 0.4021 & YAS < 0.7083 **O Halde Sınıf 0**
3.Kural:	**Eğer** GS ϵ (0.1836, 0.2432) & OGT ϵ (0.4913, 0.2928) & DKB ϵ (0.9119, 0.6754) & TDKK ϵ (0.6061, 0.6406) & SSI ϵ (0.6066, 0.7546) & VKI ϵ (0.2446, 0.4818) & DPF ϵ (0.1693, 0.2837) & YAS < 0.1868 **O Halde Sınıf 0**
4.Kural:	**Eğer** GS ϵ (0.9823, 0.9993) & OGT ϵ (0.3923, 0.2418) & DKB ϵ (0.3013, 0.703) & TDKK ϵ (0.2365, 0.7003) & SSI ϵ (0.133, 0.3582) & VKI ϵ (0.0017, 0.9869) & DPF ϵ (0.3004, 0.5262) & YAS < 0.9808 **O Halde Sınıf 0**
5.Kural:	**Eğer** GS ϵ (0.0885, 0.3556) & OGT ϵ (0.7624, 0.1342) & DKB ϵ (0.8789, 0.5274) & TDKK ϵ (0.1968, 0.6949) & SSI ϵ (0.4328, 0.6143) & VKI ϵ (0.3048, 0.3221) & DPF ϵ (0.2219, 0.5749) & YAS < 0.9786 **O Halde Sınıf 0**
6.Kural:	**Eğer** GS ϵ (0.3599, 0.7113) & OGT ϵ (0.6649, 0.1004) & DKB ϵ (0.6556, 0.2586) & TDKK ϵ (0.3908, 0.884) & SSI ϵ (0.2168, 0.7583) & VKI ϵ (0.4101, 0.2712) & DPF < 0.3988 & YAS < 0.8903 **O Halde Sınıf 0**
7.Kural:	**Eğer** GS ϵ (0.4774, 0.3345) & OGT ϵ (0.2994, 0.4846) & DKB ϵ (0.1956, 0.8924) & TDKK ϵ (0.403, 0.4619) & SSI ϵ (0.5286, 0.6449) & VKI ϵ (0.1837, 0.9334) & DPF ϵ (0.1219, 0.1967) & YAS < 0.9176 **O Halde Sınıf 0**
8.Kural:	**Eğer** GS ϵ (0.1136, 0.6346) & OGT ϵ (0.582, 0.2301) & DKB ϵ (0.7915, 0.4961) & TDKK ϵ (0.2618, 0.5293) & SSI ϵ (0.693, 0.8952) & VKI ϵ (0.3077, 0.2943) & DPF ϵ (0.1453, 0.1268) & YAS < 0.8321 **O Halde Sınıf 0**

9.Kural: **Eğer** GS ϵ (0.7386, 0.7881) & OGT ϵ (0.8952, 0.2771) & DKB ϵ (0.5157, 0.6923) & TDKK ϵ (0.0535, 0.5582) & SSI ϵ (0.627, 0.6513) & VKI ϵ (0.1795, 0.8467) & DPF ϵ (0.5501, 0.9463) & YAS < 0.7183 **O Halde Sınıf 1**

10.Kural: **Eğer** GS ϵ (0.7386, 0.7881) & OGT ϵ (0.8952, 0.2771) & DKB ϵ (0.5157, 0.6923) & TDKK ϵ (0.0535, 0.5582) & SSI ϵ (0.627, 0.6513) & VKI ϵ (0.1795, 0.8467) & DPF ϵ (0.5501, 0.9463) & YAS < 0.7183 **O Halde Sınıf 1**

11.Kural: **Eğer** GS ϵ (0.0527, 0.8259) & OGT ϵ (0.7374, 0.8156) & DKB ϵ (0.386, 0.4366) & TDKK ϵ (0.2599, 0.8688) & SSI ϵ (0.4937, 0.6775) & VKI ϵ (0.8631, 0.4558) & DPF ϵ (0.8275, 0.862) & YAS < 0.8076 **O Halde Sınıf 1**

12.Kural: **Eğer** GS ϵ (0.1079, 0.5681) & OGT ϵ (0.904, 0.2711) & DKB ϵ (0.0032, 0.8595) & TDKK ϵ (0.4371, 0.8655) & SSI ϵ (0.2982, 0.9957) & VKI ϵ (0.6957, 0.3584) & DPF ϵ (0.5522, 0.9344) & YAS < 0.8922 **O Halde Sınıf 1**

13.Kural: **Eğer** GS ϵ (0.8777, 0.706) & OGT ϵ (0.3049, 0.3613) & DKB ϵ (0.1522, 0.9297) & TDKK ϵ (0.1951, 0.2867) & SSI ϵ (0.0182, 0.7545) & VKI ϵ (0.7372, 0.3483) & DPF ϵ (0.6567, 0.6881) & YAS < 0.8657 **O Halde Sınıf 1**

14.Kural: **Eğer** GS ϵ (0.3304, 0.4167) & OGT ϵ (0.9395, 0.4217) & DKB ϵ (0.0191, 0.6525) & TDKK ϵ (0.5463, 0.9747) & SSI ϵ (0.1693, 0.2235) & VKI ϵ (0.5722, 0.1869) & DPF ϵ (0.3728, 0.3631) & YAS < 0.6486 **O Halde Sınıf 1**

15.Kural: **Eğer** GS ϵ (0.2866, 0.6553) & OGT ϵ (0.5778, 0.0681) & DKB ϵ (0.2998, 0.5443) & TDKK ϵ (0.3937, 0.8088) & SSI ϵ (0.5377, 0.8291) & VKI ϵ (0.0293, 0.8864) & DPF ϵ (0.1371, 0.6474) & YAS < 0.8696 **O Halde Sınıf 1**

16.Kural: **Eğer** GS ϵ (0.1855, 0.9405) & OGT ϵ (0.5863, 0.0764) & DKB ϵ (0.2427, 0.5126) & TDKK ϵ (0.1956, 0.5142) & SSI ϵ (0.2622, 0.8577) & VKI ϵ (0.6326, 0.2718) & DPF ϵ (0.1799, 0.9462) & YAS < 0.4614 **O Halde Sınıf 1**

INDISC yöntemi bu veri kümesi için örneklerin % 80.34'nü doğru bir şekilde sınıflandırmıştır. INDISC yönteminin sınıflandırma başarısı ve literatürdeki çalışmaların başarısı Çizelge 6.3'de gösterilmiştir (Anonymous, 2011a). Çizelge 6.3.'den de anlaşılacağı üzere INDISC yöntemi kural çıkarımı için en iyi sonuçlardan birisi olmuştur.

Çizelge 6.3. Pima yerlileri diyabet hastalığı veri kümesine ait deneysel başarılar

Yöntem	Doğruluk%	Referans
INDISC	**80.34**	**Köklü, 2014**
*** Kural çıkarma**	**77.20**	**Köklü ve ark. (2012)**
SSV 5 nodes/BF	75.3±4.8	WD, Ghostminer
SSV opt nodes/3CV/BF	74.7±3.5	WD, Ghostminer
SSV opt prune/3CV/BS	74.6±3.3	WD, Ghostminer
SSV opt prune/3CV/BF	74.0±4.1	WD, Ghostminer
SSV opt nodes/3CV/BS	72.9±4.3	WD, Ghostminer
SSV 5 nodes/BF	74.9±4.8	WD, Ghostminer
SSV 3 nodes/BF	74.6±5.2	WD, Ghostminer
CART	74.5	Stalog
DB-CART	74.4	Shang & Breiman
ASR	74.3	Ster & Dobnikar
CART	72.8	Ster & Dobnikar
C4.5	73.0	Stalog

*** Açıklama:** 2012 yılında yayınlanan makalede bu veri kümesi için toplam 26 kural (Her sınıf için 13 kural) çıkartılmış ve önerilen yöntem veri kümesini % 77.20'sini doğru bir şekilde sınıflandırmıştır. Bu makaleden sonra yönteme budama bölümü eklenerek kural sayısı 16'ya düşürülmüş ve sınıflandırma başarı oranı da %80.34'e çıkartılmıştır.

6.1.2. Wisconsin göğüs kanseri (orijinal) veri kümesi

İkinci deneysel uygulama çalışması için seçilen Machine Learning Repository (UCI)'den alınan Wisconsin göğüs kanseri (orijinal) (Wisconsin breast cancer (original)) veri kümesidir. (Frank ve Asuncion, 2011).

Veri kümesi iki sınıflı ve dokuz özniteliğe sahiptir. Bunlar:

- Parça Kalınlığı,
- Tek düzey Hücre Boyutu,
- Hücre Şekil Benzerliği,
- Marjinal Yapışma,
- Tekli Epitel Hücre Boyutu,
- Sade Çekirdekler,
- Tatsız Kromatin,
- Normal nükleol,
- Mitoz.

Sınıflar 0 ve 1 olarak verilmiştir. Sınıf 0 (458 örnek) ve sınıf 1 (241 örnek), sırasıyla iyi huylu ve kötü huylu kanser tipini belirtmişlerdir. Veri kümesi toplam 699 örneklemden oluşmaktadır. Öznitelikleri veritabanında aynı aralıkta değerlere sahip olup [1-10] aralığında değişmektedir. Çizelge 6.4'de bu veri tabanına ait öznitelikler ve onların aralık değerleri görülmektedir.

Çizelge 6.4. Wisconsin göğüs kanseri (orijinal) veri kümesinin öznitelik ve aralık değerleri

Öznitelik	**Aralık Değerleri**
Parça Kalınlığı (PK)	1 – 10
Tek düzey Hücre Boyutu (TD)	1 – 10
Hücre Şekil Benzerliği (HSB)	1 – 10
Marjinal Yapışma (MY)	1 – 10
Tekli Epitel Hücre Boyutu (TE)	1 – 10
Sade Çekirdekler (SC)	1 – 10
Tatsız Kromatin (TK)	1 – 10
Normal nükleol (NN)	1 – 10
Mitoz (MIT)	1 – 10

INDISC yöntemi Wisconsin göğüs kanseri (orijinal) veri kümesine uygulanmıştır. Sonuç olarak, 14 kural (7 kural sınıf 0 ve 7 kural sınıf 1 için) oluşturulmuştur. Kuralları görmek için, çıkarılan vektörler çözümlenmiştir. Çözümlemeden sonra çıkan kurallar Çizelge 6.5'de sunulmuştur.

Çizelge 6.5. Wisconsin göğüs kanseri (orijinal) veri kümesine ait kurallar

Kural Numarası	Kurallar
1.Kural:	**Eğer** PK < 6.1806 & HSB < 4.9681 & MY < 6.4639 & TE < 9.4427 & SC < 5.0168 **O Halde Sınıf 0**
2.Kural:	**Eğer** PK < 8.6413 & HSB < 9.2499 & MY < 6.4947 & TE < 8.2715 & SC < 7.4848 & TK < 8.1397 & NN < 8.0438 & MIT < 5.4044 **O Halde Sınıf 0**
3.Kural:	**Eğer** PK < 8.8527 & TD < 9.4587 & HSB > 2.1517 & MY < 6.7815 & TK ϵ (1.3024, 4.0324) & NN < 7.958 & MIT < 7.7248 **O Halde Sınıf 0**
4.Kural:	**Eğer** PK < 8.8058 & HSB < 8.5428 & MY < 9.0144 & TE > 2.3297 & SC > 2.6043 & TK ϵ (1.0339, 4.1587) & NN ϵ (1.1963, 9.4871) & MIT < 8.008 **O Halde Sınıf 0**
5.Kural:	**Eğer** PK < 6.4134 & TD < 6.8315 & HSB < 7.5832 & SC < 5.6345 & NN < 7.6222 & MIT < 7.0605 **O Halde Sınıf 0**
6.Kural:	**Eğer** TD < 5.9008 & HSB < 8.2048 & MY < 8.8073 & SC > 5.0516 & TK > 1.2406 & MIT < 9.0814 **O Halde Sınıf 0**
7.Kural:	**Eğer** PK > 5.4016 & TD < 8.0457 & HSB ϵ (2.0608, 4.4028) & SC > 1.717 & NN > 1.6965 & MIT < 4.661 **O Halde Sınıf 0**
8.Kural:	**Eğer** TD > 2.0511 & HSB > 2.5275 **O Halde Sınıf 1**
9.Kural:	**Eğer** PK > 1.0975 & TD < 3.9711 & HSB < 4.1132 & TE < 8.8444 & SC > 2.1484 **O Halde Sınıf 1**
10.Kural:	**Eğer** PK > 3.2977 & TD < 7.4377 & MY < 6.3324 & SC ϵ (1.2796, 8.7596) & TK ϵ (1.0327, 5.3567) **O Halde Sınıf 1**
11.Kural:	**Eğer** TD < 5.9235 & HSB < 6.0508 & MY < 6.7546 & TE < 7.7936 & TK < 6.2862 & NN > 2.7653 & MIT < 5.2392 **O Halde Sınıf 1**
12.Kural:	**Eğer** TD < 7.5291 & MY < 5.6644 & TE < 9.9899 & SC ϵ (1.6207, 5.2861) & TK < 8.9338 & NN < 6.9707 & MIT < 3.2737 **O Halde Sınıf 1**
13.Kural:	**Eğer** PK > 1.9854 & TD > 1.1421 & TK > 1.1352 **O Halde Sınıf 1**
14.Kural:	**Eğer** PK > 6.4043 & HSB ϵ (3.1242, 7.0012) & MY < 7.7544 & TE < 6.8467 & NN < 5.0717 **O Halde Sınıf 1**

INDISC yöntemi bu veri kümesi için örneklerin % 99.12'si doğru bir şekilde sınıflandırmıştır. INDISC yönteminin sınıflandırma başarısı ve literatürdeki çalışmaların başarı Çizelge 6.6'da gösterilmiştir (Anonymous, 2011a). Çizelge 6.6.'dan da görüleceği üzere INDISC yöntemi kural çıkarımı için en iyi sonuçlardan birisi olmuştur.

Çizelge 6.6. Wisconsin göğüs kanseri (orijinal) veri kümesine ait deneysel başarılar

Yöntem	Doğruluk%	Referans
INDISC	**99.12**	**Köklü, 2014**
C-MLP2LN	99.0	
*** Kural çıkarma**	**98.54**	**Köklü ve ark. (2012)**
FSM	98.3	RA
C4.5 (decision tree)	96.0	Hamilton ve ark.
RIAC (prob. inductive)	95.0	Hamilton ve ark.

*** Açıklama:** 2012 yılında yayınlanan makalede bu veri kümesi için toplam 20 kural (Her sınıf için 10 kural) çıkartılmış ve önerilen yöntem veri kümesini % 98.54'ünü doğru bir şekilde sınıflandırmıştır. Bu makaleden sonra yönteme budama bölümü eklenerek kural sayısını 14'e düşürülmüş ve sınıflandırma başarı oranı da %99.12'ye çıkartılmıştır.

6.1.3. Wisconsin göğüs kanseri (teşhis) veri kümesi

Üçüncü deneysel uygulama çalışması için seçilen Machine Learning Repository (UCI)'den alınan Wisconsin göğüs kanseri (teşhis) (Wisconsin breast cancer (diagnostic)) veri kümesidir. (Frank ve Asuncion, 2011).

Veri kümesi aslında temelde 10 özniteliğe sahiptir. Fakat her özniteliğin kendi içinde ortalaması, standart hatası ve en büyüğü (üç büyük değerlerin ortalaması) olmak üzere üç ayrı değere sahiptir bu yüzdendir ki veri kümesindeki öznitelik sayısı 30'a çıkmaktadır. Bunlar:

1. Yarıçap,
2. Doku (gri-ölçek değerleri standart sapma),
3. Çevre,
4. Alan,
5. Pürüzsüzlük (yarıçap uzunluklarındaki bölgesel farklılıklar),
6. Yoğunluk (çevre ^ 2 / alan - 1.0),
7. İçbükeylik (konturun içbükey kısımlarının derecesi),
8. İçbükey noktaları (kontur içbükey bölümlerinin sayısı),
9. Simetri,
10. Fraktal boyut ("kıyı şeridi yaklaşımı" - 1)

Sınıflar 0 ve 1 olarak verilmiştir. Sınıf 0 (357 örnek) ve sınıf 1 (212 örnek), sırasıyla iyi huylu (sınıf 0) ve kötü huylu (sınıf 1) kanser olan hastaları temsil etmektedir. Toplam 569 örneklemden oluşmaktadır. Wisconsin göğüs kanseri (teşhis) veri kümesinin öznitelik ve aralık değerleri Çizelge 6.7'de verilmiştir.

Çizelge 6.7. Wisconsin göğüs kanseri (teşhis) veri kümesinin öznitelik ve aralık değerleri

Öznitelik	Aralık Değerleri
Ortalama Yarıçap (OYC)	6.981 – 28.11
Standart Hatalı Yarıçap (SHYC)	9.71 – 39.28
En Büyük Yarıçap (EBYC)	43.79 – 188.5
Ortalama Doku (OD)	143.5 – 2501
Standart Hatalı Doku(SHD)	0.0526 – 0.1634
En Büyük Doku (EBD)	0.0194 – 0.3454
Ortalama Çevre (OC)	0 – 0.4268
Standart Hatalı Çevre (SHC)	0 – 0.2012
En Büyük Çevre (EBC)	0.1060 – 0.3040
Ortalama Alan (OA)	0.05 – 0.0974
Standart Hatalı Alan (SHA)	0.1115 – 2.8730
En Büyük Alan (EBA)	0.3602 – 4.8850
Ortalama Pürüzsüzlük (ODU	0.7570 – 21.98
Standart Hatalı Pürüzsüzlük (SHP)	6.8020 – 542.2
En Büyük Pürüzsüzlük (EBP)	0.0017 – 0.0311
Ortalama Yoğunluk (OY)	0.0023 – 0.1354
Standart Hatalı Yoğunluk (SHY)	0 – 0.3960
En Büyük Yoğunluk (EBY)	0 – 0.0528
Ortalama İçbükeylik (OIB)	0.0079 – 0.0790
Standart Hatalı İçbükeylik(SHIB)	0 – 0.0298
En Büyük İçbükeylik (EBIB)	7.93 – 36.04
Ortalama İçbükey Noktaları (OIN)	12.02 – 49.54
Standart Hatalı İçbükey Noktaları (SHIO)	50.41 – 251.2
En Büyük İçbükey Noktaları (EBIO)	185.2 – 4254
Ortalama Simetri (OS)	0.0712 – 0.2226
Standart Hatalı Simetri (SHS)	0.0273 – 1.0580
En Büyük Simetri (EBS)	0 – 1.2520
Ortalama Fraktal Boyut (OFB)	0 – 0.2910
Standart Hatalı Fraktal Boyut (SHFB)	0.1565 – 0.6638
En Büyük Fraktal Boyut (EBFB)	0.055 – 0.2075

INDISC yöntemi Wisconsin göğüs kanseri (teşhis) veri kümesine uygulanmıştır. Sonuç olarak, 14 kural (7 kural sınıf 0 ve 7 kural sınıf 1 için) oluşturulmuştur. Kuralları görmek için, oluşan vektörler çözümlenmiştir. Çözümlemeden sonra çıkan kurallar Çizelge 6.8'de sunulmuştur.

Çizelge 6.8. Wisconsin göğüs kanseri (teşhis) veri kümesine ait kurallar

Kural Numarası	Kurallar
1.Kural:	**Eğer** OD < 1999.2256 & SHD < 0.158 & EBD < 0.2667 & OC < 0.315 & OA > 0.051 & SHA < 2.6518 & EBA < 4.8261 & ODU < 19.9975 & SHP < 532.9538 & EBP < 0.0489 & SHY < 0.3442 & EBY < 0.0442 & OIB < 0.0404 & EBIB ϵ (11.5835, 19.4892) & OIN < 39.7186 & SHIO < 211.9052 & OS < 0.1624 & SHS < 1.0043 & EBS < 0.4433 **O Halde Sınıf 0**
2.Kural:	**Eğer** OYC < 17.5986 & EBYC < 179.6769 & SHD < 0.1328 & EBC < 0.2721 & EBA < 3.9408 & ODU < 6.1518 & SHP < 184.5135 & EBP < 0.0449 & OY < 0.1086 & SHY < 0.3365 & OIB < 0.0551 & SHIB < 0.0248 & EBIB < 35.7174 & SHIO ϵ (50.4107, 174.2148) & EBIO < 1730.5859 & OS < 0.1507 & SHS < 0.7125 **O Halde Sınıf 0**
3.Kural:	**Eğer** OYC < 20.8866 & EBYC < 105.7718 & OD > 191.8657 & SHD < 0.1403 & EBD < 0.1867 & SHC < 0.1005 & EBC < 0.2574 & SHA < 2.3672 & EBP < 0.0464 & OY < 0.1158 & SHY < 0.2013 & EBIB ϵ (9.5443, 16.154) & OS < 0.173 & SHFB < 0.6279 **O Halde Sınıf 0**
4.Kural:	**Eğer** OYC < 27.114 & SHYC < 31.3837 & OD < 1269.6683 & SHC < 0.0804 & SHA < 2.8306 & EBA < 3.4095 & EBP < 0.0386 & OY < 0.1077 & EBY < 0.0317 & OIB > 0.0102 & SHIB < 0.0252 & EBIB < 33.7426 & OIN < 46.4441 & SHS < 0.6096 & EBS < 1.2 & OFB < 0.1698 & EBFB < 0.1952 **O Halde Sınıf 0**
5.Kural:	**Eğer** OYC < 26.7424 & OD < 2197.2068 & OC < 0.3763 & SHC < 0.1241 & OA < 0.0844 & SHA < 1.7119 & EBA < 4.2766 & ODU < 11.322 & SHP < 425.7396 & EBP < 0.025 & SHY < 0.3594 & EBY < 0.0495 & OIB < 0.0699 & EBIB < 16.902 & OFB < 0.1556 **O Halde Sınıf 0**
6.Kural:	**Eğer** OYC < 22.4954 & EBC < 0.2295 & EBA < 3.1165 & ODU < 20.3743 & EBP < 0.0367 & EBY < 0.0343 & OIB < 0.0404 & EBIB < 35.2658 & OIN < 46.9219 & SHIO < 121.5319 & SHS < 0.4455 & OFB < 0.2808 & EBFB < 0.1986 **O Halde Sınıf 0**
7.Kural:	**Eğer** OYC < 6.981 & SHYC < 9.71 & EBYC < 43.79 & OD < 143.5 & SHD < 0.0526 & EBD < 0.0194 & OC < 0 & SHC < 0 & EBC < 0.106 & OA < 0.05 & SHA < 0.1115 & EBA < 0.3602 & ODU < 0.757 & SHP < 6.802 & EBP < 0.0017 & OY < 0.0023 & SHY < 0 & EBY < 0 & OIB < 0.0079 & SHIB ϵ (0.0009, 0.0009) & EBIB < 7.93 & OIN < 12.02 & SHIO < 50.41 & EBIO < 185.2 & OS < 0.0712 & SHS < 0.0273 & EBS < 0 & OFB < 0 & SHFB < 0.1565 & EBFB < 0.055 **O Halde Sınıf 0**
8.Kural:	**Eğer** OYC > 13.7628 & EBYC > 64.0516 & OC < 0.3199 & SHC > 0.0161 & EBC < 0.2534 & SHA < 2.6897 & ODU < 20.5539 & SHP < 266.0112 & EBP < 0.0541 & OY < 0.1119 & SHIB < 0.0292 & OIN > 21.8862 & SHIO > 108.3034 & EBIO < 3435.2802 & OFB > 0.0879 & SHFB < 0.6141 **O Halde Sınıf 1**
9.Kural:	**Eğer** OYC ϵ (8.5503, 25.7744) & SHYC > 14.2787 & EBYC < 141.759 & OD < 1644.2557 & EBD < 0.2687 & SHC > 0.0399 & EBC > 0.1188 & EBA > 0.3606 & ODU < 12.988 & SHP < 320.6909 & EBP < 0.0555 & OY < 0.0967 & EBY < 0.0429 & EBIB > 12.4304 & OIN < 37.5319 & EBIO < 3538.766 & SHS < 0.6673 & EBS < 0.9393 & SHFB < 0.4521 & EBFB < 0.191 **O Halde Sınıf 1**
10.Kural:	**Eğer** OYC > 7.7217 & EBD < 0.3431 & OC > 0.0677 & OA < 0.0794 & SHA < 1.6835 & EBA < 3.8098 & ODU < 21.3234 & EBP < 0.0288 & OY < 0.0892 & SHY < 0.1829 & EBY < 0.0501 & OIB < 0.061 & SHIB < 0.0284 & EBIB > 15.4281 & EBIO > 411.0368 & OS > 0.1013 & OFB > 0.0224 **O Halde Sınıf 1**
11.Kural:	**Eğer** SHYC < 35.9635 & OD < 2000.5069 & SHD < 0.1266 & EBD > 0.0325 & OC > 0.039 & EBC > 0.1438 & OA < 0.0899 & EBA < 4.0313 & ODU < 18.637 & SHP > 10.5214 & EBP < 0.026 & OY > 0.005 & SHY < 0.231 & EBY < 0.0405 & OIB < 0.0686 & EBIB > 15.4295 & OIN > 14.3877 & SHIO < 216.302 & EBIO < 3455.7145 & SHS < 1.0273 & EBS < 0.9853 & OFB > 0.0107 & SHFB > 0.1877 & EBFB < 0.1834 **O Halde Sınıf 1**

12.Kural: **Eğer** OYC > 10.7654 & SHYC < 30.6706 & EBYC > 51.2123 & SHD ϵ (0.068, 0.1606) & EBD > 0.0231 & OC > 0.0403 & SHC > 0.053 & ODU < 14.5544 & SHP < 437.3418 & EBP < 0.0416 & SHY < 0.1999 & EBY < 0.0316 & OIB < 0.0717 & SHIB < 0.0266 & EBIB > 13.633 & OIN > 21.1489 & SHIO < 230.2728 & OS > 0.0944 & EBS < 1.2445 & OFB > 0.0729 **O Halde Sınıf 1**

13.Kural: **Eğer** OD < 2319.6609 & EBD < 0.3426 & EBC < 0.2421 & EBA < 4.4866 & SHP < 502.8012 & EBP < 0.0342 & OY < 0.116 & EBIB ϵ (18.0406, 30.0062) & OIN < 32.6381 & EBS ϵ (0.1206, 1.1754) & OFB < 0.266 & EBFB < 0.2047 **O Halde Sınıf 1**

14.Kural: **Eğer** OYC > 12.0613 & SHYC > 10.2982 & EBYC > 71.4105 & OD > 244.7401 & SHD > 0.0659 & SHC > 0.0292 & SHA < 2.7269 & EBA < 4.1367 & EBP < 0.0266 & OIB < 0.0765 & SHIB < 0.0291 & EBIB > 8.6631 & OIN > 15.6995 & EBIO > 702.7465 & OS > 0.0815 & EBS > 0.17 & SHFB < 0.6447 **O Halde Sınıf 1**

INDISC yöntemi bu veri kümesi için örneklerin % 96.31'ni doğru bir şekilde sınıflandırmıştır. Literatürde Wisconsin göğüs kanseri (teşhis) veri kümesi ile ilgili sınıflandırma çalışmaları yapılmış fakat kural çıkarma çalışmalarına Köklü ve ark. (2012) yılında yapmış oldukları çalışmadan başka bir ikinci çalışmaya rastlanamamıştır.

Çizelge 6.9. Wisconsin göğüs kanseri (teşhis) veri kümesine ait deneysel başarılar

Yöntem	Doğruluk%	Referans
INDISC	**96.31**	**Köklü, 2014**
*** Kural çıkarma**	**95.61**	**Köklü ve ark. (2012)**

*** Açıklama:** 2012 yılında yayınlanan makalede bu veri kümesi için toplam 32 kural (Her sınıf için 16 kural) çıkartılmış ve önerilen yöntem veri kümesini % 95.61'ünü doğru bir şekilde sınıflandırmıştır. Bu makaleden sonra yönteme budama bölümü eklenerek kural sayısını 14'e düşürülmüş ve sınıflandırma başarı oranı da %96.31'e çıkartılmıştır.

6.2. Çok Sınıflı Veri Setleri Üzerindeki Uygulamalar

INDISC yöntemini uygulamak için çok sınıflı veriler Irvine California Üniversitesi (UCI) makine öğrenmesi veri deposundan elde edilmiştir. Çok sınıflı kural çıkarımı, için literatürde birçok yöntem önerilmiştir. Ancak tarafımızdan önerilen yöntemin literatürdeki diğer yöntemlerden daha iyi sonuçlar verdiği gözlemlenmiştir.

Uygulaması yapılan çok sınıflı veri kümeleri şunlardır:

1. Deniz kabuğu (Abalone) veri kümesi,
2. Süsen çiçeği (Iris) veri kümesi,
3. Cam kimliklendirme (Glass identification) veri kümesi,
4. Şarap (Wine) veri kümesi,
5. Tiroid hastalığı (Thyroid) veri kümesi,

6.2.1. Deniz kabuğu veri kümesi

Çok sınıflı veri kümelerinden birinci deneysel uygulama çalışması için seçilen Machine Learning Repository (UCI)'den alınan Deniz kabuğu (Abalone) veri kümesidir. Bu veri kümesinde deniz kabuklarının birtakım özellikleri ve yaşı bulunmaktadır. Bir deniz kabuğunun yaşını hesaplamak için kabuk konisi dikey olarak kesildikten sonra boyanarak halkalar sayılmaktadır. Halka sayımı işlemi mikroskop kullanılarak yapılmaktadır. Halkalarının sayısına 1.5 eklendikten sonra deniz kabuğunun yaşı elde edilir. Bu iş pahalı ve zaman alıcı bir işlemdir. Dolaysıyla yaşların belirlenmesi için bilgisayarlı yöntemlerin kullanılmasının çok daha etkili, hızlı ve ucuz olacağı öngörülmektedir (Frank ve Asuncion, 2011).

Deniz kabuğu veri kümesi 8 öznitelik ve 29 sınıftan oluşmaktadır. Bu sınıflar Clark ve arkadaşlarının çalışmalarında sınıfları bazı sınıflardaki verilerin az olması sebebiyle yeniden düzenlemişlerdir (Clark ve ark., 1996). Sınıf 1 (1'den 8'e kadar olan sınıflar), Sınıf 2 (Sınıf 9 ve sınıf 10) ve Sınıf 3 (11'den 29'a kadar olan sınıflar) olarak belirlenmiştir. Veri kümesi 4177 örneklemden oluşmaktadır. Deniz kabuğu veri kümesinin öznitelikleri ve değer aralıkları Çizelge 6.10'da gösterilmiştir.

Çizelge 6.10. Deniz kabuğu veri kümesinin öznitelik ve aralık değerleri

Öznitelik	Aralık Değerleri
Cinsiyet – (Erkek –Dişi – Çocuk)	E, D, C
Uzunluk – (UZ)	0.075 – 0.815
Çap – (CA)	0.055 – 0.650
Yükseklik – (YU)	0 – 1.130
Toplam Ağırlık – (TA)	0.002 – 2.825
Kabuksuz Ağırlık – (KZA)	0.001 –1.488
İç Organ Ağırlığı – (IOA)	0.0005 – 0.760
Kabuk Ağırlığı – (KKA)	0.0015 – 1.005

INDISC yöntemi deniz kabuğu veri kümesine uygulanmıştır. Bunun için önce ilk popülasyon oluşturulurmuştur. Bu veri kümesinde sadece cinsiyet ikili değer (E, D, C), diğer 7 özellik gerçel değerler aldığı için vektörün ikili kısmının uzunluğu 3, gerçel kısımının uzunluğu 14'tür. İlk popülasyonu oluşturmak için cinsiyet özelliğine rasgele bir değer atamak amacıyla ikili kısımdaki değerlerden rastgele birinin değeri 1 yapılmıştır, gerçel kısım "random" fonksiyonu kullanılarak üretilmiştir. Deniz Kabuğu veri kümesi için oluşturulan ilk popülasyondaki vektörlerden biri aşağıdaki gibi oluşmuştur: [1, 0, 0, 0.6551, 0.3112, 0.9027, 0.8116, 0.5211, 0.0424, 0.6663, 0.5985, 0.4624, 0.8699, 0.6834, 0.9296, 0.0426, 0.3478].

Bir sonraki aşamada CLONALG çalıştırılmıştır. Deniz kabuğu veri kümesi için Sınıf 1 için geçerli olmak üzere ilk kural {1, 0, 0, -0.5229, -0.0475, -0.3153, -0.0377, -0.3003, -0.0024, -0.2508, 0.7149, 0.4724, 0.5999, 5.4553, 0.7939, 0.7301, 0.4138} olarak bulunmuştur. İlk üç sayı {1, 0, 0} cinsiyeti belirler, burada yalnızca ilk sayı 1 olduğundan cinsiyet erkek anlamındadır. Elde edilen vektörün 4-10 elemanları öznitelikler için aralıkların orta noktalarını, 11-17 elemanları ise aralıkların genişlemelerini temsil etmektedir. Vektörlerin kodlarının çözülmesiyle "Eğer Cinsiyet=Erkek & UZ ϵ (-0.5229, 0.7149) & CA ϵ (-0.0475, 0.4724) & YU ϵ (-0.3153, 0.5999) & TA ϵ (-0.0377, 5.4553) & KZA ϵ (-0.3003, 0.7939) & IOA ϵ (-0.0024, 0.7301) & KKA ϵ (-0.2508, 0.4138) O halde Sınıf 1" kuralı elde edilir.

Deniz kabuğu veri kümesinde UZ özniteliği [0.075, 0.815] aralığında değerler aldığından ve -0.5229<0.075 olduğundan UZ ϵ (-0.5229, 0.7149) yerine UZ<0.7149 yazılabilir. Benzer bir mantıkla CA<0.4724, YU<0.5999, KZA<0.7939, IOA<0.7301 ve KKA<0.4138 olarak yazılabilir. TA özniteliği [0.002, 2.825] aralığında değerler aldığından ve [0.002, 2.825] $\subset$(-0.0377, 5.4553) olduğundan TA ϵ (-0.039669, 5.457325) ifadesinin bir anlamı kalmamaktadır ve bu koşulun kuraldan kaldırılması başarıyı etkilemeyeceği gibi karar verme hızı artıracaktır. Sonuç olarak kodları çözülmüş kural

kısaca "Eğer Cinsiyet = Erkek & UZ<0.7149& CA<0.4724 & YU<0.5999 & KZA <0.7939 & IOA <0.7301 & KKA<0.4138 O halde Sınıf 1" şeklinde ifade edilir.

Kuralın ilk temsili yanlış değildir ama okunması zordur. Bununla beraber ikincisi daha anlaşılır ve kısadır. Sonuç olarak 54 kural (Her bir sınıf için 18 kural) oluşturulmuştur. Veri kümesinden çıkan kurallar Çizelge 6.11'de verilmiştir.

Çizelge 6.11. Deniz kabuğu veri kümesi kuralları

Kural Numarası	Kurallar
1.Kural:	**Eğer** Cinsiyet = Erkek & UZ < 0.7149 & CA < 0.4724 & YU < 0.5999 & KZA < 0.7939 & IOA < 0.7301 & KKA < 0.4138 **O Halde Sınıf 1**
2.Kural:	**Eğer** UZ < 0.7823 & CA < 0.5504 & TA < 1.9368 & KZA < 1.0222 & KKA < 0.1724 **O Halde Sınıf 1**
3.Kural:	**Eğer** UZ ϵ (0.2009, 0.4836) & CA ϵ (0.0986, 0.568) & YU < 1.031 & KZA < 1.4442 & KKA < 0.8975 **O Halde Sınıf 1**
4.Kural:	**Eğer** Cinsiyet = Erkek & UZ < 0.7178 & TA < 1.0715 & KZA < 0.8098 **O Halde Sınıf 1**
5.Kural:	**Eğer** Cinsiyet ≠ Çocuk & UZ < 0.7669 & CA ϵ (0.0826, 0.4117) & KZA < 0.5791 & IOA < 0.5305 & KKA < 0.7579 **O Halde Sınıf 1**
6.Kural:	**Eğer** KKA < 0.1552 **O Halde Sınıf 1**
7.Kural:	**Eğer** Cinsiyet ≠ Çocuk & UZ < 0.6914 & CA < 0.4393 & KZA < 1.0146 & IOA < 0.7513 & KKA < 0.4447 **O Halde Sınıf 1**
8.Kural:	**Eğer** Cinsiyet = Erkek & IOA < 0.5129 **O Halde Sınıf 1**
9.Kural:	**Eğer** UZ < 0.5638 & TA < 2.5231 & KZA < 1.3484 & IOA < 0.1328 & KKA < 0.2869 **O Halde Sınıf 1**
10.Kural:	**Eğer** Cinsiyet ≠ Çocuk & UZ < 0.5541 & CA < 0.4546 & YU < 1.0671 & IOA < 0.7186 & KKA < 0.4764 **O Halde Sınıf 1**
11.Kural:	**Eğer** Cinsiyet ≠ Çocuk & UZ < 0.6361 & CA < 0.5897 & YU < 0.1465 & KZA < 1.439 & IOA < 0.4764 & KKA < 0.6288 **O Halde Sınıf 1**
12.Kural:	**Eğer** Cinsiyet ≠ Çocuk & CA < 0.4449 & YU < 0.3462 & TA < 2.3676 **O Halde Sınıf 1**
13.Kural:	**Eğer** Cinsiyet = Erkek & UZ < 0.5823 & YU < 0.1951 & KZA < 0.8765 & KKA < 0.9662 **O Halde Sınıf 1**
14.Kural:	**Eğer** Cinsiyet ≠ Çocuk & UZ < 0.7374 & CA < 0.4505 & TA < 0.8042 & KZA < 1.3684 & IOA < 0.6092 **O Halde Sınıf 1**
15.Kural:	**Eğer** Cinsiyet = Erkek & UZ < 0.7186 & TA < 2.5983 & KZA < 1.19 **O Halde Sınıf 1**
16.Kural:	**Eğer** Cinsiyet = Erkek & UZ < 0.667 & CA < 0.5817 & YU < 0.9775 & KZA < 0.4504 & IOA < 0.2581 & KKA < 0.7545 **O Halde Sınıf 1**
17.Kural:	**Eğer** Cinsiyet = Erkek & YU < 0.8745 & KZA < 0.9186 & IOA < 0.5828 **O Halde Sınıf 1**
18.Kural:	**Eğer** Cinsiyet ≠ Çocuk & UZ ϵ (0.1737, 0.611) & CA < 0.5528 & YU < 1.0044 & TA < 0.7018 & KZA ϵ (0.0231, 1.3615) & KKA < 0.2583 **O Halde Sınıf 1**
19.Kural:	**Eğer** UZ < 0.7418 & CA ϵ (0.4252, 0.6395) & YU < 0.9937 & TA ϵ (0.1917, 1.5756) & KZA < 1.0693 & IOA < 0.5329 **O Halde Sınıf 2**
20.Kural:	**Eğer** UZ > 0,0995 & CA < 0.5132 & TA > 0.0119 & KZA > 0.1214 & IOA > 0.1025 & KKA > 0.2038 **O Halde Sınıf 2**
21.Kural:	**Eğer** YU ϵ (0.1269, 0.7841) & TA ϵ (0.5924, 1.5392) & IOA ϵ (0.0037, 0.5099) & KKA > 0.1389 **O Halde Sınıf 2**

22.Kural: **Eğer** UZ < 0.7176 & CA ϵ (0.1031, 0.6449) & YU < 0.9833 & TA > 0.6205 & KZA > 0.3212 & KKA < 0.4938 **O Halde Sınıf 2**

23.Kural: **Eğer** UZ > 0.5257 & CA < 0.609 & TA < 1.5542 & KZA < 0.8464 **O Halde Sınıf 2**

24.Kural: **Eğer** UZ > 0.0903 & CA < 0.5744 & YU < 1.0244 & TA > 0.7811 & KZA < 1.0583 & IOA < 0.5291 & KKA < 0.4641 **O Halde Sınıf 2**

25.Kural: **Eğer** UZ < 0.678 & CA > 0.2832 & KZA < 0.925 & IOA > 0.1582 & KKA < 0.4773 **O Halde Sınıf 2**

26.Kural: **Eğer** UZ ϵ (0.1042, 0.6745) & CA > 0.2107 & YU < 0.5079 & TA ϵ (0.5478, 2.4358) & KKA ϵ (0.0591, 0.495) **O Halde Sınıf 2**

27.Kural: **Eğer** Cinsiyet ≠ Erkek CA < 0.5328 & YU < 0.6775 & TA > 0.3121 & KZA > 0.129 **O Halde Sınıf 2**

28.Kural: **Eğer** Cinsiyet ≠ Erkek TA > 0.1126 & IOA < 0.3769 & KKA < 0.3633 **O Halde Sınıf 2**

29.Kural: **Eğer** UZ > 0.3166 & CA > 0.1894 & TA ϵ (0.5032, 1.5115) & KZA > 0.2828 **O Halde Sınıf 2**

30.Kural: **Eğer** UZ < 0.6535 & YU < 0.6771 & KZA > 0.3283 & KKA > 0.0829 **O Halde Sınıf 2**

31.Kural: **Eğer** UZ < 0.6674 & CA > 0.394 & YU < 1.0766 & TA > 0.5122 & IOA < 0.6783 **O Halde Sınıf 2**

32.Kural: **Eğer** UZ < 0.6622 & CA ϵ (0.2843, 0.5779) & TA ϵ (0.6103, 2.0605) & IOA ϵ (0.1297, 0.6133) & KKA > 0.1635 **O Halde Sınıf 2**

33.Kural: **Eğer** UZ > 0.5458 & CA ϵ (0.2023, 0.5837) & YU < 0.4008 & TA > 0.3758 & KZA < 0.8386 **O Halde Sınıf 2**

34.Kural: **Eğer** UZ < 0.7183 & CA > 0.3838 & TA ϵ (0.2827, 1.9732) & KZA < 1.0793 & IOA ϵ (0.0966, 0.4358) & KKA > 0.1657 **O Halde Sınıf 2**

35.Kural: **Eğer** UZ ϵ (0.3869, 0.7015) & CA < 0.5702 & YU ϵ (0.1191, 0.6609) & TA ϵ (0.6628, 1.6795) & KZA ϵ (0.2437, 1.4214) & IOA ϵ (0.0324, 0.5472) & KKA < 0.7689 **O Halde Sınıf 2**

36.Kural: **Eğer** UZ < 0.6593 & KZA > 0.3077 & KKA > 0.1181 **O Halde Sınıf 2**

37.Kural: **Eğer** CA > 0.0672 & YU > 0.1445 & TA > 0.4614 & KZA > 0.1635 **O Halde Sınıf 3**

38.Kural: **Eğer** Cinsiyet ≠ Erkek CA > 0.2144 & TA > 0.6124 & KZA > 0.148 & KKA < 0.5461 **O Halde Sınıf 3**

39.Kural: **Eğer** YU < 0.9146 & TA ϵ (0.7634, 2.1709) & IOA < 0.5928 **O Halde Sınıf 3**

40.Kural: **Eğer** YU ϵ (0.0804, 0.7074) & TA > 0.2162 & KKA > 0.2484 **O Halde Sınıf 3**

41.Kural: **Eğer** Cinsiyet ≠ Erkek CA > 0.1765 & YU > 0.0749 & KZA < 0.9238 & IOA < 0.6244 & KKA ϵ (0.1975, 0.8315) **O Halde Sınıf 3**

42.Kural: **Eğer** Cinsiyet ≠ Erkek UZ > 0.1355 & CA > 0.3057 & YU ϵ (0.0241, 0.5362) & TA > 0.5772 & IOA > 0.09 **O Halde Sınıf 3**

43.Kural: **Eğer** Cinsiyet ≠ Erkek UZ > 0.1194 & CA > 0.3845 & YU > 0.0975 **O Halde Sınıf 3**

44.Kural: **Eğer** UZ > 0.375 & CA > 0.206 & TA < 2.1859 & IOA < 0.3695 & KKA > 0.211 **O Halde Sınıf 3**

45.Kural: **Eğer** UZ < 0.7926 & CA > 0.3185 & YU < 0.8282 & TA ϵ (0.7328, 2.1619) & IOA < 0.5521 & KKA > 0.0197 **O Halde Sınıf 3**

46.Kural: **Eğer** UZ > 0.354 & CA > 0.0818 & IOA ϵ (0.0887, 0.6156) & KKA > 0.2497 **O Halde Sınıf 3**

47.Kural: **Eğer** Cinsiyet ≠ Erkek UZ > 0.1038 & CA > 0.3485 & YU < 0.808 & IOA > 0.1176 & KKA > 0.0254 **O Halde Sınıf 3**

48.Kural: **Eğer** Cinsiyet ≠ Erkek YU < 0.5167 & IOA ϵ (0.0306, 0.745) & KKA ϵ (0.1465, 0.968) **O Halde Sınıf 3**

49.Kural:	**Eğer** UZ > 0.1241 & CA > 0.1506 & YU < 0.6032 & KZA > 0.0222 & KKA ϵ (0.2201, 0.873) **O Halde Sınıf 3**
50.Kural:	**Eğer** Cinsiyet ≠ Erkek CA > 0.1991 & YU < 0.3639 & KZA ϵ (0.1942, 1.4755) & IOA > 0.1506 **O Halde Sınıf 3**
51.Kural:	**Eğer** Cinsiyet ≠ Erkek UZ > 0.3321 & YU < 0.2929 & TA < 2.0363 & IOA ϵ (0.0926, 0.6798) & KKA > 0.1759 **O Halde Sınıf 3**
52.Kural:	**Eğer** Cinsiyet ≠ Erkek UZ > 0.4696 & CA > 0.1281 & YU > 0.0576 & TA > 0.1366 & KZA < 1.2167 & IOA > 0.0257 **O Halde Sınıf 3**
53.Kural:	**Eğer** UZ > 0.2443 & YU < 0.7353 & IOA < 0.5923 & KKA > 0.2597 **O Halde Sınıf 3**
54.Kural:	**Eğer** CA > 0.1823 & YU > 0.1428 & KZA < 1.1079 & IOA ϵ (0.0757, 0.5776) **O Halde Sınıf 3**

Bu veri kümesinde sınıflandırma başarısı %62,59 olarak elde edilmiştir. Literatürde bu veri kümesi üzerinde herhangi bir kural çıkarım çalışmasına rastlanammadığından dolayı sonuçlar karşılaştırılamamıştır.

6.2.2. Süsen çiçeği veri kümesi

Çok sınıflı veri kümelerinden ikinci uygulama çalışması için seçilen Machine Learning Repository (UCI)'den alınan ve R.A. Fisher'in oluşturduğu Süsen çiçeği (iris) veri kümesidir. Süsen çiçeği veri kümesi 3 sınıflıdır ve 4 özniteliği ölçülmüştür. Bu veri kümesi her sınıftan 50 örnek olmak üzere toplam 150 örneğe sahiptir. Sınıf 1 (Iris Setosa), Sınıf 2 (Iris Versicolor) ve Sınıf 3 (Iris Virginica) olarak sınıflandırılmıştır (Frank ve Asuncion, 2011). Süsen çiçeği veri kümesinin öznitelik ve aralık değerleri Çizelge 6.12'de gösterilmiştir.

Çizelge 6.12. Süsen çiçeği veri kümesinin öznitelik ve aralık değerleri

Öznitelik	Aralık Değerleri
Çanak Yaprağının Uzunluğu - (CYU)	4.3 – 7.9
Çanak Yaprağının Genişliği - (CYG)	2.0 – 4.4
Taç Yaprağının Uzunluğu - (TYU)	1.0 – 6.9
Taç Yaprağının Genişliği - (TYG)	0.1 – 2.5

INDISC yöntemi süsen çiçeği veri kümesine uygulanmıştır. Sonuç olarak, 7 kural (1 kural sınıf 1, 3 kural sınıf 2 ve 3 kural sınıf 3 için) oluşturulmuştur. Kuralları görmek için, oluşan vektörler çözümlenmiştir. Çözümlemeden sonra çıkan kurallar Çizelge 6.13'de verilmiştir.

Çizelge 6.13. Süsen çiçeği veri kümesi kuralları

Kural Numarası	Kurallar
1. Kural	**Eğer** CYG > 2.2961 & TYU < 2.996 & TYG < 1.4778 **O Halde Sınıf 1**
2. Kural	**Eğer** CYG < 4.094 & TYU ϵ (2.9529, 5.5484) & TYG < 1.6001 **O Halde Sınıf 2**
3. Kural	**Eğer** CYG ϵ (2.8232, 3.3694) & TYU ϵ (4.6651, 5.0689) **O Halde Sınıf 2**
4. Kural	**Eğer** CYG ϵ (2.3993, 2.7015) & TYU < 5.2246 & TYG ϵ (0.6250, 1.6500) **O Halde Sınıf 2**
5. Kural	**Eğer** TYU > 4.7428 **O Halde Sınıf 3**
6. Kural	**Eğer** CYU < 7.6554 & CYG < 2.9679 & TYU > 3.3016 & TYG > 1.4345 **O Halde Sınıf 3**
7. Kural	**Eğer** CYU ϵ (5.9947, 6.7277) & CYG < 3.7341 & TYG > 1.7034 **O Halde Sınıf 3**

Bu veri kümesi için örneklerin %100'ü doğru olarak sınıflandırılmıştır. Önerilen yöntemin sınıflandırma sonuçları literatürdeki çalışmalarla karşılaştırılmış ve Çizelge 6.13'de verilmiştir. Çizelge 6.14'de de görüldüğü üzere önerilen yöntem kural çıkarımı için en iyi sonuçlardan birisi olmuştur.

Çizelge 6.14. Süsen çiçeği veri kümesine ait deneysel başarılar

Yöntem	Doğruluk%	Referans
INDISC	**100.00**	**Köklü ve ark. (2014)**
Taco-Miner	100.00	Özbakır ve ark. (2009)
PSO-Miner	100.00	Özbakır ve Delice (2011)
Grobian (rough)	100.00	Browne ve ark. (1998)
PVM 2 Rules	98.00	Weiss ve Kapouleas (1990)
C-MLP2LN	98.00	Duch ve ark. (2001)
SSV	98.00	Duch ve ark. (2001)
PVM 1 Rule	97.30	Weiss ve Kapouleas (1990)
Minerva	97.00	Huysmans ve ark. (2006)
NEFCLASS	96.67	Nauck ve ark. (1996)
C4.5	96.64	Sarkar ve ark. (2012)
CART	96.00	Weiss ve Kapouleas (1990)
IRID3 Rule Base	96.00	Pal ve Chakraborty (2001)
HREM	95.78	Wang ve ark. (2007)
RBC using RS	95.37	Sarkar ve ark. (2012)
RS	95.10	Wang ve ark. (2007)
Ripper	95.00	Huysmans ve ark. (2006)
ADERF	94.67	Castellanno ve ark. (2004)

6.2.3. Cam kimliklendirme veri kümesi

Çok sınıflı veri kümelerinden üçüncü deneysel uygulama çalışması için seçilen Machine Learning Repository (UCI)'den alınan Cam kimliklendirme (glass identification) veri kümesidir (Frank ve Asuncion, 2011). Veri kümesi on niteliğe sahiptir. Bu veri kümesi Diagnostic Products Corporation'dan Vina Spiehler tarafından oluşturulmuş ve kriminolojik inceleme amaçlı kullanılmıştır. Cam kimliklendirme veri

kümesi 9 öznitelik, 7 sınıftan ve 214 örneklemden oluşmaktadır. Fakat 4. sınıfta veri bulunmadığından uygulamada kullanılmamıştır.

Cam tipleri, sınıfları ve örneklem sayıları aşağıdakilerdir (Frank ve Asuncion, 2011):

- Sınıf 1: Bina penceresi yayılarak işlenmiş (70 adet),
- Sınıf 2: Bina penceresi yayılmadan işlenmiş (76 adet),
- Sınıf 3: Araç camı yayılarak işlenmiş (17 adet),
- Sınıf 4: Araç camı yayılmadan işlenmiş (0 adet, veri kümesinde hiç veri yoktur),
- Sınıf 5: Kavanozlar (13 adet),
- Sınıf 6: Sofra takımları (9 adet),
- Sınıf 7: Gece lambaları (29 adet).

Cam kimliklendirme veri kümesinin öznitelik ve aralık değerleri Çizelge 6.15'de gösterilmiştir.

Çizelge 6.15. Cam kimliklendirme veri kümesinin öznitelik ve aralık değerleri

Öznitelik	Aralık Değerleri
Kırılma Katsayısı - (KK)	1.5112 – 1.5339
Sodyum - (NA)	10.73 – 17.38
Magnezyum - (MG)	0 – 4.49
Aliminyum- (AL)	0.29 – 3.5
Silikon - (SI)	69.81 – 75.41
Potasyum - (K)	0 – 6.21
Kalsiyum - (CA)	5.43 – 16.19
Baryum- (BA)	0 – 3.15
Demir - (FE)	0 – 0.51

INDISC yöntemi Cam kimliklendirme veri kümesine uygulanmıştır. Sonuç olarak 78 kural (her bir sınıf için 13 kural, veri örneği olmaması nedeniyle sınıf dört için kural oluşturulmamıştır) oluşturulmuştur. Çıkarılmış kurallar Çizelge 6.16'da verilmiştir.

Çizelge 6.16. Cam kimliklendirme veri kümesi kuralları

Kural Numarası	Kurallar
1.Kural:	**Eğer** KK > 1.5152 & NA < 17.2839 & AL < 1.4478 & SI < 74.8362 & CA < 10.2991 & BA < 0.9945 & FE < 0.2756 **O Halde Sınıf 1**
2.Kural:	**Eğer** KK > 1.5171 & NA < 15.9681 & MG > 2.3951 & AL < 2.55 & SI > 71.1554 & CA > 6.0973 **O Halde Sınıf 1**
3.Kural:	**Eğer** KK > 1.5164 & MG ϵ (2.795**,** 4.4635) & SI > 70.6997 & CA < 10.5309 & BA < 0.0057 **O Halde Sınıf 1**
4.Kural:	**Eğer** KK ϵ (1.5172, 1.5282) & NA > 12.0101 & MG ϵ (1.9917**,** 3.9383) & AL < 3.3553 & SI > 70.7825 & K < 4.0357 & CA > 6.9792 **O Halde Sınıf 1**
5.Kural:	**Eğer** KK ϵ (1.517, 1.5312) & NA > 11.4606 & MG ϵ (2.379, 3.8908) & AL < 2.97 & K < 3.6267 & CA ϵ (6.5907, 12.3804) **O Halde Sınıf 1**
6.Kural:	**Eğer** MG > 2.3768 & AL < 1.3712 & SI > 70.8835 & K < 0.6748 & CA < 11.7204 & BA < 2.6059 & FE < 0.1704 **O Halde Sınıf 1**
7.Kural:	**Eğer** KK ϵ (1.5171, 1.5294) & MG > 2.0914 & AL < 2.6918 & SI > 70.6174 & CA < 16.1552 & FE < 0.3011 **O Halde Sınıf 1**
8.Kural:	**Eğer** KK < 1.5335 & NA ϵ (10.9078, 14.4337) & AL ϵ (0.4145**,** 1.4367) & SI ϵ (70.0345, 74.762) & CA ϵ (8.1821**,** 12.5016) & BA < 1.5294 & FE < 0.5076 **O Halde Sınıf 1**
9.Kural:	**Eğer** KK > 1.5122 & NA > 11.4459 & AL < 1.3541 & SI < 75.1812 & CA < 10.4012 & FE < 0.2376 **O Halde Sınıf 1**
10.Kural:	**Eğer** KK ϵ (1.5169, 1.5306) & NA < 17.2695 & MG > 2.6111 & AL < 3.3301 & SI > 70.9505 & K < 5.8426 & CA ϵ (7.436**,** 13.4227) **O Halde Sınıf 1**
11.Kural:	**Eğer** KK ϵ (1.5171, 1.5284) & MG > 2.3833 & CA < 10.9883 & BA < 1.4667 & FE < 0.2303 **O Halde Sınıf 1**
12.Kural:	**Eğer** KK > 1.5173 & NA < 15.0457 & AL < 1.5397 & SI > 71.1956 & K < 5.7606 & FE < 0.3074 **O Halde Sınıf 1**
13.Kural:	**Eğer** KK > 1.5167 & NA > 11.3557 & MG > 3.2185 & AL < 1.6717 & CA < 14.6315 & BA < 2.7159 **O Halde Sınıf 1**
14.Kural:	**Eğer** KK > 1.5144 & NA > 12.0406 & MG ϵ (1.9662, 4.3776) & AL > 1.0244 & SI < 73.5057 & K < 5.9072 & CA ϵ (7.632, 8.6153) & BA < 1.9393 & FE < 0.3321 **O Halde Sınıf 2**
15.Kural:	**Eğer** KK > 1.512 & NA < 14.6442 & MG > 2.3839 & AL ϵ (0.4976, 3.4992) & SI > 69.8541 & CA < 8.4429 & FE < 0.5053 **O Halde Sınıf 2**
16.Kural:	**Eğer** KK < 1.5174 & MG > 1.1259 & AL ϵ (1.0678, 2.3493) & SI > 70.0028 & CA > 5.7863 & BA < 2.4109 **O Halde Sınıf 2**
17.Kural:	**Eğer** KK ϵ (1.512, 1.5295) & MG > 1.8332 & AL > 1.381 & SI ϵ (71.7957, 73.9955) & BA < 1.953 & FE < 0.3781 **O Halde Sınıf 2**
18.Kural:	**Eğer** KK < 1.5308 & MG > 2.0725 & AL > 1.3281 & SI > 71.6518 & K < 2.9772 & CA > 6.2018 & BA < 1.6834 **O Halde Sınıf 2**
19.Kural:	**Eğer** KK ϵ (1.5135, 1.5306) & MG > 0.3697 & AL ϵ (0.8661, 3.195) & K < 5.4795 & CA < 8.3703 **O Halde Sınıf 2**
20.Kural:	**Eğer** NA ϵ (11.2806, 16.8854) & AL > 1.3535 & SI ϵ (71.8001, 73.3818) & K > 0.2119 & CA < 11.3464 & BA < 2.2567 & FE < 0.3426 **O Halde Sınıf 2**
21.Kural:	**Eğer** KK < 1.5245 & NA > 11.8143 & MG ϵ (0.748, 3.7195) & AL > 1.3544 & CA > 6.8797 **O Halde Sınıf 2**
22.Kural:	**Eğer** NA > 11.6614 & MG > 1.0869 & AL ϵ (1.4013**,** 3.2281) & K < 1.5431 & CA ϵ (6.7388**,** 15.2741) & FE < 0.3743 **O Halde Sınıf 2**
23.Kural:	**Eğer** NA > 11.8126 & MG > 1.0803 & AL ϵ (1.3892**,** 3.3773) & SI > 71.143 & K > 0.1197 & BA < 2.4997 **O Halde Sınıf 2**
24.Kural:	**Eğer** KK < 1.5267 & MG ϵ (2.6068**,** 4.454) & AL ϵ (0.7627**,** 2.1485) & SI < 74.8503 & K < 4.9227 & CA ϵ (6.8625**,** 8.3302) & BA < 2.9714 & FE < 0.3681 **O Halde Sınıf 2**

25.Kural: **Eğer** NA > 11.1702 & MG > 2.0658 & AL > 1.4077 & SI < 75.4081 & K < 1.8774 & CA ϵ (6.7698, 9.9588) & BA < 2.6124 & FE < 0.3964 **O Halde Sınıf 2**

26.Kural: **Eğer** KK ϵ (1.5134, 1.5283) & MG ϵ (2.087, 4.4198) & AL > 1.3641 & SI > 70.1009 & K < 5.8163 & CA ϵ (5.7796, 10.4039) & FE < 0.403 **O Halde Sınıf 2**

27.Kural: **Eğer** KK ϵ (1.5134, 1.5195) & NA < 17.1445 & MG ϵ (2.8642, 4.4483) & SI ϵ (71.8428, 73.0494) & K < 2.7148 & CA < 9.0695 & BA < 3.1004 **O Halde Sınıf 3**

28.Kural: **Eğer** NA < 14.9026 & MG > 0.8775 & SI ϵ (71.1526, 72.9138) & K < 4.2289 & CA > 8.2201 & BA < 0.6282 **O Halde Sınıf 3**

29.Kural: **Eğer** NA > 12.9729 & MG > 2.9093 & AL < 1.8782 & SI > 70.228 & K < 5.1781 & CA > 8.0018 & BA < 2.4751 & FE < 0.4894 **O Halde Sınıf 3**

30.Kural: **Eğer** NA > 12.2225 & MG ϵ (2.8522, 4.3624) & AL > 0.3021 & SI > 70.8418 & K < 4.8332 & CA > 8.5518 **O Halde Sınıf 3**

31.Kural: **Eğer** KK < 1.5226 & NA > 12.004 & MG > 2.4782 & CA ϵ (8.256, 13.2918) & BA < 1.4923 **O Halde Sınıf 3**

32.Kural: **Eğer** KK < 1.5221 & NA ϵ (13.0183, 15.6869) & MG > 2.2087 & AL ϵ (0.421, 2.8034) & CA > 7.6248 & FE < 0.4238 **O Halde Sınıf 3**

33.Kural: **Eğer** KK < 1.5321 & NA ϵ (13.104, 16.1802) & MG ϵ (1.867, 3.7719) & K < 5.9783 & CA ϵ (8.2461, 11.6077) & FE < 0.3497 **O Halde Sınıf 3**

34.Kural: **Eğer** KK < 1.5262 & MG > 2.7129 & AL < 2.1182 & SI < 72.9602 & CA > 7.2446 & BA < 2.3585 **O Halde Sınıf 3**

35.Kural: **Eğer** NA > 12.9951 & MG ϵ (2.1071, 3.9341) & AL < 3.0827 & K < 4.7201 & CA ϵ (8.2915, 9.1686) **O Halde Sınıf 3**

36.Kural: **Eğer** NA < 15.5723 & MG ϵ (1.6098, 4.0734) & AL < 3.0955 & SI < 72.7181 & K < 5.0675 **O Halde Sınıf 3**

37.Kural: **Eğer** MG ϵ (3.2729, 4.2178) & SI < 74.1633 & K < 5.7021 & CA ϵ (8.2849, 13.9681) **O Halde Sınıf 3**

38.Kural: **Eğer** KK < 1.5288 & NA ϵ (13.2134, 15.7918) & MG > 2.599 & K < 2.6112 & CA ϵ (8.1402, 11.4138) & BA < 3.0266 & FE < 0.4895 **O Halde Sınıf 3**

39.Kural: **Eğer** KK < 1.5247 & MG > 2.7304 & AL > 0.4061 & SI ϵ (71.115, 73.0358) & K < 5.3095 & CA ϵ (8.0502, 15.9095) **O Halde Sınıf 3**

40.Kural: **Eğer** KK < 1.5332 & MG < 2.5352 & AL ϵ (1.2216, 3.1499) & SI > 70.1034 & CA < 13.1166 & BA < 0.1974 **O Halde Sınıf 5**

41.Kural: **Eğer** KK < 1.5297 & NA < 16.5299 & MG < 2.0387 & CA < 13.1572 **O Halde Sınıf 5**

42.Kural: **Eğer** KK < 1.5266 & NA < 14.3178 & MG < 3.2502 & K > 0.2713 & FE < 0.1994 **O Halde Sınıf 5**

43.Kural: **Eğer** MG < 2.2895 & AL ϵ (1.3348, 3.1985) & SI > 71.2013 & K < 5.6726 & CA > 10.0457 **O Halde Sınıf 5**

44.Kural: **Eğer** NA ϵ (11.0111, 13.6024) & MG < 2.3385 & AL ϵ (1.0377, 3.2977) & SI > 71.2786 & K < 5.7992 & CA ϵ (9.2987, 14.6755) **O Halde Sınıf 5**

45.Kural: **Eğer** NA < 14.489 & MG < 2.3984 **O Halde Sınıf 5**

46.Kural: **Eğer** KK < 1.5246 & NA < 15.5903 & MG < 2.7838 & AL > 1.2038 & FE < 0.3121 **O Halde Sınıf 5**

47.Kural: **Eğer** KK < 1.5313 & NA < 16.9726 & MG < 2.6289 & K > 0.2173 & CA > 6.8553 & BA < 2.4131 **O Halde Sınıf 5**

48.Kural: **Eğer** KK < 1.5256 & MG < 3.1124 & AL > 0.7101 & SI < 74.0246 & CA < 15.3724 **O Halde Sınıf 5**

49.Kural: **Eğer** NA < 17.1539 & MG < 2.0765 & AL > 0.8698 & SI < 75.1403 & BA < 1.9311 **O Halde Sınıf 5**

50.Kural: **Eğer** KK > 1.5157 & MG < 3.6623 & AL > 1.3512 & SI > 71.6543 & K < 5.697 & CA > 9.5866 & BA < 1.4547 **O Halde Sınıf 5**

51.Kural: **Eğer** KK > 1.5123 & MG < 1.9197 & AL ϵ (1.1305, 3.1465) & SI ϵ (70.1657, 74.1361) & K > 0.0594 & BA < 2.3797 **O Halde Sınıf 5**

52.Kural: **Eğer** AL > 0.7664 & SI > 71.5262 & K < 1.2321 & CA > 9.27 **O Halde Sınıf 5**

53.Kural: **Eğer** NA > 13.3564 & MG < 2.9461 & SI > 71.8245 & K < 5.9972 & BA < 1.3884 & FE < 0.3049 **O Halde Sınıf 6**

54.Kural: **Eğer** KK < 1.5205 & NA > 13.1579 & MG < 3.1222 & AL < 2.4158 & SI > 71.0933 & K < 5.6578 & FE < 0.4248 **O Halde Sınıf 6**

55.Kural: **Eğer** KK < 1.5207 & NA > 11.399 & MG < 2.8712 & AL < 3.4525 & K < 5.3486 & CA < 13.7772 & BA < 2.8677 **O Halde Sınıf 6**

56.Kural: **Eğer** KK < 1.5219 & MG < 2.6392 & AL < 2.4507 & SI > 72.3259 & K < 4.9088 & CA < 12.8678 & FE < 0.2177 **O Halde Sınıf 6**

57.Kural: **Eğer** KK < 1.5338 & MG < 2.794 & AL < 2.357 & CA < 11.2578 & FE < 0.3129 **O Halde Sınıf 6**

58.Kural: **Eğer** KK < 1.5232 & NA > 13.5064 & MG < 3.6274 & BA < 1.1381 **O Halde Sınıf 6**

59.Kural: **Eğer** KK < 1.5268 & MG < 3.2936 & AL ϵ (0.3091, 2.0991) & K < 1.0466 & BA < 0.6254 **O Halde Sınıf 6**

60.Kural: **Eğer** KK < 1.5304 & MG < 3.3254 & CA < 11.394 & BA < 1.3916 & FE < 0.2274 **O Halde Sınıf 6**

61.Kural: **Eğer** NA > 13.5886 & MG < 3.4791 & AL < 3.3321 & SI > 70.0657 & K < 3.6402 & CA < 13.4964 **O Halde Sınıf 6**

62.Kural: **Eğer** KK < 1.5316 & MG < 2.8225 & AL < 2.3298 & SI > 72.2969 & CA < 13.6019 & BA < 0.6157 **O Halde Sınıf 6**

63.Kural: **Eğer** MG < 2.9445 & AL ϵ (0.299, 2.5111) & SI > 71.4827 & K < 2.703 & CA < 13.8037 & FE < 0.2656 **O Halde Sınıf 6**

64.Kural: **Eğer** KK < 1.5199 & MG < 2.4757 & K < 4.5207 & BA < 1.133 **O Halde Sınıf 6**

65.Kural: **Eğer** KK < 1.5246 & NA > 10.9186 & MG < 2.9483 & AL < 2.6814 & SI > 70.0072 & K < 3.9752 & CA > 6.6347 & FE < 0.3243 **O Halde Sınıf 6**

66.Kural: **Eğer** KK < 1.5305 & NA > 11.2075 & AL ϵ (1.0141, 3.2659) & BA > 0.1908 & FE < 0.1854 **O Halde Sınıf 7**

67.Kural: **Eğer** NA < 15.8686 & AL > 0.5414 & SI > 70.2815 & K < 3.8942 & BA > 0.3739 & FE < 0.3821 **O Halde Sınıf 7**

68.Kural: **Eğer** NA > 11.1302 & SI < 73.7961 & K < 2.6427 & CA < 14.6022 & BA ϵ (0.2206, 3.0033) **O Halde Sınıf 7**

69.Kural: **Eğer** KK < 1.5302 & MG < 3.393 & AL ϵ (1.0516, 3.4638) & CA < 15.5513 & BA ϵ (0.3693, 2.9236) & FE < 0.2741 **O Halde Sınıf 7**

70.Kural: **Eğer** NA ϵ (10.9912, 17.0915) & AL > 0.8179 & SI > 70.1573 & BA > 0.2043 **O Halde Sınıf 7**

71.Kural: **Eğer** KK < 1.5286 & NA > 12.856 & CA < 14.287 & BA ϵ (0.3624, 2.8908) & FE < 0.3033 **O Halde Sınıf 7**

72.Kural: **Eğer** KK < 1.5286 & NA ϵ (13.1876, 17.0313) & CA < 15.5821 & BA > 0.1104 & FE < 0.197 **O Halde Sınıf 7**

73.Kural: **Eğer** KK < 1.5296 & NA < 16.7419 & SI > 70.2138 & K < 5.9555 & BA > 0.3112 **O Halde Sınıf 7**

74.Kural: **Eğer** NA > 12.0292 & MG < 4.2269 & AL < 3.3399 & BA > 0.2966 & FE < 0.359 **O Halde Sınıf 7**

75.Kural: **Eğer** NA > 12.3401 & CA < 11.9261 & BA ϵ (0.3329, 2.0407) & FE < 0.3551 **O Halde Sınıf 7**

76.Kural: **Eğer** KK < 1.5309 & NA ϵ (11.9952, 16.2091) & MG < 3.7057 & AL > 1.7242 & SI > 70.6598 & CA < 15.582 & BA > 0.0035 **O Halde Sınıf 7**

77.Kural: **Eğer** NA > 11.2538 & AL > 1.1628 & K < 4.7081 & CA < 13.1134 & BA ϵ (0.3601, 3.0609) **O Halde Sınıf 7**

78.Kural: **Eğer** KK < 1.5298 & AL < 3.2222 & K < 4.3309 & CA < 13.0742 & BA > 0.2869 & FE < 0.495 **O Halde Sınıf 7**

Bu veri kümesi için örneklerin %77.10'u doğru olarak sınıflandırılmıştır. Önerilen yöntemin sınıflandırma sonuçları literatürdeki çalışmalarla karşılaştırılmış ve Çizelge 6.17'de verilmiştir. Çizelge 6.17'den de görüldüğü üzere önerilen yöntem kural çıkarımı için en iyi sonucu vermiştir.

Çizelge 6.17. Cam kimliklendirme veri kümesine ait deneysel başarılar

Yöntem	Doğruluk%	Referans
INDISC	**77.10**	**Köklü ve ark. (2014)**
C4.5	72.85	Sarkar ve ark. (2012)
RBC using RS	71.87	Sarkar ve ark. (2012)
NN	69.94	Sarkar ve ark. (2012)
HREM	63.89	Wang ve ark. (2007)
RS	60.42	Wang ve ark. (2007)
FRBC	58.28	Kim ve ark. (2005)
IRID3 Rule Base	52.70	Pal ve Chakraborty (2001)
ADERF	50.95	Castellanno ve ark. (2004)
GGP	35.20	Pappa ve Freitas (2009)

6.2.4. Şarap veri kümesi

Çok sınıflı veri kümelerinden dördüncü deneysel uygulama çalışması için seçilen Machine Learning Repository (UCI)'den alınan şarap (wine) veri kümesidir. Bu veri kümesinin on üç özniteliği vardır. Veri kümesi Stefan Aeberhard, M. Forina ve arkadaşları tarafından oluşturulmuştur. Bu veriler aslında İtalya'nın aynı yöresinde yetişen üzümlerden yapılmış şarapların kimyasal analizlerinin sonuçlarıdır ve üç farklı yetiştiriciden elde edilen verilerdir. Analizlerde üç tip şarap için bulunan 13 bileşenin miktarı belirtilmiştir (Frank ve Asuncion, 2011). Sınıflar; sınıf 1, sınıf 2 ve sınıf 3 olarak verilmiştir. Veri kümesi 178 örneğe sahiptir. Şarap veri kümesinin öznitelik ve aralık değerleri Çizelge 6.18'de gösterilmiştir.

Çizelge 6.18. Şarap veri kümesinin öznitelik ve aralık değerleri

Öznitelik	Aralık Değerleri
Alkol - (AL)	11.03 – 14.83
Malik Asit - (MA)	0.74 – 5.80
Ash - (AS)	1.36 – 3.23
Bazik Ash - (BAS)	10.60 – 30.00
Magnezyum - (MG)	70.00 – 162.00
Toplam Phenols – (TP)	0.98 – 3.88
Flavanoids - (FL)	0.34 – 5.08
Nonflavanoid phenols - (NP)	0.13 – 0.66
Proanthocyanins - (PRT)	0.41 – 3.58
Renk Yoğunluğu – (RY)	1.28 – 13.00
Renk - (RE)	0.48 – 1.71
OD280/OD315 ile Sulandırılmış - (ODS)	1.27 – 4.00
Proline - (PRL)	278.00 – 1680.00

Önerilen yöntem bu verilere uygulanmış ve sonuç olarak 15 kural (Her sınıf için 5 kural) oluşturulmuştur. Bu veri kümesinden çıkarılmış olan kurallar Çizelge 6.19'de verilmiştir.

Çizelge 6.19. Şarap veri kümesi kuralları

Kural Numarası	Kurallar
1.Kural:	**Eğer** AL > 11.7181 & MA > 1.9503 & FL > 2.0453 & RY ϵ (3.3683, 9.7618) & RE ϵ (0.784, 1.5591) **O Halde Sınıf 1**
2.Kural:	**Eğer** AL > 12.3359 & MA < 5.7084 & BAS < 21.8451 & MG > 76.3762 & FL > 0.7239 & NP < 0.6055 & PRT < 2.9912 & RY < 9.5682 & RE > 0.7964 & ODS > 2.1572 & PRL > 685.9059 **O Halde Sınıf 1**
3.Kural:	**Eğer** MA > 1.0893 & MA < 3.0286 & BAS < 22.1934 & TPH > 1.8561 & NP < 0.4366 & RY > 3.5638 & RE > 0.7532 & ODS > 1.8097 **O Halde Sınıf 1**
4.Kural:	**Eğer** AL > 11.2136 & MA > 1.4562 & TPH > 2.2345 & FL < 4.6903 & NP > 0.1423 & PRT < 2.9679 & RE < 1.3462 & ODS > 2.3532 & PRL > 612.7642 **O Halde Sınıf 1**
5.Kural:	**Eğer** AL ϵ (13.0361, 14.7619) & MA > 1.9564 & FL < 3.9603 & PRT > 1.1112 & RY ϵ (2.6155, 10.6656) & RE > 0.5406 & ODS > 2.4263 **O Halde Sınıf 1**
6.Kural:	**Eğer** AL < 13.0541 & BAS > 12.8089 & MG < 158.7653 & FL > 0.7 & RY < 9.5164 & RE > 0.5674 & ODS > 1.6852 **O Halde Sınıf 2**
7.Kural:	**Eğer** BAS < 28.5601 & TPH > 1.0261 & FL < 2.6819 & RE ϵ (0.7921, 1.6072) & ODS > 1.4223 & PRL < 1016.9881 **O Halde Sınıf 2**
8.Kural:	**Eğer** AL < 14.7273 & MA < 5.7162 & MA < 2.6026 & BAS < 24.5526 & MG < 161.1681 & TPH < 3.8209 & NP < 0.6236 & RY < 6.8685 & ODS > 1.9803 & PRL < 900.1877 **O Halde Sınıf 2**
9.Kural:	**Eğer** AL < 14.2958 & MA < 4.7857 & MG < 154.3325 & PRT < 2.8322 & RY < 7.5158 & RE > 0.4807 & ODS > 1.6179 & PRL < 685.0241 **O Halde Sınıf 2**
10.Kural:	**Eğer** AL < 13.131 & MA < 5.3626 & MA > 1.6998 & BAS > 11.1109 & FL ϵ (0.9147, 4.4531) & NP > 0.1837 & RE > 0.6127 **O Halde Sınıf 2**
11.Kural:	**Eğer** AL ϵ (12.0835, 14.7932) & MG ϵ (74.4699, 138.1886) & TPH < 2.26 & FL < 3.8919 & ODS < 3.2486 **O Halde Sınıf 3**
12.Kural:	**Eğer** MA > 1.1863 & MA > 2.0526 & MG < 142.4527 & TPH < 3.8601 & FL < 2.4885 & PRT > 0.522 & RY > 3.6158 & ODS < 3.3188 **O Halde Sınıf 3**
13.Kural:	**Eğer** BAS > 16.4458 & MG ϵ (75.3395, 150.1264) & FL < 3.1378 & PRT < 2.4837 & RE < 0.9441 **O Halde Sınıf 3**
14.Kural:	**Eğer** AL > 12.1524 & MA > 1.2505 & BAS > 17.3026 & MG < 147.4015 & TPH < 3.5894 & FL < 2.9736 & RY ϵ (3.6477, 11.7838) & PRL ϵ (457.9776, 937.6957) **O Halde Sınıf 3**
15.Kural:	**Eğer** BAS ϵ (12.5736, 27.5178) & NP > 0.2037 & PRT > 0.4512 & RY > 4.7667 & ODS < 2.722 **O Halde Sınıf 3**

Bu veri kümesindeki örneklerin %99.44'u doğru olarak sınıflandırılmıştır. Önerilen yöntemin sınıflandırma sonuçları literatürdeki çalışmalarla karşılaştırılmış ve Çizelge 6.20'de verilmiştir. Çizelge 6.20'den de görüldüğü üzere, INDISC yöntemi kural çıkarımı için en iyi sonucu vermiştir.

Çizelge 6.20. Şarap veri kümesine ait deneysel başarılar

Yöntem	Doğruluk%	Referans
INDISC	**99.44**	**Köklü ve ark. (2014)**
C4.5	97.12	Sarkar ve ark. (2012)
RISE	96.80	Rijnbeek ve Kors (2010)
RBC using RS	96.78	Sarkar ve ark. (2012)
PDGFR	95.18	Nojima ve ark. (2009)
EXPLORE	94.10	Rijnbeek ve Kors (2010)
PART	94.00	Rijnbeek ve Kors (2010)
C4.5R	93.70	Rijnbeek ve Kors (2010)
CN2	93.30	Rijnbeek ve Kors (2010)
SL^2	92.60	Rijnbeek ve Kors (2010)
RIPPER	92.10	Rijnbeek ve Kors (2010)
CBA	86.70	Rijnbeek ve Kors (2010)
IR	86.40	Rijnbeek ve Kors (2010)
CDF	86.10	Fu ve Wang (2005)

6.2.5. Tiroid hastalığı veri kümesi

Çok sınıflı veri kümelerinden beşinci ve sonuncu deneysel uygulama çalışması için seçilen Machine Learning Repository (UCI)'den alınan Tiroid hastalığı (Newthyroid) veri kümesidir. Bu veri kümesi beş öznitelikten ve üç sınıftan oluşmaktadır. Sınıflar; sınıf 1 (normal), sınıf 2 (yüksek) ve sınıf 3 (düşük) olarak verilmiştir. 150 adet Sınıf 1, 35 adet Sınıf 2 ve 30 adet Sınıf 3 olmak üzere toplam 215 örneğe sahiptir (Frank ve Asuncion, 2011).

- Nitelik_1: T3-resin alım testi (yüzdesi).
- Nitelik_2: Izotopik değiştirme yöntemi ile ölçülen toplam serum tiroksin değeri,
- Nitelik_3: Radyoimüno testi ile ölçülen serum triodothyronine değeri,
- Nitelik_4: Radyoimüno testi ile ölçülen bazal tiroid uyarıcı hormon değeri (TSH),
- Nitelik_5: TSH değerinin maksimum mutlak farkın tirotropin salgıyan hormonun 200 mikro gram enjeksiyonundan sonraki bazal değeri.

Tiroid hastalığı veri kümesinin öznitelik ve aralık değerleri Çizelge 6.21'de gösterilmiştir.

Çizelge 6.21. Tiroid hastalığı veri kümesinin öznitelik ve aralık değerleri

Öznitelik	Aralık Değerleri
Nitelik_1	65 – 144
Nitelik_2	0.5 – 25.3
Nitelik_3	0.2 – 10
Nitelik_4	0.1 – 56.4
Nitelik_5	-0.7 – 56.3

Önerilen yöntem bu verilere uygulanmış ve sonuç olarak 3 kural (Her sınıf için 1 kural) oluşturulmuştur. Bu veri kümesinden çıkarılmış olan kurallar Çizelge 6.22'de verilmiştir.

Çizelge 6.22. Tiroid hastalığı veri kümesi kuralları

Kural Numarası	Kurallar
1.Kural:	**Eğer** Nitelik_1 ϵ (89.1864, 137.8899) & Nitelik_2 ϵ (5.2947, 14.3904) & Nitelik_3 < 9.0771 & Nitelik_4 < 97.3198 & Nitelik_5 < 17.8507 **O Halde Sınıf1**
2.Kural:	**Eğer** Nitelik_2 ϵ (12.3151, 28.272) & Nitelik_4 < 63.7349 & Nitelik_5 < 46.6076 **O Halde Sınıf2**
3.Kural:	**Eğer** Nitelik_1 > 71.2129 & Nitelik_2 < 6.7686 & Nitelik_4 < 59.6451 & Nitelik_5 ϵ (-0.65, 63.0713) **O Halde Sınıf3**

Bu veri kümesindeki örneklerin %93.95'i doğru olarak sınıflandırılmıştır. Önerilen yöntemin sınıflandırma sonuçları literatürdeki çalışmalarla karşılaştırılmış ve Çizelge 6.23'de verilmiştir. Çizelge 6.23'de de görüldüğü üzere, INDISC yöntemi kural çıkarımı sınıflandırma için en iyi sonucu vermiştir.

Çizelge 6.23. Tiroid hastalığı veri kümesine ait deneysel başarılar

Yöntem	Doğruluk%	Referans
INDISC	**93.95**	**Köklü, (2014)**
*** Kural çıkarma**	**93.95**	**Köklü ve ark. (2013)**
FH-GBML-IVFS-Amp	93.49	Sanz ve ark. (2010)
RBC Using RS	91.33	Sarkar ve ark. (2012)
GAGRAD	87.91	Sanz ve ark. (2010) aktardığına göre (Dombi ve Gera) (2008)
Chi-IVFS-Amp	85.12	Sanz ve ark. (2010) aktardığına göre (Chi ve ark.) (1996)
Chi	84.65	Sanz ve ark. (2010) aktardığına göre (Chi ve ark.) (1996)

*** Açıklama:** 2013 yılında uluslararası sempozyumda sunulan yayında bu veri kümesi için toplam 9 kural (her sınıf için 3 kural) çıkartılmış ve önerilen yöntem veri kümesini % 93.95'ini doğru bir şekilde sınıflandırmıştır. Daha sonra yönteme budama bölümü eklenerek kural sayısını 3'e düşürülmüş ve sınıflandırma başarı oranında da bir değişiklik olmamıştır.

6.3. Bölüm Değerlendirmesi

Bu çalışmadaki iki sınıflı ve çok sınıflı olmak üzere 8 farklı veri kümesi kullanılmıştır İki sınıflı verilerle yapılan pima yerlileri diyabet hastalığı %80.34, Wisconsin göğüs kanseri (orijinal) %99.12, Wisconsin göğüs kanseri (teşhis) %96.31 doğruluk başarısı ile sınıflandırma yapılmıştır. Litarütürde iki sınıflı wisconsin göğüs kanseri (teşhis) veri kümesi ile ilgili sınıflandırma çalışmaları yapılmış fakat kural çıkarma çalışmalarına bizden başka bir ikinci çalışmaya rastlanamamıştır.

Çok sınıflı kural çıkarımı, deniz kabuğu %62.59, Süsen çiçeği %100, Cam kimliklendirme %77.10, Şarap %99.44 ve Tiroid hastalığı %93.95 doğruluk başarısı ile sınıflandırma yapılmıştır Litarütürde çok sınıflı deniz kabuğu veri kümesi ile ilgili sınıflandırma çalışmaları yapılmış fakat kural çıkarma çalışmalarına bizden başka bir ikinci çalışmaya rastlanamamıştır.

Çok sınıflı kural çıkarımı, için literatürde birçok yöntem önerilmiştir. Ancak çalışmada önerilen yöntem kullanıldığında literatürdeki yöntemlerden daha iyi başarı oranları elde edildiği gözlemlenmiştir.

7. SONUÇ VE ÖNERİLER

7.1. Sonuçlar

Sınıflandırma kuralları keşfi veri madenciliğinde önemli bir alandır. Literatürde birçok kural çıkarım çalışması vardır. Bu kural çıkarım çalışmalarının çoğunluğu YSA veya SVM tabanlı çalışmalardır. Bu çalışmalarda önce bir yöntem kullanılarak sınıflandırma yapılmış, sonra oluşan model optimize edilerek kurallar üretilmiştir. Son yıllarda ara bir sınıflayıcı olmadan direkt veriden kural çıkarma yöntemleri gündeme gelmiştir. Bu nedenle bu çalışmada kural keşfi verilerden yapılmıştır.

Çalışmadaki ikinci ve en önemli fark kodlama yöntemidir. Bugüne kadarki çalışmalarda öznitelikler bazen bulanık yöntemlerle, çoğunlukla ise ikili olarak kodlanmıştır. Bunun için öznitelik programcı tarafından belirlenen sayıda aralıklara bölünerek tamamen sübjektif olarak kodlanmıştır. Veri madenciliği alanında kullanılan verilerin tıp, finans, tüm mühendislik alanları, medya vb. kaynaklı olması nedeniyle bir uzmanın bütün bu alanlara hakim olmayacağına göre bu tür kodlamanın en doğru sonuca götüremeyeceği aşikardır.

Bu nedenle bu çalışmada aralıkların tespiti optimizasyon yöntemiyle yapılmıştır. Bu alt aralıklar optimizasyon aracılığıyla belirlendiği için, daha etkin şekilde oluşturulmaktadır. Böylece bu yöntem her öznitelik için aralıkları optimizasyon kullanarak hassasiyetle belirlemektedir.

Çalışmada ayrık ve gerçel değerli öznitelikler farklı yöntemlerle kodlanmış olup, sınıflandırma başarısı uygunluk fonksiyonu kabul edilerek optimizasyon uygulanmıştır. Optimizasyon sonucu elde edilen vektörlerin her birinin birer kural belirttiği kabul edilmiştir. Bir sonraki aşamada oluşan bu kurallar budamaya tabi tutulup, sayısı en aza

indirgenmiştir. Son olarak ise kuralların kodu çözülmüş, veri tabanındaki ilk haline dönüştürülmüştür. Önerilen bu yönteme isim olarak "Aralık Keşfi" anlamına gelen INDISC (INterval DISCovery) verilmiştir. INDISC aşağıdaki adımlardan oluşur;

- Özniteliklerin ayrık ve gerçel değerli olarak ayrılması
- Gerçel değerli özelliklerin [0,1] aralığında normalize edilmesi.
- Ayrık özelliklerin ikili kodlanması
- Popülasyonun rastgele oluşturulması
- Sınıflandırılma kurallarının optimize edilerek çıkarılması
- Oluşturulan kuralların budanması
- Elde edilen sınıflandırma kuralların kodunun çözülmesi

Optimizasyon amacıyla CLONALG algoritması kullanılmıştır. Sunulmuş yaklaşım, sekiz gerçek durum sınıflandırma problemine uygulanmıştır. Veriler Irvine California Üniversitesi (UCI) makine öğrenmesi veri deposudan sitesinden alınmıştır (Frank ve Asuncion, 2011). INDISC yöntemi, Pima yerlileri diyabet hastalığı %80.34, Wisconsin göğüs kanseri (orijinal) %99.12, Wisconsin göğüs kanseri (teşhis) %96.31, Deniz kabuğu %62.59, Süsen çiçeği %100, Cam kimliklendirme %77.10, Şarap %99.44 ve Tiroid hastalığı %93.95 doğruluk başarısı ile sınıflandırma yapılmıştır Önceki araştırmalarla karşılaştırıldığında, önerilen metot ile daha başarılı sonuçlar elde edildiği görülmüştür.

Bu çalışmadaki problemlerde iki sınıflı verilerin yanı sıra çok sınıflı veri kümeleri de kullanılmıştır. Çok sınıflı kural çıkarımı, için literatürde birçok yöntem önerilmiştir. Ancak tarafımızdan önerilen yöntem literatürdeki yöntemlerden daha iyi sonuçlar verdiği gözlemlenmiştir.

Sistemin başarısını test etmek için sekiz farklı veri tabanında uygulanarak elde edilen sonuçlar literatürde rapor edilen sonuçlarla karşılaştırılmıştır. Genel olarak bakıldığında doğru tanı oranı yüksek olan bir sistem geliştirildiği görülmektedir. Çalışmada alınan önemli sonuçlardan bazıları şunlardır:

1. Verinin çoğaldığı zamanımızda verinin bilgiye dönüştürülmesi çok önemli olduğu göz önünde bulundurulduğunda, bu tip bir sınıflandırma sisteminin önemi daha iyi anlaşılmaktadır.
2. Başarılı sınıflandırma yapmak için kural keşfi amaçlayan bu çalışmada, yeni bir kodlama yöntemi önerilmiştir. Kural çıkarmak için bir yapay bağışıklık algoritması olan CLONALG kullanılarak optimizasyonu yapılmaktadır.
3. Kodlama yönteminin sınıflandırma başarısına etkisi gösterilmiştir.
4. Başarı ölçümlerinin güvenilirliğinin sağlanması amacıyla sekiz farklı veritabanı kullanılmıştır. Sekiz veritabanı da araştırmalara açık olduğu için üzerlerinde birçok uygulama yapılmaktadır. Buda bize çalışmamızın sonucunu karşılaştırma olanağı sunmaktadır.

Bu tez çalışmasında elde edilen sonuçlardan SCI expanded kapsamında taranan dergilerde iki makale ve bir uluslararası konferans bildirisi kabul edilmiş ve yayınlanmıştır. (Köklü ve ark., 2012; Köklü ve ark., 2013; Köklü ve ark., 2014).

7.2. Öneriler

Elde edilen araştırma sonuçları ışığında, teklif edilen kodlama yönteminin sınıflandırma başarısını yükselttiği gözlenmiştir. Ayrık ve gerçel verilerin farklı yöntemlerle kodlanmasının sonuçlar üzerinde etkili olduğu görülmüştür. Araştırmada farklı kodlama yöntemleri ve optimizasyon algoritmaları denenmiştir. Optimizasyon yöntemine oranla kodlama yönteminin başarı üzerinde etkisinin daha fazla olduğu gözlenmiştir. Ama %100 başarı oranı yakalanamamıştır. Bu nedenle bu konu halen üzerinde araştırma ve çalışmalar yapılması gereken popüler bir alan olmaya devam etmektedir. Veri kümesinin yapısını modelleyen özniteliklerin incelenmesi ve farklı

makine öğrenmesi yöntemlerinin çok değişik şekillerde kullanımları sonucunda gelecekte daha başarılı sınıflandırma yöntemlerinin geliştirilebileceği öngörülmektedir.

EKLER

EK-1 Süsen Çiçeği (İris) Veritabanı Dosyası

5,1	3,5	1,4	0,2	1
4,9	3,0	1,4	0,2	1
4,7	3,2	1,3	0,2	1
4,6	3,1	1,5	0,2	1
5,0	3,6	1,4	0,2	1
5,4	3,9	1,7	0,4	1
4,6	3,4	1,4	0,3	1
5,0	3,4	1,5	0,2	1
4,4	2,9	1,4	0,2	1
4,9	3,1	1,5	0,1	1
5,4	3,7	1,5	0,2	1
4,8	3,4	1,6	0,2	1
4,8	3,0	1,4	0,1	1
4,3	3,0	1,1	0,1	1
5,8	4,0	1,2	0,2	1
5,7	4,4	1,5	0,4	1
5,4	3,9	1,3	0,4	1
5,1	3,5	1,4	0,3	1
5,7	3,8	1,7	0,3	1
5,1	3,8	1,5	0,3	1
5,4	3,4	1,7	0,2	1
5,1	3,7	1,5	0,4	1
4,6	3,6	1,0	0,2	1

5,1	3,3	1,7	0,5	1
4,8	3,4	1,9	0,2	1
5,0	3,0	1,6	0,2	1
5,0	3,4	1,6	0,4	1
5,2	3,5	1,5	0,2	1
5,2	3,4	1,4	0,2	1
4,7	3,2	1,6	0,2	1
4,8	3,1	1,6	0,2	1
5,4	3,4	1,5	0,4	1
5,2	4,1	1,5	0,1	1
5,5	4,2	1,4	0,2	1
4,9	3,1	1,5	0,1	1
5,0	3,2	1,2	0,2	1
5,5	3,5	1,3	0,2	1
4,9	3,1	1,5	0,1	1
4,4	3,0	1,3	0,2	1
5,1	3,4	1,5	0,2	1
5,0	3,5	1,3	0,3	1
4,5	2,3	1,3	0,3	1
4,4	3,2	1,3	0,2	1
5,0	3,5	1,6	0,6	1
5,1	3,8	1,9	0,4	1
4,8	3,0	1,4	0,3	1
5,1	3,8	1,6	0,2	1
4,6	3,2	1,4	0,2	1
5,3	3,7	1,5	0,2	1

5,0	3,3	1,4	0,2	1
7,0	3,2	4,7	1,4	2
6,4	3,2	4,5	1,5	2
6,9	3,1	4,9	1,5	2
5,5	2,3	4,0	1,3	2
6,5	2,8	4,6	1,5	2
5,7	2,8	4,5	1,3	2
6,3	3,3	4,7	1,6	2
4,9	2,4	3,3	1,0	2
6,6	2,9	4,6	1,3	2
5,2	2,7	3,9	1,4	2
5,0	2,0	3,5	1,0	2
5,9	3,0	4,2	1,5	2
6,0	2,2	4,0	1,0	2
6,1	2,9	4,7	1,4	2
5,6	2,9	3,6	1,3	2
6,7	3,1	4,4	1,4	2
5,6	3,0	4,5	1,5	2
5,8	2,7	4,1	1,0	2
6,2	2,2	4,5	1,5	2
5,6	2,5	3,9	1,1	2
5,9	3,2	4,8	1,8	2
6,1	2,8	4,0	1,3	2
6,3	2,5	4,9	1,5	2
6,1	2,8	4,7	1,2	2
6,4	2,9	4,3	1,3	2

6,6	3,0	4,4	1,4	2
6,8	2,8	4,8	1,4	2
6,7	3,0	5,0	1,7	2
6,0	2,9	4,5	1,5	2
5,7	2,6	3,5	1,0	2
5,5	2,4	3,8	1,1	2
5,5	2,4	3,7	1,0	2
5,8	2,7	3,9	1,2	2
6,0	2,7	5,1	1,6	2
5,4	3,0	4,5	1,5	2
6,0	3,4	4,5	1,6	2
6,7	3,1	4,7	1,5	2
6,3	2,3	4,4	1,3	2
5,6	3,0	4,1	1,3	2
5,5	2,5	4,0	1,3	2
5,5	2,6	4,4	1,2	2
6,1	3,0	4,6	1,4	2
5,8	2,6	4,0	1,2	2
5,0	2,3	3,3	1,0	2
5,6	2,7	4,2	1,3	2
5,7	3,0	4,2	1,2	2
5,7	2,9	4,2	1,3	2
6,2	2,9	4,3	1,3	2
5,1	2,5	3,0	1,1	2
5,7	2,8	4,1	1,3	2
6,3	3,3	6,0	2,5	3

5,8	2,7	5,1	1,9	3
7,1	3,0	5,9	2,1	3
6,3	2,9	5,6	1,8	3
6,5	3,0	5,8	2,2	3
7,6	3,0	6,6	2,1	3
4,9	2,5	4,5	1,7	3
7,3	2,9	6,3	1,8	3
6,7	2,5	5,8	1,8	3
7,2	3,6	6,1	2,5	3
6,5	3,2	5,1	2,0	3
6,4	2,7	5,3	1,9	3
6,8	3,0	5,5	2,1	3
5,7	2,5	5,0	2,0	3
5,8	2,8	5,1	2,4	3
6,4	3,2	5,3	2,3	3
6,5	3,0	5,5	1,8	3
7,7	3,8	6,7	2,2	3
7,7	2,6	6,9	2,3	3
6,0	2,2	5,0	1,5	3
6,9	3,2	5,7	2,3	3
5,6	2,8	4,9	2,0	3
7,7	2,8	6,7	2,0	3
6,3	2,7	4,9	1,8	3
6,7	3,3	5,7	2,1	3
7,2	3,2	6,0	1,8	3
6,2	2,8	4,8	1,8	3

6,1	3,0	4,9	1,8	3
6,4	2,8	5,6	2,1	3
7,2	3,0	5,8	1,6	3
7,4	2,8	6,1	1,9	3
7,9	3,8	6,4	2,0	3
6,4	2,8	5,6	2,2	3
6,3	2,8	5,1	1,5	3
6,1	2,6	5,6	1,4	3
7,7	3,0	6,1	2,3	3
6,3	3,4	5,6	2,4	3
6,4	3,1	5,5	1,8	3
6,0	3,0	4,8	1,8	3
6,9	3,1	5,4	2,1	3
6,7	3,1	5,6	2,4	3
6,9	3,1	5,1	2,3	3
5,8	2,7	5,1	1,9	3
6,8	3,2	5,9	2,3	3
6,7	3,3	5,7	2,5	3
6,7	3,0	5,2	2,3	3
6,3	2,5	5,0	1,9	3
6,5	3,0	5,2	2,0	3
6,2	3,4	5,4	2,3	3
5,9	3,0	5,1	1,8	3

EK-2 Süsen Çiçeği (İris) Veri Kümesinin Öznitelik ve Aralık Değerler Dosyası

CYU 4.3 7.9

CYG 2.0 4.4

TYU 1.0 6.9

TYG 0.1 2.5

Açıklama:

CYU Çanak Yaprağının Uzunluğu,

CYG Çanak Yaprağının Genişliği,

TYU Taç Yaprağının Uzunluğu,

TYG Taç Yaprağının Genişliği

anlamına gelmektedir.

KAYNAKLAR

Abbas, Abul K., Lichtman, Andrew H. and Pober, Jordan S., 1994, Celluar and Molecular Immunology, Eds. W. B. Sounders Com., Fourth edition A27A52A1:A25A1:A32A1:A37A52A1:A25A1:A40A52A1:A25A1:A42A1:A41

Akgöbek, Ö. ve Öztemel, E., 2006, Endüktif Öğrenme Algoritmalarının Kural Üretme Yöntemleri ve Performanslarının Karşılaştırılması, SAÜ Fen Bilimleri Enstitüsü Dergisi, 10 (1), 1-9.

Akşit F., Akgün Y. And Kiraz N., 1996, Mikrobiyoloji, Açıköğretim Fakültesi Yayınları.

Aliev R.A., Aliev R.R., Guirimov B. and Uyar K., 2008, Dynamic Data Mining Technique for Rules Extraction in a Process of Battery Charging, Applied Soft Computing, 8, 1252–1258.

Andrews, R., Diederich, J. and Tickle A. B., 1995, A Survey and Critique of Techniques for extracting rules from trained artificial neural networks, Knowledge-Based Systems, 8 (6), 373-389.

Ang J.H., Tan K.C. and Mamun A.A., 2010, An Evolutionary Memetic Algorithm For Rule Extraction, Expert Systems with Applications, 37 (2), 1302-1315.

Anonymous, 2011a Logical rules extracted from data, http://www.is.umk.pl/projects/rules.html, [Ziyaret Tarihi: Haziran 2011].

Avşar G., 2007, Kaba Kümeler Yardımıyla Önemsiz ve Kayıp Türdeki Eksik Verilerden Bulanık Kuralların Çıkarılması, Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, Elâzığ.

Aydoğan, F., 2003, E-ticarette Veri Madenciliği Yaklaşımlarıyla Müşteriye Hizmet Sunan Akıllı Modüllerin Tasarımı ve Gerçekleştirimi, Yüksek Lisans Tezi , Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Baysal, B. 2001, Temel İmmünoloji Ders Notları, Selçuk Üniversitesi Tıp Fakültesi Mikrobiyoloji Anabilim Dalı.

Bioch, J. C., O. van der Meer, et al., 1996, Classification using Bayesian neural nets, The 1996 IEEE International Conference on Neural Networks, Institute of Electrical and Electronics Engineers, Washington, DC, 1488-1493.

Browne C., Duntsch I. and Gediga G., 1998, IRIS revisited: A comparison of discriminant and enhanced rough set data analysis, Rough Sets in Knowledge Discovery 2, Physica-Verlag Heidelberg, 19, 345-368.

Brownlee, J., 2005, Clonal Selection Theory & CLONALG the Clonal Selection Classification Algorithm (CSCA), Tecnical Report.

Carpenter, G. A., N. Markuzon, 1998, ARTMAP-IC and medical diagnosis: instance counting and inconsistent cases, Neural Networks, 11 (2), 323-336.

Castellano G., Fanelli A. M., and Mencar C., 2004, An empirical risk functional to improve learning in a neuro-fuzzy Classifier. IEEE Trans. Syst., Man, Cybern. B. 34, 725-731.

Cheung, N., 2001, Machine learning techniques for medical analysis, BsC Thesis, School of Information Technology and Electrical Engineering, University of Queenland.

Chi Z., Yan H. and Pham T., 1996, Fuzzy Algorithms with Applications to Image Processing and Pattern Recognition, World Scientific.

Chu Kiong Loo, 2005, Accurate and Reliable Diagnosis and Classification Using Probabilistic Ensemble Simplified Fuzzy ARTMAP, IEEE Transactions On Knowledge And Data Engineering, 17 (11), 1589-1593.

Clark D., Schreter Z. and Adams A., 1996, A Quantitative Comparison of Dystal and Backpropagation, Submitted to the Australian Conference on Neural Networks.

Coenen, F., Leng, P., 2007, The Effect of Threshold Values on Association Rule Based Classification Accuracy, Data & Knowledge Engineering, 60, 345-360.

Craven, M.W., Shavlik, J.W., 1996, Extracting tree-structured representations of trained networks, Advances in Neural Information Processing Systems, 8, MIT Press, Cambridge, MA, 24–30.

Çiftçi M.T., 2011, Büyük Boyutlu Sınıflandırma Problemlerin Çözümü İçin Yeni Bir Matematiksel Yaklaşımları, Yüksek Lisans Tezi, Anadolu Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, Eskişehir.

Das, S., Mozer, M., 1998, Dynamic on-line clustering and state extraction:an approach to symbolic learning, Neural Networks, 11 (1), 53–64.

De Castro L. N. and Timmis, J., 2002, Artificial Immune Systems: A New Computational Intelligence Approach, Springer, London.

De Castro, L. N. and Von Zuben, F. J., 1999, Artificial Immune Systems: Part I- Basic Theory and Aplications, Technical Report.

De Castro, L. N. and Von Zuben, F. J., 2000, The Clonal Selection Algorithm with Engineering Applications, GECCO'00-Workshop Proceedings, 36-37.

De Castro, L. N., Von Zuben, F. J. and De Deus Jr., G. A., 2000, The construction of a Boolean competitive neural network using ideas from immunology, Neurocomputing - IJON, 50, 51-85.

De Falco, I., Della Cioppa, A., Tarantino, E., 2007, Facing Classification Problems with Particle Swarm Optimization, Applied Soft Computing, 7 (3), 652-658.

Dehuri, S., Patnaik, S., Ghosh, A., Mall, R., 2008, Application of Elitist Multi-Objective Genetic Algorithm for Classification Rule Generation, Applied Soft Computing, 8, 477-487.

Delice Y., 2008, Parçacık Sürü Optimizasyonu ile Yapay Sinir Ağlarından Sınıflandırma Kuralı Çıkarımı, Yüksek Lisans Tezi, Erciyes Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı, Kayseri.

Do Wan Kim, JinBaePark, and Young Hoon Joo, 2005, Design of Fuzzy Rule-Based Classifier: Pruning and Learning, Fuzzy Systems and Knowledge Discovery, Lecture Notes in Computer Science, 3613, 416-425.

Dombi J., Gera Z., 2008, Rule based fuzzy classification using squashing functions, Journal of Intelligent & Fuzzy Systems, 19 (1), 3–8.

Dorado, J., Rabunal, J. R., Rivero, D., Santos, A. Pazos, A., 2002, Automatic Recurrent ANN Rule Extraction with Genetic Programming, Proceedings of The 2002 International Joint Conference on Neural Networks-IJCNN'02, 1552-1557.

Duch W., Adamczak R., Grąbczewski K., 2001, A new methodology of extraction, optimization and application of crisp and fuzzy logical rules. IEEE Transactions on Neural Networks, 12, 277-306.

Duch, W., Adamczak, R. and Grabczewski, K., 1998, Extraction of logical rules from training data using backpropagation networks, Neural Process., 7, 1–9.

Duch, W., Adamczak, R., Grabczewski, K. and Zal, G., 1999, Hybrid neuralglobal minimization method of logical rule extraction, J. Adv. Comput. Intell., 3, 348–356.

Duman E., Akın E., 2006, Yapay Bağışıklık Sistemi ile Bulanık Yaklaştırıcı Kural Tabanı Optimizasyonu, ELECO06, Bursa.

Eklund, P. W. and A. Hoang, 1998, Classifier Selection and Training Set Features: LMDT, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.34.9095&rep=rep1&type=pdf , [Ziyaret Tarihi: Haziran 2014].

Elalfi, E., Haque, R., Elalami, M. E., 2004, Extracting Rules From Trained Neural Network Using GA For Managing E-business, Applied Soft Computing, 4, 65-77.

Forrest, S., A. Perelson, Allen, L. And Cherukuri, R., 1994, Self-Nonself Discrimination in a Computer, Proc. of the IEEE Symposium on Research in Security and Privacy, 202-212.

Frank, A. and Asuncion, A., 2011, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, [Erişim Tarihi: February 2011].

Ful X. and Wang L., 2005, A Rule Extraction System with Class-Dependent Features, Studies in Fuzziness and Soft Computing, DOI: 10.1007/3-540-32358-9_5, 163/2005, 79-99.

Gallant, S. I., 1993, Neural Network Learning and Expert Systems. MIT Press, Cambridge, MA.

Garcez, A.S.D., Broda, K. and Gabbay, D.M., 2001, Symbolic knowledge extraction from trained neural networks: a sound approach, Artif. Intell., 125, 155–207.

Han J. and Kamber M., 2001, Data mining: Concepts and techniques, Morgan Kaufmann Publishers, San Francisco.

Hruschka, E. R., Ebecken, F. F., 2006, Extracting Rules From Multilayer Perceptrons in Classification Problems: A Clustering-Based Approach, 7th Brazilian Symposium on Neural Networks-SBRN '04, 70, 384-397.

Huysmans J., Baesens B. and Vanthienen J, 2006, Using Rule Extraction to Improve the Comprehensibility of Predictive Models. Open Access publications from Katholieke Universiteit Leuven with number urn:hdl:123456789/121060, KBI-0612.

Jerne, N. K., 1973, The Immune System, Scientific American, 229 (1), 52-60.

Jo H., Han I., and Lee H., 1997, Bankruptcy prediction using case-based reasoning, neural networks and discriminant analysis, Expert Syst. Appl., 13, 97–108.

Kahramanlı H., 2008, Hibrit Bulanık Sinir Ağını Kullanarak Bir Sınıflandırma ve Kural Çıkarma Sisteminin Geliştirilmesi, Doktora Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü Matematik Anabilim Dalı, Konya.

Kahramanli H. and Allahverdi N., 2009a, Extracting rules for classification problems: AIS based approach, Expert Systems with Applications 36, 10494–10502.

Kahramanli H. and Allahverdi N., 2009b, Rule extraction from trained adaptive neural networks using artificial immune systems, Expert Systems with Applications 36, 1513–1522.

Karabatak M., 2008, Özellik Seçimi, Sınıflama ve Öngörü Uygulamalarına Yönelik Birliktelik Kuralı Çıkarımı ve Yazılım Geliştirilmesi, Doktora Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik Mühendisliği Anabilim Dalı, Elâzığ.

Karaboga D. and Akay B., 2009, A comparative study of Artificial Bee Colony algorithm, Applied Mathematics and Computation 214, 108–132.

Keedwell, E., Narayanan, A. and Savic, D., 2000, Creating rules from trained neural networks using genetic algorithms, Int. J. Comput. Syst. Signals, 1 (1), 30–42.

Kınacı A.C., 2013, Yapay Sinir Ağlar Üzerinde Kural Çıkarma İşlemi İçin Bağlantın Melez Bir Zeki Sistem Geliştirilmesi, Doktora Tezi, Ege Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, İzmir.

Kim D.W., Park J.B., and Joo Y.H., 2005, Design of Fuzzy Rule-Based Classifier: Pruning and Learning, Fuzzy Systems and Knowledge Discovery Lecture Notes in Computer Science, 3613, 416–425, Springer.

Köklü M., Kahramanlı H., ve Allahverdi N., 2012, A New Approach to Classification Rule Extraction Problem by the Real Value Coding, International Journal of Innovative Computing, Information and Control (IJICIC), 8 (9), 6303-6315.

Köklü M., Kahramanlı H., ve Allahverdi N., 2013, Classification Rule Extraction Using Artificial Immune System: Clonalg and its Application, 4th International Conference on Mathematical and Computational Applications (ICMCA-2013), Celal Bayar University, 11-13 June 2013, 283-292, Manisa.

Köklü M., Kahramanlı H., ve Allahverdi N., 2014, Sınıflandırma Kurallarının Çıkarımı İçin Etkin ve Hassas Yeni Bir Yaklaşım, Journal of the Faculty of Engineering and Architecture of Gazi University, 29 (3), 477-486.

Kulluk S., 2009, Karınca Koloni Optimizasyonu İle Yapay Sinir Ağlarından Kural Çıkarımı, Doktora Tezi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü Makine Mühendisliği Anabilim Dalı, Kayseri.

Kuttiyil, A. S., 2004, Survey of Rule Extraction Methods, Thesis, Wayne State University, Detroit.

Markowska-Kaczmar, U., 2005, The Influence of Prameters in Evolutionary Based Rule Extraction Method From Neural Network, Proceedings of the 2005 5th International Conference on Intelligent Systems Design and Applications-ISDA'05, 106 – 111.

Leandro N. De Castro and Fernando J. Von Zuben, 2002, Learning and Optimization Using the Clonal Selection Principle, IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems, 6, 239-251.

Lee, Y. J., and Mangasarian, O. L., 2001a, RSVM: Reduced support vector machines, In Proceedings of the first SIAM international conference on data mining.

Lee, Y. J., and Mangasarian, O. L., 2001b, SSVM: A smooth support vector machine for classification, Computational Optimization and Applications, 20 (1), 5–22.

Liu, B., 1998, Integrating Classification and Association Rule Mining. KDD-98, Knowledge Discovery and Data Minng, New York, 80-86.

Loo Chu K., 2005, Accurate and Reliable Diagnosis and Classification Using Probabilistic Ensemble Simplified Fuzzy ARTMAP, IEEE Transactions On Knowledge And Data Engineering, 17 (11), 1589-1593.

Lu, H., Setiono, R. and Liu, H., 1996, Effective data mining using neural networks, IEEE Trans Knowledge Data Eng, 8 (6), 957–961.

Michie, D., D. J. Spiegelhalter, et al., 1994, Machine learning, neural and statistical classification. New York, Ellis Horwood.

Mitra, S. and Acharya, T., 2003, Data Mining: Multimedia, Soft Computing and Bioinformatics, Wiley-Interscience, Inc. Hoboken, New Jersey.

Nauck D., Nauck U. and Kruse R., 1996, Generating Classification Rules with the Neuro-Fuzzy System NEFCLASS., Proc. Biennial Conf. of the North American Fuzzy Information Processing Society-NAFIPS'96, Berkeley.

Nojima Y., Ishibuchi H. and Kuwajima I., 2009, Parallel Distributed Genetic Fuzzy Rule Selection (FDGFRS), Soft Computing, 13, 511–519.

Oates, T., 1994, MSDD as a Tool for Classification. EKSL Memorandum 94-29, Department of Computer Science, University of Massachusetts at Amherst.

Olmeda, I. And Fernandez, E., 1997, Hybrid classifiers for financial multicriteria decision making: the case of bankruptcy prediction, Comput. Econ., 10, 317–335.

Otero, F. E., Freitas, A. A., Johnson, C. G., 2008, cAnt-Miner: An Ant Colony Classification Algorithm to Cope with Continuous Attributes, Proceedings of the 6th International Conference on Ant Colony Optimization And Swarm Intelligence, Brussels, Belgium, 5217, 48-59.

Özbakır, L., Baykasoğlu, A., Kulluk, S., Yapıcı, 2009, H., TACO-miner: An Ant Colony Based Algorithm for Rule Extraction from Trained Neural Networks, Expert Systems with Applications, 36, 12295-12305.

Özbakır L., Baykasoglu A., Kulluk S., 2010, A soft computing-based approach for integrated training and rule extraction from artificial neural networks: DIFACONN-miner, Applied Soft Computing, 10, 304–317.

Özbakır, L., Delice,Y., 2011, Exploring comprehensible classification rules from trained neural Networks integrated with a time-varying binary particle swarm optimizer, Engineering Applications of Artificial Intelligence, 24, 491-500.

Özbal, Y., 1994, Temel İmmunoloji, Nobel Tıp Kitabevleri Tic. LTD. Şti, 1. Baskı

Pal N. R. and Chakraborty S., 2001, Fuzzy Rule Extraction From ID3-Type Decision Trees for Real Data, IEEE Trans. Syst., Man, Cybern., 31, 745-754.

Palade, V., Neagu, D. and Puscasu, G., 2000, Rule extraction from neural networks by interval propagation, in: Proceedings of the Fourthh IEEE International Conference on Knowledge-Based Intelligent Engineering Systems, Brighton-UK, 217–220.

Papageorgiou E.I., 2011, A New Methodology for Decisions in Medical Informatics Using Fuzzy Cognitive Maps Based on Fuzzy Rule-Extraction Techniques, Applied Soft Computing, 11, 500–513.

Pappa G.L. and Freitas A. A., 2009, Evolving Rule Induction Algorithms with Multi-Objective Grammar-Based Genetic Programming, Knowledge and Information Systems, DOI 10.1007/s10115-008-0171-1, 19, 283–309.

Parpinelli, R.S., Lopes, H.S., Freitas, A.A., 2001, An Ant Colony Based System For Data Mining: Applications To Medical Data, Proceedings of the Genetic and Evolutionary Computation Conference, San Francisco, California, 791–798.

Patterson, G., Zhang, M., 2007, Fittness Functions in Genetic Programming for Classification with Unbalanced Data, Springer Berlin Heidelberg, 4830, 769-775.

Perelson, A. S. & Oster, G. F., 1979, Theoretical Studies of Clonal Selection: Minimal Antibody Repertuarie Size and Reliability of Self-Nonself Discrimination, J. Theor. Biol., 81, 645-670.

Pham, D. T., Dimov, S. S., and Salem, Z., 2000, Technique for selecting examples in inductive learning, In European symposium on intelligent techniques, Aachen-Germany, 119–127.

Quinlan, J. R., 1983, Learning Efficient Classification Procedures and Their Application to Chess End Games, Machine Learning - An Artificial Intelligence Approach, Springer Berlin Heidelberg, 463-482.

Quinlan, J. R., 1993, C4.5: Programs for Machine Learning, Morgan Kaufmann, San Francisco, CA, USA.

Rijnbeek P.R. and Kors J.A., 2010, Finding a Short and Accurate Decision Rule in Disjunctive Normal Form by Exhaustive Search, Machine Learning, 80, DOI 10.1007/s10994-010-5168-9, 33–62.

Rodríguez M., Escalante D. M., and Peregrín A., 2011, Efficient Distributed Genetic Algorithm for Rule extraction, Applied Soft Computing, 11 (1), 733-743.

Rodríguez M., Escalante D. M., and Peregrín A., 2011, Efficient Distributed Genetic Algorithm for Rule Extraction, Applied Soft Computing, 11 (1), 733-743.

Saito, K. and Nakano, R., 1988, Medical diagnostic expert system based on PDP model, In Proceedings of the IEEE International Conference on Neural Networks, San Diego, CA. IEEE Pres, 255-262.

Sanz J.A., Fernández A., Bustince H. and Herrera F., 2010, Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning, Information Sciences, 180 (19), 3674–3685.

Sarkar B.K., Sana S.S. and Chaudhuri K., 2012, A genetic algorithm-based rule extraction system, Applied Soft Computing, 12 (1), 238–254.

Özşen S., 2008, Biyomedikal Sınıflama Problemleri için Problem-Tabanlı Bir Yapay Bağışıklık Sisteminin Geliştirilmesi ve Biyomedikal Sınıflama Problemlerine Uygulanması, Doktora Tezi, Fen Bilimleri Enstitüsü, Konya.

Sethi I. and Yoo J., 1996, Multi-valued logic mapping of neurons in feedforward networks, Eng Intell. Syst., 4 (4), 243–253.

Shin, K.-S., Lee, T.S., and Kim, H.-J., 2005, An application of support vector machines in bankruptcy prediction model, Expert Syst. Appl., 28, 127–135.

Smith, J. W., J. E. Everhart, et al., 1988, Using the ADAP learning algorithm to forecast the onset of diabetes mellitus, Proceedings of the Symposium on Computer Applications and Medical Care, 261-265.

Snyders, S. and Omlin, C., 2001, Rule Extraction from Knowledge-Based Neural Networks with Adaptive Inductive Bias, in: Proceedings of the Eighth International Conference on Neural Information Processing, 1, 143–148.

Tan, K. C., Yu, Q., Ang, J. H, 2006, Dual-Objective Evolutionary Algorithm for Rules Extraction in Data Mining, Computational Optimization and Applications, 34, 273-294.

Tang, Y., Jin, B., Sun,Y. and Zhang, Y-Q. 2004, Granular Support Vector Machines for Medical Binary Classification Problems, IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, 73-78.

Tchoumatchenko, I. and Ganascia, J.-G., 1994, A Bayesian framework to integrate symbolic and neural learning, In Proceedings of the Eleventh International Conference on Machine Learning, New Brunswick, NJ. Morgan Kaufmann, 302-308.

Thrun, S., 1995, Extracting rules from artificial neural networks with distributed representations, Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, 7, 1995.

Thabtah, F. A., Cowling, P. I., 2007, A Greedy Classification Algorithm Based on Association Rule, Applied Soft Computing, 7, 1102-1111.

Thangavel, K., Jaganathan, P., 2007, Rule Mining Algorithm with a New Ant Colony Optimization Algorithm, International Conference on Computational Intelligence and Multimedia Applications, 2, 135-140.

Tokinaga, S., Lu, J., Ikeda, Y., 2005, Neural Network Rule Extraction by Using The Genetic Programming and Its Applications to Explanatory Classifications, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E88-A (10), 2627-2635.

Towell, G.G. and Shavlik, J., 1993, Extracting refined rules from knowledgebased neural networks, Mach. Learn., 13, 71–101.

Turna F., 2011, Veri Madenciliği Teknikleriyle Tramvay Arıza Kayıtlarından Kural Çıkarımı, Yüksek Lisans Tezi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı, Kayseri.

Van Gestel, T., Suykens, J. A. K., Lanckriet, G., Lambrechts, A., De Moor, B., and Vandewalle, J., 2002, Bayesian framework for least squares support vector machine classifiers, gaussian processes and kernel fisher discriminant analysis, Neural Computation, 14 (5), 1115–1147.

Wahba, G., C. Gu, et al., 1992, Soft Classification, a.k.a. Risk Estimation, via Penalized Log Likelihood and Smoothing Spline Analysis of Variance, The mathematics of generalization: the proceedings of the SFI/CNLS Workshop on Formal Approaches to Supervised Learning, Santa Fe, Addison-Wesley Pub. Co., 331-360.

Wang S., Wu G. and Pan J., 2007, A Hybrid Rule Extraction Method Using Rough Sets and Neural Networks, Lecture Notes in Computer Science Part II, 4492, 352–361.

Weijters, T., Bosh, A.V.D. and Herik, J.V.D., 1997, Intelligible neural Networks with BP_SOM, Proceedings of the Ninth Dutch Conference on Artificial Intelligence, 27–36.

Weiss S.M. and Kapouleas, I., 1990, An empirical comparison of pattern recognition, neural nets and machine learning classification methods, J.W. Shavlik and T.G. Dietterich, Readings in Machine Learning, Morgan Kauffman Publ, CA.

Zahiri, S.-H., Seyedin, S.-A., 2007, Swarm Intelligence Based Classifiers, Journal of the Franklin Institute, 344, 362-376.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı	**:**	Murat KÖKLÜ
Uyruğu	**:**	T.C.
Doğum Yeri ve Tarihi	**:**	Cihanbeyli, 1979
Telefon	**:**	0332 223 33 35
Faks	**:**	0332 241 21 79
e-mail	**:**	mkoklu@selcuk.edu.tr

EĞİTİM

Derece		**Adı, İlçe, İl**	**Bitirme Yılı**
Lise	:	Konya Meram Gazi Lisesi	1997
Üniversite	:	Selçuk Üniversitesi Teknik Eğitim Fakültesi, Bilgisayar Sistemleri Öğretmenliği Bölümü	2002
Yüksek Lisans	:	Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Elektronik ve Bilgisayar Eğitimi A.B.D.	2005
Doktora	:	Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği A.B.D.	2014

İŞ DENEYİMLERİ

Yıl	**Kurum**	**Görevi**
Eylül 2002 – Aralık 2002 (3 Ay)	Afyonkarahisar Gazi Teknik ve Endüstri Meslek Lisesi	Teknik Öğretmen
Aralık 2002 – Devam Ediyor	Selçuk Üniversitesi Teknik Eğitim Fakültesi	Araştırma Görevlisi

YABANCI DİLLER

İngilizce, ÜDS Mart 2005 - 73.75

YAYINLAR

1.1. Uluslararası hakemli dergilerde yayınlanan makaleler (SCI & SSCI & Arts ve Humanities)

1. **Köklü M.,** Kahramanlı H., ve Allahverdi N., 2012, A New Approach to Classification Rule Extraction Problem by the Real Value Coding, International Journal of Innovative Computing, Information and Control (IJICIC), 8 (9), 6303-6315. **(Doktara Tezinden)**

2. **Köklü M.,** Kahramanlı H., ve Allahverdi N., 2014, Sınıflandırma Kurallarının Çıkarımı İçin Etkin ve Hassas Yeni Bir Yaklaşım, Journal of the Faculty of Engineering and Architecture of Gazi University, 29 (3), 477-486. **(Doktara Tezinden)**

1.2. Uluslararası bilimsel toplantılarda sunulan ve bildiri kitabında (*Proceedings*) basılan bildiriler

1. **Köklü M.,** Kahramanlı H., ve Allahverdi N., 2013, Classification Rule Extraction Using Artificial Immune System: Clonalg and its Application, 4th International Conference on Mathematical and Computational Applications (ICMCA-2013), Celal Bayar University, 11-13 June 2013, 283-292, Manisa. **(Doktara Tezinden)**