ÖZET

Doktora Tezi

BİYOMEDİKAL SİNYALLERDE VERİ ÖN-İŞLEME TEKNİKLERİNİN MEDİKAL TEŞHİSTE SINIFLAMA DOĞRULUĞUNA ETKİSİNİN İNCELENMESİ

Kemal POLAT
Selçuk Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik-Elektronik Mühendisliği Anabilim Dalı
Danışman: Yrd. Doç. Dr. Salih GÜNEŞ
2008, 151 sayfa

Jüri: **Prof. Dr. Sadık KARA**
Prof. Dr. Mehmet BAYRAK
Prof. Dr. Saadetdin HERDEM
Yrd. Doç. Dr. Nihat YILMAZ
Yrd. Doç. Dr. Salih GÜNEŞ

Bu tez çalışmasında, biyomedikal veri kümelerinin sınıflandırılmasında sınıflama performansını arttırmak için veri ağırlıklandırma ve özellik seçme yöntemleri önerilmiş ve kullanılmıştır. Biyomedikal veri kümelerini sınıflamada sınıflama performansını azaltan bazı etmenler vardır. Bu etmenler gürültü, aykırı değer, lineer olmayan bir veri dağılımına sahip olma gibi durumlardır. Yukarıdaki etmenlere sahip olan veri kümelerinin sınıflama performanslarını arttırmak için çeşitli veri ön-işleme teknikleri kullanılır.

Biyomedikal veri kümelerinde, özellik çıkarımından sonra oluşturulan veri setinin boyutu fazla olabilir veya veri setinde ilgisiz/fazla özellikler olabilir. Bu özelliklerin dezavantajları; sınıflama performansını azaltır ve sınıflandırıcının hesaplama maliyetini arttırır. Yapılan çalışmalarda, özellik seçme algoritmaları ile daha yüksek genelleştirme yeteneği ve daha az işlem karışıklığı elde edilmiştir. Bu tez çalışmasında, boyut azaltımı ve özellik seçme algoritması olarak, *temel bileşen analizi*, *bilgi kazancına dayanan özellik seçme algoritması* ve *Kernel F-skor özellik seçme* yöntemleri özelik seçme algoritmaları olarak kullanılmıştır. Bu yöntemler arasında, özellik seçme olarak, *bilgi kazancına dayanan özellik seçme algoritması* ile *Kernel F-skor özellik seçme* yöntemi ön plana çıkmaktadır. Boyut azaltımı olarak da *temel bileşen analizine* ağırlık verilmiştir.

Veri ağırlıklandırma yöntemleri olarak, *bulanık ağırlıklandırma ön-işleme*, *k-NN (k-en yakın komşu) tabanlı veri ağırlıklandırma ön-işleme*, *genelleştirilmiş ayrışım analizi* ve *benzerlik tabanlı veri ağırlıklandırma ön-işleme* yöntemleri medikal veri kümelerini sınıflamada sınıflama performansını iyileştirmek için kullanılmış ve önerilmiştir.

Bu tez çalışmasında kullanılan biyomedikal veri kümeleri; kalp hastalığı, SPECT (Single Photon Emission Computed Tomography) görüntüleri ile kalp hastalığı, E.coli Promoter gen dizileri, Doppler sinyali ile damar sertliği (Atherosclerosis) hastalığı, VEP (Görsel Uyarılmış Potansiyel) sinyali ile optik sinir hastalığı ve PERG (Örüntü Retinografisi) sinyali ile Macular hastalığı veri kümeleridir. Bu veri kümeleri içinden, kalp hastalığı, SPECT (Single Photon Emission Computed Tomography) görüntüleri ile kalp hastalığı, E.coli Promoter gen dizileri veri kümeleri, UCI (University of California, Irvine) makine öğrenmesi veritabanından alınmıştır. Doppler sinyali ile damar sertliği hastalığı, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile Macular hastalığı veri kümeleri ise Fatih Üniversitesi Öğretim Üyesi Prof. Dr. Sadık Kara ve Erciyes Üniversitesi Biyomedikal Mühendisliği ekibi tarafından alınan verilerdir.

Veri ön-işleme ve özellik seçme yöntemlerinin performanslarını değerlendirmek için bu yöntemler sınıflama algoritmaları ile hibrid olarak kullanılmışlardır. Kullanılan sınıflama algoritmaları, *ANFIS* (Adaptif Ağ Tabanlı Bulanık Çıkarım Sistemi), *C4.5 karar ağacı*, *YBTS (*Yapay Bağışıklık Tanıma Sistemi*)*, *bulanık kaynak dağılım mekanizmalı YBTS* ve *yapay sinir ağlarıdır*.

Biyomedikal veri kümelerinin sınıflandırılması sonucunda, veri ağırlıklandırma yöntemleri arasında en iyi sonuçları veren yöntem, *k-NN (k- en yakın komşu) tabanlı veri ağırlıklandırma* yöntemi olmuştur. Özellik seçme yöntemleri arasında ise *temel bileşen analizi* diğer özellik seçme yöntemlere göre üstün sonuçlar elde etmiştir.

Özellik seçme yöntemleri, veri ağırlıklandırma yöntemleri ile sınıflama algoritmaları birleştirilerek 12 yeni hibrid sistem oluşturulmuş ve bu yeni hibrid sistemler tezde kullanılan 6 medikal veri kümesine uygulanmıştır. Hesaplama maliyeti ve sınıflama performansı açısından her bir medikal veri kümesi için en iyi hibrid model seçilmiştir.

Anahtar Kelimeler – Sınıflama, Özellik Seçme, Veri Ağırlıklandırma, Medikal Veri Kümeleri, Hibrid Sistemler

ABSTRACT

PhD. Thesis

THE INVESTIGATION OF EFFECT OF DATA PRE-PROCESSING TECHNIQUES TO CLASSIFICATION ACCURACY ON MEDICAL DIAGNOSIS IN BIOMEDICAL SIGNALS

Kemal POLAT

Selcuk University

Graduate School of Natural and Applied Sciences

Department of Electrical-Electronics Engineering

Supervisor: Assist. Prof. Dr. Salih GÜNEŞ

2008, 151 pages

Jury: **Prof. Dr. Sadık KARA**

Prof. Dr. Mehmet BAYRAK

Prof. Dr. Saadetdin HERDEM

Assist. Prof. Dr. Nihat YILMAZ

Assist. Prof. Dr. Salih GÜNEŞ

In this PhD. thesis, data weighting and feature selection methods are proposed and used for increasing the performance of classification of biomedical datasets. There are some factors that decrease the classification performance on classification of biomedical datasets. These factors are noise, invalid data, non-linearly separable data distribution etc. Various data pre-processing methods are used to increase the classification performance of medical datasets afflicted above factors.

In the biomedical datasets, after feature extraction, the dimension of produced dataset can be huge or biomedical datasets may contain the irrelevant or redundant

features. The disadvantages of these features are as follows: they decrease the classification performance and increase the computation cost of classifier. In the conducted studies, higher generalization ability and lesser operational complexity are achieved with feature selection and dimensionality reduction algorithms. In this thesis, *principal component analysis*, *feature selection algorithm based on information gain*, and *kernel f-score feature selection* methods are proposed and used as feature selection and dimensionality reduction algorithms. Among these methods, *feature selection algorithm based on information gain* and *kernel f-score feature selection* methods are emphasized. As for the dimensionality reduction process, more weight is given to *principal component analysis*.

As data weighting methods, *fuzzy weighted pre-processing*, *k-NN based weighted pre-processing*, *generalized discriminant analysis*, *similarity based weighted pre-processing* methods are proposed and used to improve the performance of classifier in classification of biomedical datasets. Among above methods, the proposed data weighted methods are *fuzzy weighted pre-processing*, *k-NN based weighted pre-processing*, and *similarity based weighted pre-processing* methods.

In this PhD. thesis, the used biomedical datasets are heart disease, heart disease with SPECT (Single Photon Emission Computed Tomography) images, E.coli Promoter gene sequences, Atherosclerosis disease with Doppler signals, optic nerve disease with VEP (Visual Evoked Potentials) signals, and macular disease with PERG (Pattern Electroretinography) datasets. Among datasets, heart disease, heart disease with SPECT (Single Photon Emission Computed Tomography) images, E.coli Promoter gene sequences datasets are taken from UCI (University of California, Irvine) machine learning database. The other datasets including Atherosclerosis disease with Doppler signals, optic nerve disease with VEP signals, and macular disease with PERG datasets are taken from Prof. Dr. Sadık Kara in Fatih University and biomedical engineering team in Erciyes University.

In order to evaluate the performances of data weighting and feature selection methods, these methods are used as hybrid with classifier algorithms. Used classification algorithms are *ANFIS (Adaptive Network Based Fuzzy Inference System)*, *C4.5 decision tree classifier algorithm*, *AIRS (Artificial Immune Recognition Immune System)*, *Fuzzy-AIRS (Artificial Immune Recognition Immune System with Fuzzy Resource Allocation Mechanism)* and Artificial neural network.

As a result of classifying the biomedical datasets, *k-NN based weighted method* was the best data weighting method among others. Among feature selection methods, the principal component analysis was superior to other methods.

The twelve new hybrid systems was created combining feature selection methods, data weighting methods and classifier algorithms. These novel hybrid systems were applied to six medical datasets used in this thesis. The best hybrid system in terms of computation time and classification performance was chosen for each medical dataset.

Keywords – Classification, Feature Selection, Data Weighting, Medical Datasets, Hybrid Systems.

TEŞEKKÜR

Çalışmama verdiği destek ve gösterdiği anlayıştan dolayı danışmanım sayın Yrd.Doç.Dr. Salih GÜNEŞ'e, tez süresince belirttikleri görüşler ve önerilerle tezin yönlenmesinde yardımcı olan Tez İzleme Komitesi Üyeleri sayın Prof.Dr. Sadık KARA ve sayın Prof.Dr. Saadetdin HERDEM'e, Bölüm Başkanımız sayın Prof. Dr. Mehmet BAYRAK'a, Makine Öğrenmesi alanında sağladığı değerli bilgilerden dolayı sayın Prof.Dr. Ahmet ARSLAN'a, beraberce yapmış olduğumuz çalışmalardan dolayı sayın Arş.Gör.Dr. Seral ÖZŞEN'e, sayın Arş.Gör.Dr. Halife KODAZ'a ve Arş. Gör. Bayram AKDEMİR'e, Selçuk Üniversitesi Mühendislik-Mimarlık Fakültesi Elektrik-Elektronik Mühendisliği ve Bilgisayar Mühendisliği Bölümlerindeki tüm çalışma arkadaşlarıma ve özellikle manevi desteğini hiç bir zaman benden esirgemeyen her zaman ve her koşulda hep yanımda olan aileme sonsuz teşekkür ederim.

Kemal POLAT
2008, Konya

İÇİNDEKİLER

ÖZET ..... **i**
ABSTRACT ..... **iii**
TEŞEKKÜR ..... **v**
İÇİNDEKİLER ..... **vi**
SİMGELER VE KISALTMALAR ..... **viii**
ŞEKİLLER LİSTESİ ..... **x**
TABLOLAR LİSTESİ ..... **xii**
1. GİRİŞ ..... **xii**
1.1. Literatür Araştırması 5
1.2. Çalışmanın Temel Amacı ve Literatüre Katkıları 14
1.3. Çalışmanın Kapsamı ve Organizasyonu 15
2. ÖZELLİK ÇIKARMA VE ÖZELLİK SEÇME YÖNTEMLERİ **17**
2.1. Özellik Çıkarımı İşlemi ve Welch Metodu 18
2.1.1. Pencereleme İşlemi 18
2.1.2. Fourier analizi ile güç spektrum yoğunluğunun bulunması 19
2.2. Özellik Seçme ve Boyut Azaltımı Yöntemleri 23
2.2.1. Temel bileşen analizi–TBA (Principal component analysis-PCA) 24
2.2.2. Bilgi kazancına dayanan özellik seçme algoritması 28
2.2.3. Kernel F-Skor özellik seçme yöntemi 33
3. VERİ AĞIRLIKLANDIRMA VE ÖN-İŞLEME YÖNTEMLERİ **38**
3.1. Bulanık Mantık Tabanlı Veri Ağırlıklandırma İşlemi 39
3.2. k-NN (k-en yakın) Komşu Tabanlı Veri Ağırlıklandırma İşlemi 43
3.3. Genelleştirilmiş Ayrışım Analiz (GAA) Yöntemi 46
3.4. Benzerlik Tabanlı Veri Ağırlıklandırma Yöntemi 49
4. BİYOMEDİKAL VERİLERİ SINIFLAMADA KULLANILAN YÖNTEMLER **55**
4.1. Sınıflama Algoritmaları 55
4.1.1. Çok katmanlı algılayıcı yapay sinir ağları 56
4.1.2. Geriye yayılım (Back-propagation) algoritması 57
4.2. ANFIS (Adaptive Network Based Fuzzy Inference System) 58
4.3. Yapay Bağışıklık Tanıma Sistemi (Artificial Immune Recognition System) Sınıflama Algoritması 64
4.4. Bulanık Kaynak Dağılımlı Yapay Bağışıklık Tanıma Sistemi (Fuzzy-AIRS) 65
4.4.1. Bulanık kaynak dağılım mekanizması 67

4.5. C4.5 Karar Ağaçları 70

5. KULLANILAN BİYOMEDİKAL VERİ KÜMELERİ 72

5.1. Özelliği Belirlenmiş Medikal Veri Kümeleri 72
5.1.1. Kalp hastalığı veri kümesi 73
5.1.2. E. Coli promoter gen dizileri veri kümesi 75
5.1.3. SPECT (Single Photon Emission Computed Tomography) görütüleri kalp hastalığı teşhisi veri kümesi 78
5.2. Özelliği Çıkarılmış Medikal Veri Kümeleri 80
5.2.1. Doppler sinyali ile damar sertliği veri kümesi uygulamaları 80
5.2.2. VEP (Visual Evoked Potential- Görsel Uyarılmış Potansiyel) sinyali ile optik sinir hastalığı veri kümesi 83
5.2.3. PERG (Pattern Electroretinography) sinyali ile Macular hastalığı veri kümesi 87

6. VERİ ÖN-İŞLEME YÖNTEMLERİNİN MEDİKAL VERİ KÜMELERİNİ SINIFLAMADAKİ PERFORMANSI VE UYGULAMA SONUÇLARI 90

6.1. Performans Ölçüm Kriterleri 91
6.1.1. 5-kat kez çaprazlama ile veri kümeleri ayırma 92
6.1.2. Sınıflama doğruluğu ve duyarlılık-seçicilik değerleri 93
6.1.3. ROC (Receiver Operating Characterictics-Alıcı İşlem Karakteristiği) eğrileri ve eğrinin altında kalan alan (AUC) 95
6.2. Veri Ağırlıklandırma Yöntemlerinin Etkisi 97
6.2.1. Bulanık mantık veri ağırlıklandırma yöntemi sonuçları 99
6.2.2. k-NN (k-en yakın komşu) tabanlı veri ağırlıklandırma yöntemi sonuçları 101
6.2.3. Genelleştirilmiş ayrışım analizi (Generalized Discriminant Analysis) sonuçları 108
6.2.4. Benzerlik tabanlı veri ağırlıklandırma veri ön-işleme yöntemi sonuçları 110
6.3. Boyut Azaltımı ve Özellik Seçme Yöntemlerinin Etkisi 116
6.3.1. Temel bileşen analizi ile boyut azaltımı yönteminin sonuçları 117
6.3.2. Kernel F-skor özellik seçme yönteminin sonuçları 119
6.3.3. Bilgi kazancına dayanan özellik seçme yönteminin sonuçları 126
6.4. Ön-işleme ve Özellik Seçme Yöntemleri Birleştirilerek Oluşturulan Hibrid Sistemler 129

7. SONUÇLAR VE ÖNERİLER 138

7.1. Sonuçlar 138
7.2. Öneriler 142

8. KAYNAKLAR 144

SİMGELER VE KISALTMALAR

ab	**:** Tek bir antikoru ifade etmektedir.
AB	**:** Antikor kümesini ifade etmektedir.
ab.c	**:** Tek bir *ab* antikorunun sınıf değerini temsil etmektedir.
ab.uyarım	**:** Tek bir *ab* antikorunun uyarım seviyesini ifade etmektedir.
ab.kaynak	**:** Tek bir *ab* antikoruna tahsis edilen kaynak sayısını ifade etmektedir.
ADD	**:** Ayrık dalgacık dönüşümü
ag	**:** Tek bir *ag* antijenini ifade etmektedir.
ag.c	**:** Tek bir *ag* antijenin sınıfını ifade etmektedir.
ag.f	**:** Tek bir antijenin özellik vektörünü ifade etmektedir.
ANFIS	**:** Adaptif ağ tabanlı bulanık çıkarım sistemi (Adaptive network based fuzzy inference system)
ANN	**:** Yapay sinir ağları (Artificial neural network)
AIRS	**:** Yapay bağışıklık tanıma sistemi (Artificial immune recognition system)
AR	**:** Özbağlanımlı (Autoregressive)
ATS	**:** Duyarlılık eşik ölçüsü (Affinity threshold scalar)
AUC	**:** ROC eğrisi altında kalan alan (Area under the ROC curve)
AWAIS	**:** Özellik ağırlıklandırmalı yapay bağışıklık sistemi (Attribute weighted artificial immune system)
BBA	**:** Bağımsız bileşen analizi
BKO	**:** Bulanık kümele ortalama
BKÖS	**:** Bilgi kazancına dayanan özellik seçme algoritması
BVAÖ	**:** Bulanık veri ağırlıklandırma ön-işleme yöntemi
ECG	**:** Elektrokardiyogram
EEG	**:** Electroencephalography (Beyin sinyalleri)
Entropi	**:**Verideki rasgeleliği, belirsizliği, kirliliği, beklenmeyen durumun ortaya çıkma olasılığını karakterize eden bir ölçüdür.
FFT	**:** Hızlı fourier dönüşümü (Fast Fourier transform)
GRNN	**:** Genel regresyon sinir ağları (General regression neural network)

GUI	**:** Grafiksel kullanıcı arabirimi (Graphical user interface)
Fuzzy-AIRS	**:** Bulanık-yapay bağışıklık tanıma sistemi (Fuzzy-artificial immune recognition system)
GAA	**:** Genelleştirilmiş ayrışım analizi
ID3	**:** Iterative Dichotomiser 3
Kaynak sayısı	**:** İzin verilen sistem kaynaklarının toplam sayısını ifade etmektedir.
KFSÖS	**:** Kernel F-skor özellik seçme yöntemi
k-NN	**:** k-en yakın komşu (k-nearest neighbor)
LAA	**:** Lineer ayrışım analizi
LM	**:** Levenberg marquard
Mutasyon oranı	**:** Bir antikorun herhangi bir özelliğinin mutasyona uğrama olasılığını işaret eden *0* ile *1* arasında bir parametredir.
ÖVN	**:** Öğrenme vektör niceleme
PERG	**:** Örüntü retinografisi (Pattern Electroretinography)
RBF	**:** Radyal taban fonksiyonu (Radial basis function)
ROC	**:**Alıcı işlem karakteristiği (Receiver operating characteristics)
SPECT	**:** Tek foton emisyon hesaplanmış tomografi (Single photon emission computed tomography)
TBA	**:** Temel bileşen analizi
UCI	**:** University of California, Irvine
Uyarma eşiği	**:** Antijenin eğitimi için bir durma kriteri olarak kullanılan *0* ile *1* arasında bir parametredir.
VEP	**:** Görsel uyarılmış potansiyel (Visual evoked potential)
YBS	**:** Yapay bağışıklık sistemi
YBTS	**:** Yapay bağışıklık tanıma sistemi
YSA	**:** Yapay sinir ağları

ŞEKİLLER LİSTESİ

Şekil 1.1 Tezde kullanılan örüntü tanıma sisteminin blok diyagramı…….....3
Şekil 2.1 Özellik çıkarımı ve özellik seçmeyi gösteren bir blok diyagramı.....18
Şekil 2.2 44100 Hz' de örneklenmiş bir Doppler sinyali örneği.....21
Şekil 2.3 64 pencereye ayrılmış Doppler sinyalinin Welch yöntemi sonucu güç spektrum yoğunluğu.....22
Şekil 2.4 128 pencereye ayrılmış Doppler sinyalinin Welch yöntemi sonucu güç spektrum yoğunluğu.....22
Şekil 2.5 256 pencereye ayrılmış Doppler sinyalinin Welch yöntemi sonucu güç spektrum yoğunluğu.....23
Şekil 2.6 Temel bileşen analizinin blok diyagramı.....24
Şekil 2.7 Bilgi kazancı tabanlı özellik seçme algoritması.....29
Şekil 2.8 C4.5 veri ayrışımı.....29
Şekil 2.9 Kernel F-skor özellik seçme yönteminin akış şeması.....35
Şekil 2.10 Orijinal XOR veri dağılımı.....36
Şekil 2.11 RBF Kernel uzayına attıktan sonraki veri dağılımı.....37
Şekil 2.12 Lineer Kernel uzayına attıktan sonraki veri dağılımı.....37
Şekil 3.1 Giriş üyelik fonksiyonları.....39
Şekil 3.2 Çıkış üyelik fonksiyonları.....39
Şekil 3.3 k-NN tabanlı veri ağırlıklandırma ön-işleme metodunun blok diyagramı.....43
Şekil 3.4 Genelleştirilmiş ayrışım analizi yönteminin akış şeması.....48
Şekil 3.5 Benzerlik tabanlı veri ağırlıklandırma ön-işlemenin blok şeması.....52
Şekil 4.1 Danışmalı öğrenme sisteminin şematik gösterimi.....56
Şekil 4.2 Çok katmanlı bir perseptron.....58
Şekil 4.3 ANFIS' in öğrenme ve test akış diyagramı.....59
Şekil 4.4 İki girişi bir ANFIS mimarisi.....63
Şekil 4.5 YBTS' nin eğitme ve test akış şeması.....65
Şekil 4.6 Fuzzy-AIRS sınıflama algoritmasının akış şeması.....66

Şekil 4.7 (a) Giriş Üyelik Fonksiyonu, (b) Çıkış Üyelik Fonksiyonu.....68
Şekil 5.1 DNA nükleoitleri.....76
Şekil 5.2 Doppler verisi elde etmek için kullanılan donanımın blok diyagramı.....81
Şekil 5.3 Bileşenlerle etiketlenmiş normal VEP sinyali, N75, P100,
N145 ise gecikmelerdir.....84
Şekil 5.4 İki farklı insan gözünün VEP cevap örnekleri.....84
Şekil 5.5 VEP sinyalini elde etmek için kullanılan sistemin blok diyagramı..... 85
Şekil 5.6 İki farklı kişiye ait PERG cevabı sinyalleri.....88
Şekil 6.1 Medikal verileri sınıflamak için kullanılan sistemin blok diyagramı.....91
Şekil 6.2 Beş kat çaprazlamayı gösteren şematik bir diyagram.....92
Şekil 6.3 Farklı AUC değerlerine sahip ROC eğrileri.....95

TABLOLAR LİSTESİ

Tablo 2.1 x ve y değişkeninden oluşan bir veri kümesi.....27
Tablo 2.2 Ortalama değerden çıkarılmış veri kümesini yeni hali.....27
Tablo 2.3 Sürekli değerli üç özelliğe sahip tenis oynama durumu için veri kümesi..31
Tablo 2.4 XOR problemi veri dağılımı.....36
Tablo 3.1 Bulanık veri ağırlıklandırma ön-işleme için kullanılan bulanık kural tabanı.....41
Tablo 3.2 *x* ve *y* değişkeninden oluşan veri kümesi.....41
Tablo 3.3 *x* ve *y* değişkenin [0,1] aralığındaki normalize değerleri.....42
Tablo 3.4 *x* ve *y* değişkenin bulanık ağırlıklandırma sonrası yeni değerleri.....42
Tablo 3.5 *x* ve *y* değişkeninden oluşan veri kümesi.....44
Tablo 3.6 *x* ve *y* değişkenin k-NN tabanlı veri ağırlıklandırma sonrası yeni dağılımı (k=3 için).....45
Tablo 3.7 *x* ve *y* değişkenin k-NN tabanlı veri ağırlıklandırma sonrası yeni dağılımı (k=4 için).....45
Tablo 3.8 *x* ve *y* değişkenin k-NN tabanlı veri ağırlıklandırma sonrası yeni dağılımı (k=5 için).....45
Tablo 3.9 x ve y değişkeninden oluşan veri kümesi.....48
Tablo 3.10 x ve y değişkeninden oluşan veri kümesine GAA uygulandıktan sonraki veri dağılımı.....49
Tablo 3.11 x ve y değişkeninden oluşan veri kümesi.....53
Tablo 3.12 *x* ve *y* değişkenin [0,1] aralığındaki normalize değerleri.....53
Tablo 3.13 *x* ve *y* değişkenin, kosinüs genliği benzerlik fonksiyonuna göre veri kümesinin ağırlıklandırılmış yeni dağılımı.....53
Tablo 3.14 *x* ve *y* değişkenin, mutlak üstel benzerlik fonksiyonuna göre veri kümesinin ağırlıklandırılmış yeni dağılımı.....54
Tablo 3.15 *x* ve *y* değişkenin, üstel benzerlik katsayı benzerlik fonksiyonuna göre veri kümesinin ağırlıklandırılmış yeni dağılımı.....54
Tablo 4.1. ANFIS için hibrid öğrenme prosedüründeki iki geçiş.....63

Tablo 4.2 Bulanık kaynak dağılımı için kural tabanı.....69
Tablo 4.3 Giriş ve çıkış üyelik fonksiyonları için dilsel değerler.....69
Tablo 4.4 Fuzzy kaynak dağılımı için örnek bir uygulama.....70
Tablo 5.1 Kalp hastalığı veri kümesindeki özellikleri dağılımı.....73
Tablo 5.2 Kalp hastalığı veritabanında her bir durum için çıkış değerlerine bağlı özellikler.....74
Tablo 5.3 Kalp hastalığı veri kümesinde bulunan özelliklerin istatistiki değerleri....74
Table 5.4 Promoterlar için bir alan teorisi.....76
Tablo 5.5 E. Coli promoter gen dizileri veritabanında her bir durum için çıkış değerlerine bağlı özellikler.....76
Tablo 5.6 DNA dizilerindeki nükleoitlerin görülme sıklıkları.....76
Tablo 5.7 E. Coli promoter gen dizileri veri kümesinde bulunan özelliklerin istatistiki değerleri.....77
Tablo 5.8 Kalp ile ilgili SPECT görüntüleri veritabanında her bir durum için çıkış değerlerine bağlı özellikler.....79
Tablo 5.9 Kalp ile ilgili SPECT görüntüleri veri kümesinde bulunan özelliklerin istatistiki değerleri.....79
Tablo 5.10 Doppler sinyali ile damar sertliği veri kümesinde her bir durum için çıkış değerlerine bağlı özellikler.....81
Tablo 5.11 Doppler sinyali ile damar sertliği veri kümesinde bulunan özelliklerin istatistiki değerleri..... 82
Tablo 5.12 VEP sinyali ile optik sinir hastalığı veri kümesinde her bir durum için çıkış değerlerine bağlı özellikler.....85
Tablo 5.13 VEP sinyali ile optik sinir hastalığı veri kümesinde bulunan özelliklerin istatistiki değerleri.....86
Tablo 5.14 PERG sinyali ile Macular hastalığı veri kümesinde her bir durum için çıkış değerlerine bağlı özellikler.....88
Tablo 5.15 PERG sinyali ile Macular hastalığı veri kümesinde bulunan özelliklerin istatistiki değerleri..... 88
Tablo 6.1 Karışıklık matrisinin gösterimi.....93
Tablo 6.2 Sınıflama sonuç tablosu.....94
Tablo 6.3 Kalp hastalığı veri kümesinin sınıflandırılmasında ön-işlemesiz

sınıflandırıcılardan elde edilen sonuçlar.....97
Tablo 6.4 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında ön-işlemesiz sınıflandırıcılardan elde edilen sonuçlar.....98
Tablo 6.5 SPECT görüntüleri veri kümesinin sınıflandırılmasında ön-işlemesiz sınıflandırıcılardan elde edilen sonuçlar.....98
Tablo 6.6 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında ön-işlemesiz sınıflandırıcılardan elde edilen sonuçlar.....98
Tablo 6.7 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında ön işlemesiz sınıflandırıcılardan elde edilen sonuçlar.....98
Tablo 6.8 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında ön-işlemesiz sınıflandırıcılardan elde edilen sonuçlar.....99
Tablo 6.9 Kalp hastalığı veri kümesinin sınıflandırılmasında, bulanık mantık veri ağırlıklandırma veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....99
Tablo 6.10 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, bulanık mantık veri ağırlıklandırma veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....101
Tablo 6.11 SPECT görüntüleri veri kümesinin sınıflandırılmasında, bulanık mantık veri ağırlıklandırma veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar..... 100
Tablo 6.12 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, bulanık mantık veri ağırlıklandırma veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar..... 100
Tablo 6.13 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, bulanık mantık veri ağırlıklandırma veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.... 100
Tablo 6.14 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, bulanık mantık veri ağırlıklandırma veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....101
Tablo 6.15 Kalp hastalığı veri kümesinin sınıflandırılmasında,

k-NN tabanlı veri ağırlıklandırma veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....102
Tablo 6.16 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, k-NN tabanlı veri ağırlıklandırma veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....103
Tablo 6.17 SPECT görüntüleri veri kümesinin sınıflandırılmasında, k-NN tabanlı veri ağırlıklandırma veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....104
Tablo 6.18 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, k-NN tabanlı veri ağırlıklandırma veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....105
Tablo 6.19 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, k-NN tabanlı veri ağırlıklandırma veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....106
Tablo 6.20 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, k-NN tabanlı veri ağırlıklandırma veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....107
Tablo 6.21 Kalp hastalığı veri kümesinin sınıflandırılmasında, genelleştirilmiş ayrışım analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....109
Tablo 6.22 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, genelleştirilmiş ayrışım analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....109
Tablo 6.23 SPECT görüntüleri veri kümesinin sınıflandırılmasında, genelleştirilmiş ayrışım analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar..... 109
Tablo 6.24 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, genelleştirilmiş ayrışım analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar..... 109
Tablo 6.25 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, genelleştirilmiş ayrışım analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....110

Tablo 6.26 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, genelleştirilmiş ayrışım analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....110
Tablo 6.27 Kalp hastalığı veri kümesinin sınıflandırılmasında, benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....111
Tablo 6.28 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....112
Tablo 6.29 SPECT görüntüleri veri kümesinin sınıflandırılmasında, benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....113
Tablo 6.30 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....114
Tablo 6.31 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....115
Tablo 6.32 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....116
Tablo 6.33 Kalp hastalığı veri kümesinin sınıflandırılmasında, temel bileşen analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....118
Tablo 6.34 E.coli promoter gen dizileri veri kümesin sınıflandırılmasında, temel bileşen analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....118
Tablo 6.35 SPECT görüntüleri veri kümesinin sınıflandırılmasında, temel bileşen analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....118
Tablo 6.36 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, temel bileşen analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....118

Tablo 6.37 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, temel bileşen analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....119
Tablo 6.38 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, temel bileşen analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....119
Tablo 6.39 Kalp hastalığı veri kümesinin sınıflandırılmasında, f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....120
Tablo 6.40 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....120
Tablo 6.41 SPECT görüntüleri veri kümesinin sınıflandırılmasında, f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....120
Tablo 6.42 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....121
Tablo 6.43 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....121
Tablo 6.44 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....121
Tablo 6.45 Kalp hastalığı veri kümesinin sınıflandırılmasında, Lineer kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....122
Tablo 6.46 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, Lineer kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....122
Tablo 6.47 SPECT görüntüleri veri kümesinin sınıflandırılmasında, Lineer kernel f-skor özellik seçme ile sınıflandırıcıların

birleşiminden elde edilen sonuçlar.....122
Tablo 6.48 Doppler siyali ile damar sertliği veri kümesinin sınıflandırılmasında, Lineer kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....123
Tablo 6.49 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, Lineer kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....123
Tablo 6.50 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, Lineer kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....123
Tablo 6.51 Kalp hastalığı veri kümesinin sınıflandırılmasında, RBF kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....124
Tablo 6.52 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, RBF kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....124
Tablo 6.53 SPECT görüntüleri veri kümesinin sınıflandırılmasında, RBF kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....124
Tablo 6.54 Doppler siyali ile damar sertliği veri kümesinin sınıflandırılmasında, RBF kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....125
Tablo 6.55 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, RBF kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....125
Tablo 6.56 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, RBF kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....125
Tablo 6.57 Kalp hastalığı veri kümesinin sınıflandırılmasında, bilgi kazancına dayanan özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....126
Tablo 6.58 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında,

bilgi kazancına dayanan özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....127
Tablo 6.59 SPECT görüntüleri veri kümesinin sınıflandırılmasında, bilgi kazancına dayanan özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....127
Tablo 6.60 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, bilgi kazancına dayanan özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....128
Tablo 6.61 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, bilgi kazancına dayanan özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar.....129
Tablo 6.62 Önerilen hibrid sistemler.....130
Tablo 6.63 Medikal veri kümelerini sınıflandırılmasında en iyi sonucu elde eden hibrid sistemler.....131
Tablo 6.64 Kalp hastalığı veri kümesinin sınıflandırılmasında önerilen hibrid sistemlerden elde edilen sınıflama doğrulukları.....132
Tablo 6.65 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında önerilen hibrid sistemlerden elde edilen sınıflama doğrulukları.....133
Tablo 6.66 SPECT görüntüleri veri kümesinin sınıflandırılmasında önerilen hibrid sistemlerden elde edilen sınıflama doğrulukları.....134
Tablo 6.67 Doppler siyali ile damar sertliği veri kümesinin sınıflandırılmasında önerilen hibrid sistemlerden elde edilen sınıflama doğrulukları.....135
Tablo 6.68 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında önerilen hibrid sistemlerden elde edilen sınıflama doğrulukları.....136
Tablo 6.69 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında önerilen hibrid sistemlerden elde edilen sınıflama doğrulukları.....137

1. GİRİŞ

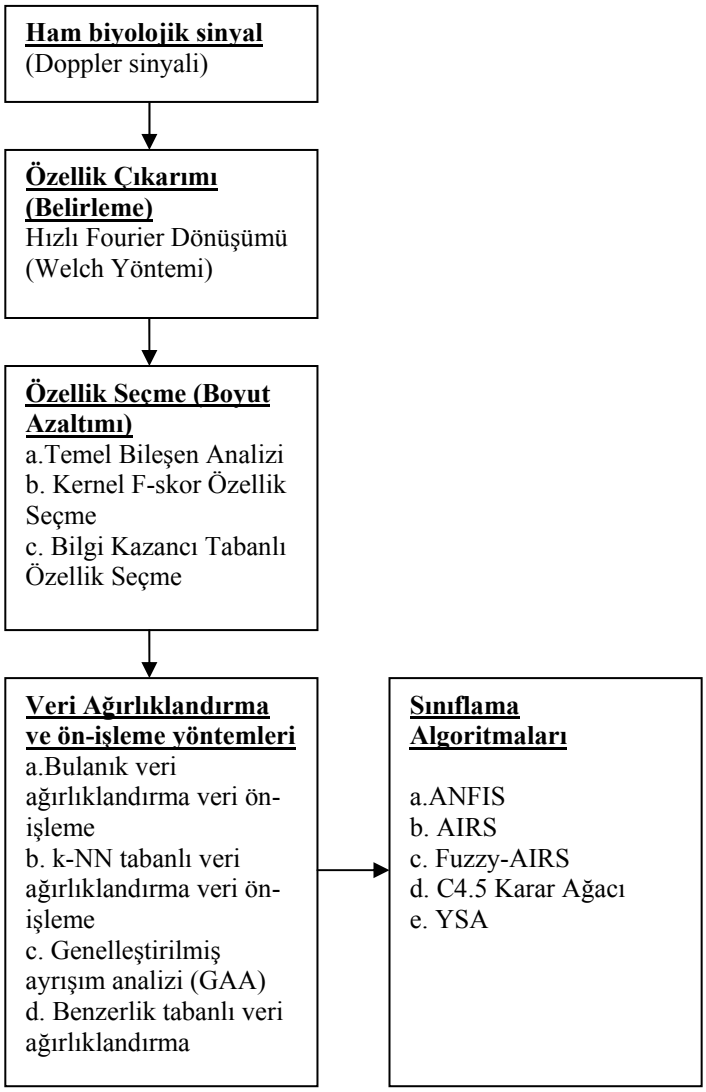
Veri ön-işleme, danışmanlı bir makine öğrenmesi algoritmasının performansını geliştirmede önemli bir etkiye sahiptir. Gürültülü örneklerin uzaklaştırılması, eksik veri, eksik özellik, aykırı değere sahip veri kümeleri, makine öğrenmesi algoritmalarında karşılaşılan en zor problemler arasındadır. Genel olarak, veri kümesinden uzaklaşan örnekler, birçok boş özellik değerine sahip olan ve oldukça ayrık olan örneklerdir. Eksik veriyi ele alma, veri hazırlama adımlarında sıklıkla ilgilenilen diğer bir konudur (Kotsiantis ve ark. 2006).

Medikal karar verme sistemleri için geliştirilen yöntemlerin bütünü genel olarak birkaç ayrık işlem parçalarına ayrılabilir: ön-işleme, özellik çıkarımı veya seçme ve sınıflama aşamalarıdır. Sinyal veya görüntü elde etme, artifakt (insan eliyle yapılan gürültü) ayıklama, ortalama, eşiklendirme, sinyal veya görüntü geliştirme ve kenar belirlemesi, ön-işleme kapsamında kullanılan temel operatörlerdir. İşaretler (sinyaller), özellik çıkarımı modülü tarafından işlenir. Özellik seçme modülü, özellik vektörünün sadece boyutta azaldığı isteğe göre seçilen bir konumdur. Sınıflandırma modülü, otomatik teşhiste son konumdur. Bu modül, giriş özellik vektörünü inceler ve kendi yapısına göre önerici bir hipotez oluşturur. (Kordylewski et al. 2001, Kwak ve ark. 2002, Güler ve ark. 2006).

Özellik, bir örüntünün parçalanmasında yapılan farklı veya karakteristik bir ölçüm, dönüşüm, yapısal bir bileşendir. Özellikler, önemli bilginin minimum kayıplı örüntülerini göstermek için kullanılır. Bir örüntüyü açıklamak için kullanılan bütün özellik kümesinden oluşan özellik vektörü, o örüntünün azaltılmış boyutlu bir gösterimidir. Bu durum, verilen bir örüntüyü açıklamak için kullanılan bütün özellik kümesinde gerçekten belirlenenlerle sınırlıdır anlamına gelir. Boyut azaltımının amacı, yazılım ve donanım kompleksliği, hesaplama maliyeti ve örüntü bilgisini sıkıştırabilme yeteneğindeki mühendislik sınırlamalarıyla karşılaşmaktır. Ayrıca, örüntü önemli özelliklerle gösterilip basitleştirilirse, sınıflama doğruluğu daha da artacaktır (Kordylewski et al. 2001, Kwak ve ark. 2002, West ve ark. 2000).

Gerçek dünya verilerinde, veri gösterimi, birçok özellik ile gösterilir fakat bunlardan birkaçı sadece hedef içerikle ilgili olabilir. Veri kümesindeki özelliklerin birbirleriyle ilişkili olması fazlalık olabilir. Böylece, veri kümesinin sınıflandırılmasında bütün özelliklerinin içerilmesine gerek yoktur (Guyon, 2003). Özellik alt küme seçimi, mümkün olduğu kadar verilerdeki ilgisiz ve gereksiz bilgi içerenleri tanımlama ve uzaklaştırma işlemidir. Bu işlem, verinin boyutunu azaltır ve öğrenme algoritmalarının daha hızlı ve daha verimli işlem yapmasına izin verebilir. (Kotsiantis ve ark. 2006).

Bu tez çalışmasındaki amaç, veri ön-işleme yöntemleri, boyut azaltımı veya özellik seçme yöntemleri ve sınıflama algoritmaları kullanılarak medikal verileri en iyi sınıflayan modeli bulmaktır. Kullanılan ve önerilen veri ön-işleme ve ağırlıklandırma yöntemleri, *bulanık ağırlıklandırma ön-işleme*, *k-NN (k- en yakın komşu) tabanlı veri ağırlıklandırma ön-işleme*, g*enelleştirilmiş ayrışım analizi (Kernel Fisher Ayrışım Analizi)* ve *Benzerlik Tabanlı Veri Ağırlıklandırma Ön-işleme* yöntemleri kullanılmış ve önerilmiştir. Bu yöntemler arasında, özellik seçme olarak, *bilgi kazancına dayanan özellik seçme algoritması* ve *Kernel F-skor özellik seçme* yöntemleri kullanılmıştır. Boyut azaltımı olarak da *temel bileşen analizi* kullanılmıştır. Kullanılan medikal veri kümeleri özelliği belirlenmiş olanlar Kalp hastılığı veri kümesi, SPECT görüntüleri kalp hastalığı veri kümesi ve Promoter gen dizileri veri kümeleridir. Özelliği çıkarılmış olan medikal veri kümeleri ise Atherosclerosis (damar sertliği) veri kümesi, Macular hastalığı veri kümesi ve Optik Sinir hastalığı veri kümeleridir. Kullanılan medikal veri kümeleri, normal ve anormal (hasta) olmak üzere iki sınıfa sahiplerdir. Tez çalışmasında kullanılan örüntü tanıma sisteminin genelleştirilmiş blok diyagramı Şekil 1.1’ de verilmiştir.



Şekil 1.1 Tezde kullanılan örüntü tanıma sisteminin blok diyagramı

Bu blok diyagramı beş aşamadan oluşmaktadır. İlk aşamada, işlenecek veya özelliği çıkarılacak ham bir biyolojik sinyal kullanılmaktadır. Tez çalışmasında, Damar sertliğini tespit etmek için kullanılan Doppler sinyalleri biyolojik işaret olarak kullanılmıştır. Hasta ve sağlıklı kişilerden alınan bu Doppler sinyalleri zaman ekseninde kaydedildiği için özellik taşıyan bilgiler içinde gömülü bulunmaktadır. O yüzden bu sinyallerin spektrum analiz yöntemleri ile Dopppler sinyallerini zaman ekseninden frekans eksenine taşımak gerekir. Bu aşamada, hızlı Fourier dönüşümü (Welch yöntemi) kullanılarak zaman ekseninde alınmış Doppler sinyalleri frekans eksenine taşınarak bu sinyallerden özellikler çıkarılabilir. Tezde kullanılan diğer medikal veri kümelerine özellik çıkarımı işlemi uygulanmamıştır. Çünkü bu medikal

veri kümelerin özellikleri çıkarıldığı veya belli olduğundan bu veri kümelerine direk olarak Şekil 1.1' de gösterilen diyagramdaki üçüncü aşamada kullanılan metodlar kullanıldı. Özellik seçme yöntemleri, veri kümelerinde bulunan ilgisiz ve gereksiz özellikleri uzaklaştırmak ve sınıflandırıcının hesaplama maliyetini azaltmak ve sınıflama performansını arttırmak için uygulanır. Dördüncü aşamada bulunan veri ağırlıklındırma ve ön-işleme yöntemleri, non-lineer veri dağılımına sahip medikal veri kümelerini lineer olarak ayrılabilir bir veriye dönüştürmek ve veri kümelerinde bulunan aykırı değerler ile gürültülü verileri ayıklamak için uygulanır. Son aşamada ise, çeşitli ön-işlemelerden geçmiş medikal veri kümelerini sınıflamak için çeşitli sınıflama algoritmaları kullanılmıştır.

Bu tez çalışmasındaki diğer bir amaç da, veri ön-işleme ve ağırlıklandırma yöntemleri ile özellik seçme yöntemlerinin, medikal veri kümelerini sınıflamada hangi veri kümelerinde iyi sonuç vereceğini bulmaktır ve bu sayede bir medikal veri kümesini sınıflamada kullanılacak uygun model belirlenmektedir. Veri ön-işleme teknikleri ve özellik seçme (boyut azaltımı) algoritmalarının performanslarını değerlendirmek için çeşitli sınıflama algoritmaları ile hibrid olarak kullanılmışlardır. Bu sınıflama algoritmaları, *ANFIS (Adaptive Neuro-Fuzzy Inference System- Adaptif Sinirsel Bulanık Çıkarım Sistemi)*, *C4.5 karar ağacı sınıflama algoritması*, *YBTS (Yapay Bağışıklık Tanıma Sistemi)*, *bulanık mantık tabanlı kaynak dağılım mekanizmalı YBTS* ve *yapay sinir ağlarıdır*. Veri ön-işleme teknikleri ve özellik seçme (boyut azaltımı) algoritmalarının performanslarını değerlendirmek için, sınıflama doğruluğu, hassasiyet ve açıklık değerleri ve ROC (Receiver Operating Characteristics) eğrisi altında kalan alan değerleri verilmiş ve bu değerlere göre mukayese edilmişlerdir. Veri ağırlıklandırma yöntemleri arasında en iyi sonuçları veren yöntem, *k-NN tabanlı veri ağırlıklandırma ön-işleme* yöntemidir. Özellik seçme veya boyut azaltımı yöntemleri arasında ise *Temel Bileşen Analizi* (*TBA*) diğer yöntemlere göre üstün sonuçlar sağlamıştır. Ayrıca, veri ağırlıklandırma ön-işleme yöntemleri, özellik seçme algoritmaları ve sınıflama algoritmaları birleştirilerek hibrid sistemler oluşturulmuş ve her bir medikal veri kümesi için en iyi hibrid sistem seçilmiştir.

1.1. Literatür Araştırması

Örüntü tanıma ve makine öğrenmesi alanında, veri ağırlıklandırma ve boyut azaltımı veya özellik seçme yöntemleri ile ilgili literatürde çok çeşitli çalışmalar mevcuttur. Bu yöntemler, biyomedikal veri sınıflama, kümele, yazı tanıma, el yazısı tanıma, yüz tanıma, parmak izi tanıma, anormal durum teşhisi gibi uygulamalarda kullanılmıştır. Veri ağırlıklandırma ve boyut azaltımı yöntemleri, sınıflama işleminden önce yapılan bir işlemdir. Lineer olmayan veri dağılımına sahip, eksik veri değere sahip olan ve özellikleri arasında lineer bir ilişki olan veri kümelerini daha kolay sınıflandırılabilir bir hale getirmek için veri ön-işleme tekniklerinin uygulaması gerekir. Bu sayede hem daha iyi sınıflama performansını elde edilebilir hem de hesaplama maliyeti azaltılabilir. Literatür araştırmasında, tezde kullanılanılan medikal veri kümelerinin sınıflaması, otomatik teşhis sistemleri, veri ağırlıklandırma ve boyut azaltımı veya özellik seçme yöntemleri ile ilgili yapılan çalışmalara yer verilmiştir.

İlk olarak, literatürde var olan yapay bağışıklık tabanlı sınıflayıcılar ve sistemler aşağıda özetlenmiştir.

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

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

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

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

Kalp hastalığı veri kümesinin sınıflamasında ve teşhisinde, literatürde aşağıdaki çalışmalar yapılmıştır.

El-Hanjouri ve ark. (2002), kalp sesleri kullanarak kalp hastalıklarının teşhisi için bir teşhis sistemi önerdiler. Özellik çıkarımı için, dalgacık ayrışımı ve Mel-frekans cepstrum yöntemlerini kullandılar. Gizli Markov modeli kullanarak farklı kalp hastalıklarını sınıflandırdılar.

Yan ve ark. (2003), beş önemli kalp hastalığının teşhisi için bir karar destek sistemi geliştirmek için Çok Katmanlı Perseptron (*ÇKP*) sinir ağına dayanan hesaplamalı bir model önerdiler. Önerilen sistem üç katmandan oluşmaktadır. Sistemin giriş katmanı, geniş sayıda hasta durumundan belirlenen 38 giriş değişkeni içerir. Gizli katmandaki düğümlerin sayısı, kaskad bir öğrenme işlemiyle belirlenir. Çıkış katmanındaki 5 düğümün her biri, ilgili kalp hastalığından birisine uygundur. Önerilen karar destek sistemini, momentum oranlı, adaptif öğrenme oranı ve hatırlamalı mekanizmalarla donatılmış bir geri yayılım algoritmasıyla eğittiler. Deneysel sonuçlar, adapte edilen *ÇKP* tabanlı bir karar modeli, kalp hastalıklarının sınıflamasında yüksek doğruluk değerleri elde edildiğini göstermiştir.

Şahan ve ark. (2005), network (ağ) tabanlı yapay bağışıklık sistemlerinde sıklıkla kullanılan şekil uzay gösteriminde Euclidean mesafesinin hesaplamasınında karşılaşılan negatif etkileri uzaklaştırmak için Özellik Ağırlıklandırmalı Yapay

Bağışıklık Sistemi sınıflama algoritması önerdiler ve kalp hastalığı veri kümesine uyguladılar ve bu önerilen sistemle %82.59 sınıflama doğruluğu elde ettiler.

Zheng ve ark. (2006), beş önemli kalp hastalığının teşhisi için, *ÇKP* gruplamasına dayana komite (gruplama) makinalarını önerdiler. Özellikle, rastgele alt uzay ve bagging (gruplama yöntemi) yöntemlerini içeren iki gruplama yöntemi düşündüler. 352 örnekli kalp hastalığı veri tabanı, önerilen sistemlerin performanslarını değelendirmek için kullandılar. *ÇKP* gruplamalı komite makinaları, tek *ÇKP*' dan daha iyi performans elde etmiştir.

Polat ve ark. (2006a), kalp hastalığını sınıflamak için bulanık mantık tabanlı veri ağırlıklandırma ön-işleme ile YBTS sınıflama algoritmalarını birleştirdiler ve 10 kat kez çaprazlama kullanılararak kalp hastalığı teşhisinde %96.30 sınıflama doğruluğu elde ettiler.

Polat ve ark. (2007), kalp hastalığını teşhis etmek için *bilgi kazancına dayanan özellik seçme algoritması (BKÖS)*, *bulanık ağırlıklandırma ön-işleme (BAÖ)* ve *yapay bağışıklık tanıma sistemlerine (YBTS)* dayanan hibrid bir sistem önerdiler ve %50-50 eğitme ve test veri kümesi ayırımı ile %92.59 sınıflama doğruluğu elde ettiler.

Polat ve ark. (2007a), *k-NN tabanlı veri ağırlıklandırma ön-işleme yöntemi*, bulanık kaynak dağılım mekanizmalı YBTS sınıflama algoritması ile birleştirerek bir uzman sistem önerdiler ve kalp hastalığı veri kümesininin sınıflamasına uyguladılar. Oluşturdukları uzman sistemle %87 sınıflama doğruluğu elde ettiler.

Kahramanlı ve ark. (2008), yapay sinir ağları ile bulanık sinir ağlarını birleştiren hibrid bir sistem önerdiler ve kalp hastalığı teşhisi problemine uyguladılar. Bu hibrid sistemle %86. 80 sınıflama doğruluğu elde ettiler.

Özşen ve ark. (2008), yapay bağışıklık sınıflama sistemi tasarımında sistem birimleri arasındaki mesafeyi ve uygunluğu hesaplamak için çeşitli ölçüm kriterleri kullandılar ve bu oluşturdukları algoritmaları kalp hastalığı sınıflamasına uyguladılar. Kullandıkları mesafe ölçümleri, Euclidean mesafesi, Manhattan mesafesi ve hibrid benzerlik ölçümleridir. Elde edilen sınıflama doğrulukları, %83.21, %80.74% ve %83.95 sonuçlarını yukarıdaki ölçüm kriterleri ile elde edilmiştir.

E. coli promoter gen dizileri veri kümesinin sınıflandırılmasında literatürde aşağıdaki çalışmalar yapılmıştır.

Geoffrey G. Ve Towell ve ark. (1990), promoter gen dizileri veri kümesini oluşturmuşlar ve standart geri yayılım algoritması, O' Neill, En yakın komşu ve ID3 algoritmaları kullanarak promoter gen dizilerini sınıflamışlardır. Yukarıdaki yöntemler kullanarak elde ettikleri sınıflama doğrulukları sırasıyla %92.45, %88.67, %87.73 ve %82.07' dır.

Zhang ve ark. (2006), E.coli promoter' u tahmin etmek için ileri beslemeli yapay sinir ağlarının gösterimini açıkladılar. Dizi korumuna göre, 60 tabanlı diziler, pozitif örnek olarak seçilirler ve E.coli kodlama alanlarından uygun non-promoterlar, negatif örnek olarak seçilirler ve daha sonra ileri beslemeli yapay sinir ağlarına dayanan sınıflayıcı eğitilir. Elde ettikleri sonuçlar, ileri beslemeli sinir ağları verimli bir şekilde promoterların istatiksel karakteristiklerini belirleyebileciğini gösterdiler.

Akdemir ve ark. (2007), promoter gen dizilerini sınıflamak için özellik seçme, bulanık ağırlıklandırma ön-işleme ve C4.5 karar ağacına dayanan hibrid bir sistem önerdiler. C4.5 karar ağacı, bulanık ağırlıklandırma ön-işleme ve C4.5 karar ağacınını birleşimi, özellik seçme ve C4.5 karar ağacınını birleşimi ve özellik seçme-bulanık ağırlıklandırma ön-işleme ve C4.5 karar ağacınını birleşimin elde edilen sınıflama sonuçları sırasıyla 70.08%, 88.88%, 76.82% ve 93.33' dür.

Polat ve ark. (2007c), bilgi kazancına dayanan özellik seçme ve en az kareler destek vektör makinalarını kaskad birleştirerek promoter gen dizilerinin sınıflaması problemine uyguladılar ve 10 kat çapraz ile %100 sınıflama doğruluğu elde ettiler.

Polat ve ark. (2007d), bilgi kazancına dayanan özellik seçme ve Fuzzy-AIRS sınıflandırıcıyı birleştirerek yeni bir hibrid sistem önerdiler ve promoter gen dizilerinin sınıflamasına uyguladılar. Fuzzy-AIRS sınıflayıcı %50 doğruluk elde ederken, önerdikleri sistem %90 sınıflama başarısı elde etmiştir.

SPECT görüntüleri veri kümesinin sınıflandırılmasında literatürde aşağıdaki çalışmalar yapılmıştır.

Cios ve ark. (1996), SPECT (Single Photon Emission Computed Tomography) görüntülerini birkaç kategoride (normal, infarct, ischemia, infarct ve ischemia, reverse re-distribution, artifact ve equivocal) sınıflamak ve SPECT görüntülerini analiz etmek için yarı otomatik bir prosedür gerçekleştirdiler. Önerilen

prosedür, SPECT görüntülerinin incelenmesinde doktora yardımcı olan sistem iki adımdan oluşmaktadır. İlk adım, SPECT görüntülerini tekrar yapılandırır. Bu görüntüler, 16 bit çözünürlükle 64x64 piksellik parçalar içerir. Taranmış görüntüler, kenar belirleme, ilgili bölge ve segmantasyon teknikleri kullanılarak nümerik formata dönüştürülür. Her bir görüntüden, dikdörtgensel bölgeyi belirlemek için yeni bir algoritma geliştirdiler. İkinci adım, işlenen görüntüyü yedi sınıftan birine otomatik sınıflamayı içerir. Sınıflama işlemini gerçekleştirmek için, C4.5 karar ağacı, CLIP3 ve sınıflama kurallarını üretmek için bulanık modelleme kullanmışlardır.

Lukasz ve ark. (2001), CLIP3 makine öğrenmesi sınıflama algoritması kullanarak SPECT görüntülerinden kalp hastalığı teşhisinde %84 sınıflama doğruluğu elde ettiler.

Kurgan ve ark. (2002), CLIP4 makine öğrenmesi algoritması geliştirerek SPECT görüntülerinden kalp hastalığı tanısı problemine uyguladılar ve CLIP3' ün elde ettiği sınıflama doğruluğunu geliştirdiler. CLIP4 algoritması ile %86.10 sınıflama doğruluğu elde ettiler.

Ümit ve ark. (2004), Radial Basis Function (*RBF*) yapay sinir ağları ve General Regression Neural Network (*GRNN*) algoritmaları kullanarak SPECT görüntülerinden kalp hastalığı teşhisinde 88.24% ve %93.58 sınıflama doğrulukları elde ettiler.

Polat ve ark. (2007b), *Bağımsız Bileşen Analizi (BBA)* ile *YBTS*' yi birleştiren bir grup sınıflayıcı yöntemi önerdiler ve önerilen sistemi SPECT görüntüleri ile kalp hastalığı teşhis problemine uyguladılar. Elde ettikleri sınıflama doğruluğu %97.74 dür. Ayrıca çalışmalarında, *BBA*' dan 3, 4 ve 5 özellikleri elde ederek *YBTS* ile birleştirerek önerilen sistemle karşılaştırdılar ve *BBA(3-özellik)+YBTS* sistemi %95.49 sınıflama doğruluğu elde ederken, *BBA(4-özellik)+YBTS* sistemi %96.99 sınıflama doğruluğu ve *BBA(5-özellik)+YBTS* sistemi ise %89.47 sınıflama doğrulukları elde ettiler.

Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılması literatürde aşağıdaki çalışmalar yapılmıştır.

Dirgenali ve ark. (2006), *Atherosclerosis* hastalığının erken teşhisi için hibrid bir yöntem önerdiler. İlk olarak, Doppler sinyalinden özellik çıkarmak için *AR (Autoregressive)* modelleme uyguladılar. İkinci kısımda ise, elde edilen özellik

sayısını azaltmak için *TBA* özelliği çıkarılmış Doppler sinyali veri kümesine uyguladılar. Son aşamada ise, yeni oluştulmuş Doppler sinyali veri kümesini sınıflandırmak için *YSA* uyguladılar ve *Atherosclerosis* hastalığını %97 doğrulukla teşhis ettiler.

Polat ve ark. (2006b), *YBTS* sınıflama performansını arttırmak için bulanık kaynak dağılımı mekanizması önerdiler ve *Atherosclerosis* hastalığı teşhisine uyguladılar. İlk olarak, Doppler sinyaline *FFT* uygularak sinyalden özellik çıkarımı yaptılar. Daha sonra özelliği belirlenmiş Doppler sinyalini sınıflamak için *YBTS* ve *Bulanık-YBTS* uyguladılar ve sırasıyla %75 ve %100 sınıflama doğrulukları elde ettiler.

Ceylan ve ark. (2007), *Atherosclerosis* hastalığını sınıflamak için iki yaklaşım önerdiler. Bu yaklaşımlar, *TBA* ve *kompleks değerli yapay sinir ağlarının* birleşimi ve B*ulanık Kümele Ortalama (BKO)* ve *kompleks değerli yapay sinir ağlarının* birleşimidir. İlk önerilen sistemde, doppler sinyallerini sınıflamadan önce veri kümesine *TBA* uygulayarak yeni özellikli ve daha az boyutta Doppler sinyali veri kümesi elde ettiler ve sonra bu yeni oluşturulan veri kümesini *kompleks değerli yapay sinir ağları* ile sınıflandırdılar. İkinci önerdikleri çalışmada ise, *BKO* doppler sinyali veri kümesine uygulanarak veri azaltımı yaparak, veri sayısı azaltılmış Doppler sinyali veri kümesi kompleks değerli yapay sinir ağları ile sınıflandırdılar. 10 kat çaprazlama ile her iki çalışmada da %100 sınıflama doğruluğu elde ettiler.

Özşen ve ark. (2007), çalışmalarında yapay bağışıklık sisteminde yeni bir danışmalı sınıflama algoritması önerdiler ve bu yöntemi *Atherosclerosis* hastalığını teşhis etmek için Doppler sinyaline uyguladılar. İlk olarak, Doppler sinyalinden özellik çıkarmak için *AR (Autoregressive)* modelleme uyguladılar, daha sonra özelliği belirlenmiş olan Doppler sinyalini sınıflamak için *Danışmalı Duyarlılık Olgunlaşma Algoritması (DDOA)* kullanarak 10 kat çaprazlama ile %98.93 sınıflama doğruluğu elde ettiler.

Latifoğlu ve ark. (2007a), çalışmalarında, *Yapay Bağışıklık Sistemine Dayanan Özellik Ağırlıklı Yapay Bağışıklık sistemi (AWAIS)* ile gerçek dünya verisi olan Doppler sinyalinden *Atherosclerosis* hastalığını teşhis ettiler ve 10 kat çaprazlama ile %99.33 sınıflama başarısı elde ettiler.

Latifoğlu ve ark. (2007b), *yapay bağışıklık tanıma sistemi* kullanarak hasta ve normal kişiye ait doppler sinyali ile *Atherosclerosis* hastalığını teşhis etmişlerdir ve 10 kat çaprazlama ile %99.29 sınıflama doğruluğu elde etmişlerdir.

Özbay ve ark. (2007a), *kompleks dağerli dalgacık yapay sinir ağları* önerdiler ve yöntemlerini doppler sinyallerinin sınıflamasına uyguladılar. Doppler sinyalleri veri kümesi, sağlıklı ve *Atherosclerosis* hastalıklı kişilerin doppler sinyallerinin güç spektrum yoğunlukları oluşturulduktan sonra elde edilen özellik vektöründen oluşmaktaktadır. Bu önerilen sistem, *Atherosclerosis* hastalığını leave-one-out (birisi dışarıda kalan) çaprazlama ile %100 sınıflama doğruluğu elde etmişlerdir.

Özbay ve ark. (2007b), Doppler sinyallerini farklı pencere tiplerine sahip *Fast Fourier Transform (FFT)*, *Hilbert dönüşümü* ve *Welch yöntemiyle* işlediler. Bu özellik çıkarımından sonra, kompleks değerli yapay sinir ağları ile Doppler sinyallerini sınıfladılar ve ön-işlemede (özellik çıkarımı) kullanılan pencere fonksiyonlarının sınıflamaya olan etkisini incelediler.

Kara ve ark. (2007a), çalışmalarında Doppler sinyallerine *Ayrık Dalgacık Dönüşümü (ADD)* uygulayarak sinyali ayrıştırdılar ve her bir ayrışan parçasına *Welch yöntemi* uygulayarak Doppler sinyalinden özellik çıkardılar. Bu işlemden sonra, elde edilen özellik vektörüne *TBA* uyguladılar ve Doppler sinyali veri kümesinin boyutunu azalttılar. *TBA*' dan sonra, yeni elde edilen Doppler sinyaline *yapay sinir ağları* uygulayarak *Atherosclerosis* hastalığı ve sağlıklı kişi olarak Doppler sinyalini sınıflandırdılar.

Latifoğlu ve ark. (2008a), *Atherosclerosis* (damar sertliği) hastalığını teşhis etmek için *FFT*, *TBA*, *k-NN tabanlı veri ağırlıklandırma ön-işleme* ve *YBTS* ' ye dayanan hibrid bir sistem önerdiler ve *Atherosclerosis* hastalığının teşhisinde %100 sınıflama doğruluğu elde ettiler.

VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılması literatürde aşağıdaki çalışmalar yapılmıştır.

Kara ve ark. (2006), çalışmalarında PERG sinyallerinden optik sinir hastalığını sınıflamak için *çok katmanlı yapay sinir ağlarını* kullandılar. Optik sinir hastalığına sahip olan ile sağlıklı kişiye ait PERG sinyallerinin ayrımında %94.2 başarı oranı elde ettiler.

Kara ve ark. (2007b), pattern electroretinography (PERG-örüntü retinografisi) sinyallerinden *TBA* ile Macular ve optik sinir hastalığını tespit ettiler. Çalışmalarında, Macular hastalıklı PERG sinyalini ve optik sinir hastalıklı PERG sinyaline *TBA* uygulayarak bu sinyallerin birinci ve ikinci temel bileşenlerini incelediler ve birinci temel bileşenin bu iki hastalığın biriinden ayrılmasında kullanılabileceğini gösterdiler.

Kara ve ark. (2007c), PERG sinyallerinden optik sinir hastalığını teşhis etmek için *Öğrenme Vektör Nicelemeli (ÖVN) yapay sinir ağlarını* kullandılar. *ÖVN li YSA* ile elde edilen sınıflama başarısı %92 olarak elde ettiler.

Polat ve ark.(2008a), PERG sinyallerinden hem Macular hem de optik sinir hastalığını teşhisinde, *en az kareler destek vektör makinaları*, *C4.5 karar ağaçları* ve *YBTS* sınıflandırıcıları kullandılar ve karşılaştırdılar. 10 kat çaprazlama kullanılarak *C4.5 karar ağacı*, *en az kareler destek vektör makinaları* ve *YBTS* sınıflayıcılarından sırasıyla %85.9, %100 ve %81.82 sınıflama doğruluğu elde ettiler.

Güven ve ark. (2008), Visual Evoked Potentials (VEP-görsel uyarılmış potansiyeller) sinyallerinden optik sinir hastalığını teşhis etmede *Genelleştirilmiş Ayrışım Analizi (GAA)*' nin sınıflandırma doğruluğu etkisini incelediler. Önerdikleri sistem iki aşamadan oluşmaktadır. İlk aşamada sınıflamadan önce, veri ön-işleme olarak VEP sinyallerine *GAA* uyguladılar. İkinci aşamada ise, *GAA* ile ayrışımı kolay hale getirilen VEP sinyallerini sınıflamak için *C4.5 karar ağacı*, *yapay sinir ağları*, *YBTS*, *lineer ayrışım analizi* ve *destek vektör makinaları* kullanıldı. *GAA*' sız elde edilen sınıflama doğrulukları *C4.5 karar ağacı*, *yapay sinir ağları*, *YBTS*, *lineer ayrışım analizi* ve *destek vektör makinaları* sınıflama algortitmaları kullanılanılarak sırasıyla %84.37, %93.75, %75, %76.56% ve %53.125' dir. *GAA* ön-işlemeli elde edilen sınıflama doğrulukları ise sırasıyla %93.75, %93.86, %81.25, %93.75 ve %93.75 oldu.

PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılması literatürde aşağıdaki çalışmalar yapılmıştır.

Güven ve ark. (2006), PERG sinyallerinden Macular hastalığını teşhis etmek için *çok katman perseptronlu yapay sinir ağları* kullandılar ve %98 sınıflama doğruluğu elde ettiler.

Polat ve ark. (2007e), Macular hastalığını sınıflamak için *k-NN tabanlı veri ağırlıklandırma ön-işleme* ve *en az kareli destek vektör makinalarını* birleştiren hibrid bir yöntem önerdiler. Ayrıca, *k-NN tabanlı veri ağırlıklandırma ön-işlemedeki k* değerinin sınıflamaya olan etkisi de incelendi ve *k*' nın 10, 15 ve 20 değerleri için 10 kat çaprazlama ile %100 sınıflama doğruluğu elde ettiler. *k-NN tabanlı veri ağırlıklandırma ön-işleme* kullanılmadan elde edilen başarı ise %90.91 oldu.

Polat ve ark.(2008b), PERG sinyalerinden Macular hastalığını teşhis etmek için ön-işleme metodu olarak *sınıfa bağımlı özellik seçme algoritması* ve *bulanık veri ağırlıklandırma ön-işleme* yöntemlerini önerdiler ve *C4.5 karar ağacı sınıflama* algoritması ile birleştirdiler. Önerdikleri sistemin performansını ölçmek için 5, 10 ve 15 kat çaprazlama ile sınıflama doğruluğunu kullandılar. 5, 10 ve 15 kat çaprazlama ile %96.22, %96.27% ve %96.30 sınıflama doğrulukları elde ettiler.

Literatürde var olan diğer veri ön-işleme yöntemleri aşağıda özetlenmiştir.

Baudat ve ark. (2000), kernel fonksiyon operatörü kullanarak non-lineer ayrışım analizi ile ilgilenmek için *Genelleştirilmiş Ayrışım Analizi* (*Generalized Discriminant Analysis* -*GDA*) olarak adlarılan yeni bir yöntem önerdiler. Temel olarak alınan teori, *GDA* giriş vektörlerini yüksek boyutlu özellik uzayında haritalamayı sağladığı için destek vektör makinalarına yakındır. Dönüştürülmüş uzayda, lineer özellikler giriş özelliğinin genişlemesini kolaylaştırır ve klasik lineer ayrışım analizini non-lineer ayrışım analizine genelleştirir. Bu formülasyon, bir özdeğer problem çözümü olarak açıklanabilir.

Birçok faktör, verilen bir görevde makine öğrenmesinin başarısını ektiler. Örnek verinin gösterimi ve kalitesi ilktir ve başta gelir. Gürültülü veya güvensiz bilgiyi gösteren çok fazla ilgisiz ve fazla bilgi varsa, eğitme sırasında bilgi keşfi çok zordur. Çok iyi biliniyor ki, veri hazırlama ve filtreleme adımları makine öğrenmesi problemlerinde oldukça yüksek bir zaman alır. Veri ön-işleme, veri temizleme, normalizasyon, veri dönüşümü, özellik çıkarımı ve özellik seçmeyi içerir. Kotsiantis ve ark. (2006), makine öğrenmesi problemlerinin çözümünde bu veri ön-işleme algoritmalarını incelemişlerdir.

Cao ve ark. (2007), sınıflandırma ve veri analizi için en ayırt edici ve bilgi verici özellikleri seçmek için bir kernel uzayında özellik seçme problemini gösterdiler. Bu problemi çözmede, özellik seçmek için ilk adım olarak kernel

uzayında temel bir küme oluşturdular. Bu küme kullanarak, birçok özelliğin bağımlı olduğu zaman bile etkisini kanıtlayan marjin tabanlı özellik seçme algoritmasını genişlettiler.

Son zamanlarda, non-lineer boyut azaltımına olan ilgi arttı. Bu, kompleks düşük boyutlu veriyle ilgilenme yeteneğine sahip olduğu iddaa edilen değişik yeni non-lineer tekniklerin önerilmesine neden oldu. Maaten ve ark. (2007), boyut azaltımı için on non-lineer tekniği incelediler. Yapay ve gerçek dünya verilerinde boyut azaltımı için bu non-lineer tekniklerin performanslarını incelediler ve iki temel lineer teknikle (*Temel Bileşen Analizi ve Lineer Ayrışım Analizi*) performanslarını karşılaştırdılar.

Tez çalışmasında, boyut azaltımı algoritmalarında *temel bileşen analizi*, *bilgi kazancına dayanan özellik seçme* ve *Kernel F-skor özellik seçme* algoritmaları kullanılmıştır. Veri ağırlıklandırma ve ön-işleme için *bulanık ağırlıklandırma ön-işleme*, *k-NN tabanlı veri ağırlıklandırma ön-işleme*, *genelleştirilmiş ayrışım analizi* ve *benzerlik tabanlı veri ağırlıklandırma ön-işleme* önerilmiş ve kullanılmıştır. Bu ön-işleme algoritmalarının biyomedikal verileri sınıflamada performanslarını görmek ve birbirleriyle karşılaştırmak için çeşitli sınıflama algoritmaları kullanılmıştır.

1.2. Çalışmanın Temel Amacı ve Literatüre Katkıları

Bu çalışmanın amacı, özelliği belirlenmiş ve özelliği çıkarılmış olan biyomedikal verileri; sınıflamak, sınıflama performasınlarını geliştirmek için çeşitli veri ön-işleme ve boyut azaltımı (veya özellik seçme) algoritmaları geliştirmek ve bu yöntemler kullanılarak en iyi hibrid sistemi bulmaktır.

Medikal hastalıkları sınıflandırırken hastalığı belirten veya açıklayan bir veri kümesine ihtiyaç vardır. Bu veri kümesi de hasta ve normal (sağlıklı) kişilerden alınan sinyaller, görüntüler, klinik veriler, laboratuvar sonuçlarından oluşabilir. Oluşturulan bu medikal veri kümesi, hasta ve normal kişi verilerinin birbirine karıştığı ve ayırt edilmesi zor bir durum olabilir. Bu duruma sahip olan veri kümelerine non-lineer veri kümeleri denilmektedir. Non-lineer dağılıma sahip olan veri kümelerini sınıflamak için veri ağırlıklandırma (ön-işleme) ve boyut azaltımı

algoritmaları kullanılarak medikal veri kümeleri daha kolay ayrıştırılabilir hale getirilmiştir.

Veri ön-işleme teknikleri; kalp hastalığı veri kümesi, SPECT görüntüleri kalp hastalığı veri kümesi, promoter gen dizileri veri kümesi, *Atherosclerosis* hastalığı veri kümesi, Macular (Sarı Nokta Hastalığı) veri kümesi ve optik sinir hastalığı veri kümelerine uygulanmıştır.

Literatüre katkıda bulunulan ve önerilen veri ön-işleme teknikleri, *bulanık veri ağırlıklandırma ön-işleme*, *k-NN tabanlı veri ön-işleme*, ve *benzerlik tabanlı veri ağırlıklandırma ön-işleme* yöntemleridir. Özellik seçme yöntemi olarak *Kernel F-skor özellik seçme* yöntemi önerilmiştir.

1.3. Çalışmanın Kapsamı ve Organizasyonu

Biyomedikal verilerin sınıflandırılmasına sınıflandırıcının performans gelişimini arttıran veri ön-işleme tekniklerine odaklanan bu tez çalışmasının bölümleri aşağıdaki gibi organize edilmiştir:

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

İkinci bölümde, özellik çıkarma ve özellik seçme yöntemleri anlatılmıştır. Özellik çıkarma işleminde, *Atherosclerosis* hastalığının teşhisinde Doppler sinyallerinden özellik çıkarmada kullanılan ayrık Fourier dönüşümü spektral analiz yöntemi anlatılmıştır. Özellik çıkarma işlemi sadece Doppler sinyallerinde uygulanmıştır. Özellik seçme yöntemleri ise özelliği belirlenmiş veya özelliği çıkarılmış biyomedikal veri kümelerindeki ilgisiz ve birbirleri ile ilişkili olan özellikleri uzaklaştırmak veya veri kümesini yeni bir uzaya taşımak için kullanılır. Bu yöntemler, *temel bileşen analizi*, *bilgi kazancına dayanan özellik seçme algoritması* ve *Kernel F-skor özellik seçme* metodlarıdır.

Üçüncü bölümde, özelliği azaltılmış veri kümesini daha ayrıştırılabilir hale dönüştürmek için çeşitli veri ön-işleme ve ağırlıklandırma algoritmaları önerildi ve kullanıldı. Beş veri ön-işleme yöntemi anlatılmıştır.

Dördüncü bölümde ise biyomedikal verileri sınıflamak için kullanılan sınıflama algoritmaları verilmiştir. Bu sınıflama algoritmaları, *C4.5 karar ağacı*, *YSA*, *YBTS*, *bulanık kaynak mekanizmalı YBTS* ve *ANFIS*' dir.

Beşinci bölümde, tez çalışmasında kullanılan biyomedikal veri kümeleri verilmiştir. Bu veri kümelerinden kalp hastalığı veri kümesi, SPECT görüntülerinden kalp hastalığı teşhisi veri kümesi ve E. Coli promoter gen dizilerinin tahmini ile ilgili veri kümeleri "The UCI KDD Archive Information and Computer Science University of California" veri tabanından alınmıştır. *Atherosclerosis* hastalığı veri kümesi, Macular hastalığı veri kümesi ve Optik Sinir hastalığı veri kümeleri ise Fatih Üniversitesi Elektrik-Elektronik Mühendisliği Bölümü Öğretim Üyesi Prof. Dr. Sadık KARA ve Erciyes Üniversitesi Mühendislik Fakültesi Elektronik Mühendisliğinde lisanüstü tez çalışmasını yaptırdığı öğrencilerden alınmıştır.

Altıncı bölümde ise geliştirilen veri ön-işleme teknikleri ve boyut azaltımı algoritmalarının biyomedikal verileri sınıflamadaki performans sonuçları verilmiştir. Ayrıca kullanılan ön-işleme tekniklerinden biyomedikal verileri en iyi sınıflayan bir hibrid sistem önerilmiş ve elde edilen sonuçlarla üstünlüğü vurgulanmıştır. Sınıflayıcılar arasındaki performansı değerlendirmek için de sınıflama doğruluğu, hassasiyet ve açıklık analizi ve ROC eğrisi altında kalan alan değerleri verilmiştir.

Yedinci bölüm, tez çalışması ile ilgili genel sonuçları içermektedir. Kaynaklar bölümünde ise bu tez çalışmasında faydalanılan kaynaklara yer verilmektedir.

2. ÖZELLİK ÇIKARMA VE ÖZELLİK SEÇME YÖNTEMLERİ

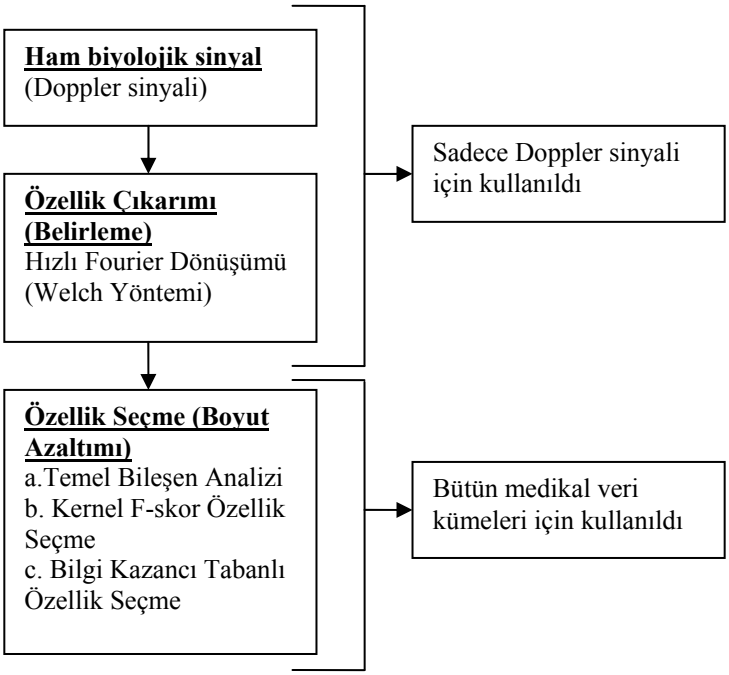
Bu bölümde, özellik çıkarma ve özellik seçme (boyut azaltımı) yöntemleri örnek uygulamalarla anlatılmıştır. Özellik çıkarma işleminde, *Atherosclerosis* (damar sertliği) hastalığının teşhisinde Doppler sinyallerinden özellik çıkarmada kullanılan *Ayrık Fourier Dönüşümü* spektral analiz yöntemi kullanılmıştır. Doppler sinyallerden özellik çıkartmak için Welch metodu ve güç spektrum yoğunluklarının elde edilişi ayrıntılı olarak anlatılmıştır.

Özellik seçme ve boyut azaltımı, özellik çıkarımından sonra oluşturulan veri setinin boyutu fazla olabilir veya veri setinde ilgisiz ve veri dağılımını bozan özellikler olabilir. Bu özellikler hem sınıflama performansını azaltır hem de sınıflama süresini arttırır. Bu yüzden özellik seçme algoritmaları, medikal veri kümelerinin sınıflandırılmasında daha yüksek genelleştirme yeteneği ve örüntü tanıma problemlerinin çözümünde daha iyi sonuçlar sunarlar.

Özellik seçme; örüntü tanıma, istatistik ve veri madenciliği konularında aktif bir araştırma konusudur. Özellik seçmenin temel amacı, çok az veya tahmin edilemeyen özelliklerin elimine edilmesiyle giriş özelliklerinin bir alt kümesini seçmektir (Wang 2003).

Tahmin edici özelliklerden doğru alt kümelerinin bulunması önemli bir problemdir. Örnek olarak, uzman, tehlikeli bir operasyonun tedavi için gerekli olup olmadığına seçilen özelliklere dayanarak bir karar verebilir (Wang 2003). Özellik seçme danışmalı ve danışmasız olmak üzere iki şekilde olabilir (Wang 2003). Bu tez çalışmasında, özellik seçme ve boyut azaltımı yöntemleri olarak, *temel bileşen analizi*, *bilgi kazancına dayanan özellik seçme* ve *Kernel F-skor özellik seçme* yöntemleri kullanılmıştır.

Şekil 2.1' de özellik çıkarımı ve özellik seçmeyi blok gösteren bir şekil verilmiştir. Bu blok diyagramda, özellik çıkarımı işlemi sadece Doppler sinyali için uygulanmıştır. Özellik seçme yöntemleri ise tezde kullanılan bütün medikal veri kümeleri için kullanılmıştır.



Şekil 2.1 Özellik çıkarımı ve özellik seçmeyi gösteren blok diyagram

2.1. Özellik Çıkarma İşlemi ve Welch Metodu

Bu bölümde, Doppler sinyallerini segmentlere ayırmak için kullanılan pencereleme işlemi ve sinyalin içerdiği frekans bileşenlerini gösteren ve güç spektral yoğunluğunu elde edebilmek için kullanılan Welch metodu anlatılmıştır.

2.1.1. Pencereleme İşlemi

Zaman fonksiyonunun bir veri segmentiyle çarpılmasına veri pencerelemesi denir. Pencereleme işlemi temel olarak, veri segmentinin FFT spektrumu hesaplanmadan önce kullanılır ve amacı sonuç spektrumunun şeklini düzeltmektir. Diğer bir amacı da spektral sızma ve lekelenmelerin zararlı etkilerini uygun bir pencere yardımıyla en aza indirmektir (Hayes 1996).

En basit haliyle pencereleme işlemi, ilk olarak bir segment, bir dalga şeklinden kesilir. Bu uygulama, bütün veriye dikdörtgen pencerenin uygulanması ile aynıdır. Analiz sadece dalganın pencerelenmiş bölümü ile sınırlıdır. Dikdörtgen dışındaki pencere şekilleri, istenilen şekille dalganın zaman domeninde çarpılmasıyla elde edilir. Bunun yanında pencereleme, hem zaman domeninde hem de frekans domeninde yapılabilir. Çünkü zaman domeninde çarpma frekans domeninde konvolüsyon işlemi ile aynıdır. Frekans domeninde pencereleme, frekans domen penceresi ile sinyal spektrumunun konvolüsyonu şeklinde uygulanır (Hayes 1996, Semmlov 2004).

Pencere şekilleri, incelenen fonksiyonlar olarak da sunulabilir. Bunun yanı sıra, pencerenin dışında kalan noktaların doğru olmasa da sıfır olduğu varsayılır. Güç spektrumu yaklaşımında pencere seçimi önemli bir faktördür. Çünkü bir veri seçildiği zaman pencerenin frekans karakteristiği spektral sonucun bir parçası olabilir. Bütün pencereler iki tip artifakt (dışarıdan istenmeyen sinyal) üretir: analob (mainlob) ve yanlob (sidelob) (Semmlov 2004). Pencere tarafından üretilen artefakt pencerenin Fourier dönüşümü alınarak bulunabilir. Gerçek spektrum analob denilen artefakt ile genişletilir ve yan lob adı verilen ek pikler üretilir. Eğer işlem, yakın frekans uzayındaki iki darbantlı sinyalin çözümü ise en dar bantlı pencere tercih edilir. Bir güçlü ve bir zayıf sinyal söz konusu ise güçlü sinyallerin yan loblarının zayıf sinyaller tarafından yokedilmesini önlemek için hızla azalan yan lobları olan pencereler tercih edilir. Eğer iki tane orta derecede güçlü sinyal var ise orta derecede dar analob ve orta derecede zayıflayan loblar en iyi çözümdür. Genellikle bir sinyale hangi tipte pencere uygulanacağının bulunması için farklı tipte pencerelerle denemeler yapılması gerekmektedir (Hayes 1996, Semmlov 2004).

2.1.2. Fourier analizi ile güç spektrum yoğunluğunun bulunması

Periyodik sayısal bir işaret, Fourier serileri ile temsil edilebilir ve bu işaret kendisinin Fourier katsayılarından yeniden elde edilebilir. Fourier dönüşümü ile işaret, farklı frekanslardaki sinüzoidal sinyallerin toplamı olarak ifade edilir. Sonlu uzunlukta olan bir sayısal işaretin hızlı Fourier dönüşümünü almak için bu işaret 64,

128, 256, 512 gibi örnekler içeren pencerelere ayrılır. Her bir pencerenin frekans spektrumu hesaplanırken öncelikle pencereleme işlemi yapılır. Pencereleme işleminin amacı gerçekte olmayan frekans bileşenin spektrumda ortaya çıkmasını önlemektir. Pencereleme işleminden sonra sayısal işarete sıfır ekleme işlemi yapılır. Sıfır ekleme işlemi, spektrumda okunabilirliği arttırmaktır (Hayes 1996, Dirgenali 2003, Semmlov 2004).

Elde edilen ayrık zamanlı periyodik işaretin ayrık Fourier dönüşümü için (2.1) eşitliği kullanılmıştır.

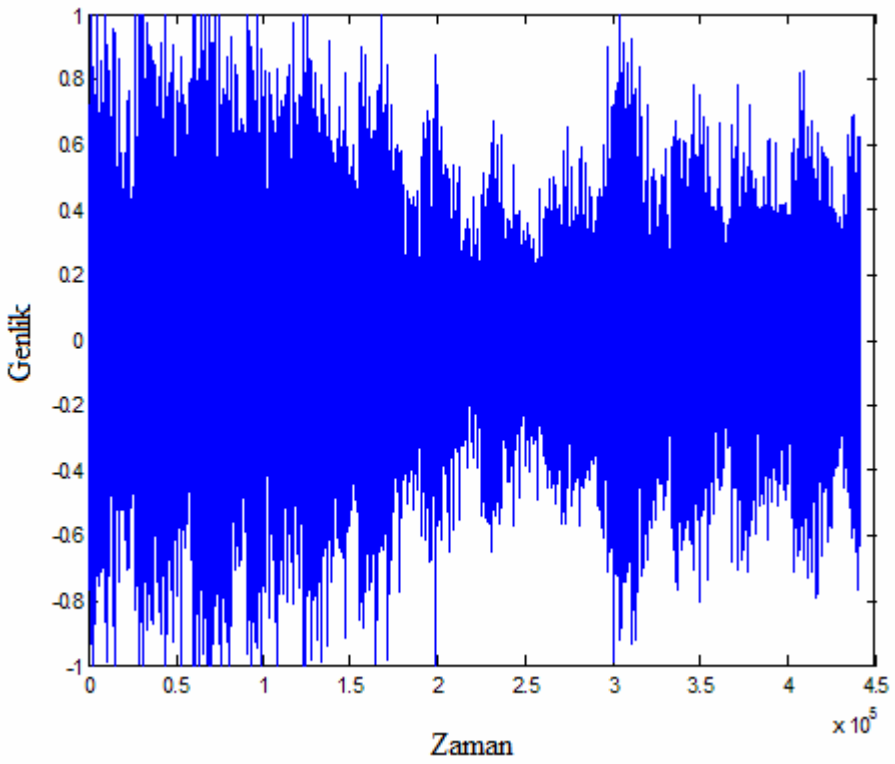
$$X_k = \sum_{n=0}^{N-1} x[n] \exp(-jkn \frac{2\pi}{N}) \qquad (2.1)$$

Buradaki X_k katsayıları ayrık Fourier katsayıları olarak adlandırılır. Burada, N pencere boyu, $x[n]$ zaman domeni giriş işaretidir. Bu işaretin güç spektral yoğunluğunu bulmak için ayrık Fourier dönüşüm katsayıları olan X_k katsayılarının mutlak değerlerinin karelerinin logaritması alınmıştır. Eşitlik (2.2), güç spektral yoğunluğunu göstermektedir (Hayes 1996, Dirgenali, 2003).

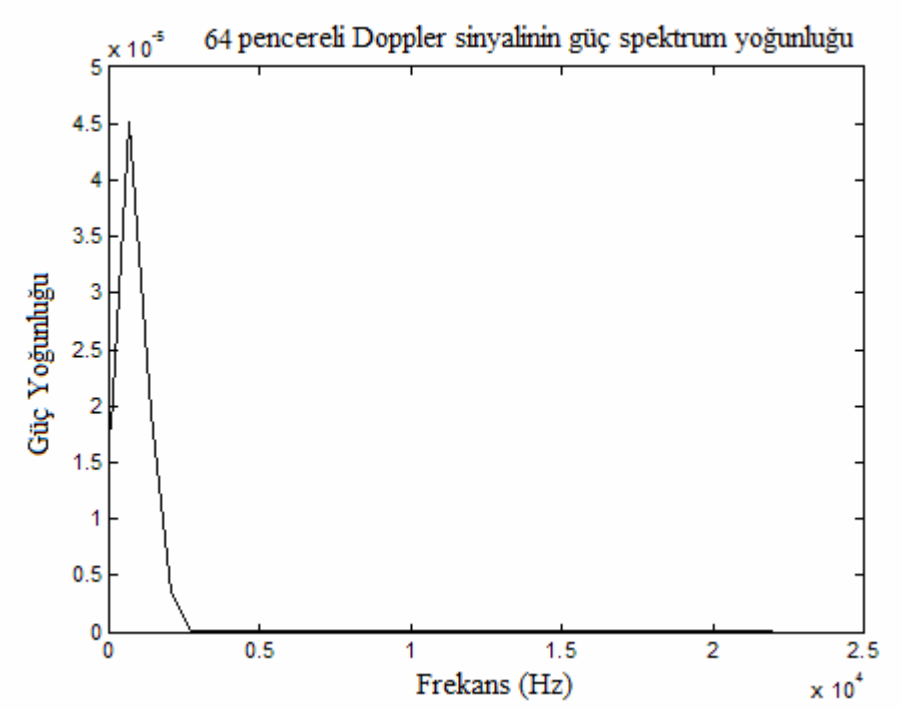
$$P(k) = 10 \log |X_k|^2 \qquad (2.2)$$

Fourier analizinin önemi, sinyalin frekans bileşenin oldukça önemli olduğu durumlarda ortaya çıkmaktadır. Tek bileşenli frekansdaki bir işarete bir çok fiziksel sistemin cevabı, diğer frekans bileşenlerinin genliğinden ve görüntüsünden bağımsız olmasıdır. Böyle sistemler giriş işaretenin büyüklüğündeki bir değişimin çıkış işaretinde de aynı oranda da değişim verdiği için doğrusal sistemlerdir. Ancak Fourier analizi frekans domenine geçerken zaman bilgisinin kaybolması gibi bir sakıncası vardır. Sinyal karakteristikleri zamanla sıkça değişmiyorsa bu sakınca çok önemli değildir. Ancak bazı sinyaller oldukça değişken (non-stationary) olabilir. Bu durumda Fourier analizinin performansı çok uygun olmayabilir. Doppler sinyallerini analiz ederken ve özellik çıkarma işlemini gerçekleştirirken Welch yöntemi kullanılmıştır (Hayes 1996, Dirgenali, 2003).

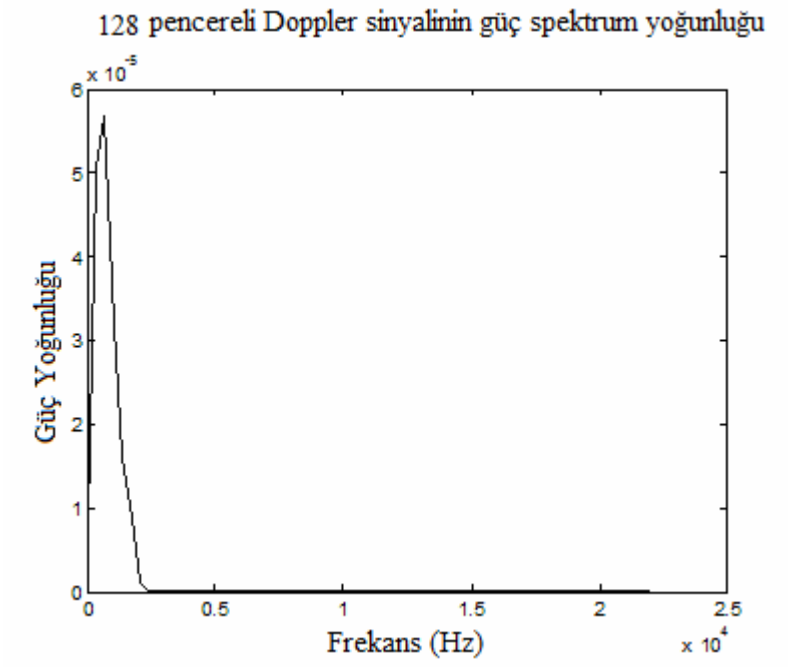
Aşağıda Welch yöntemine bir örnek verilmiştir. Bu örnekte 44100 Hz' de örneklenmiş bir kayıt olan Karotid atar damar Doppler sinyalinin, 64, 128 ve 256 pencere uzunluğuna sahip Welch yöntemi ile spektrum yoğunluğukları hesaplansın. Şekil 2.2, 44100 Hz' de örneklenmiş bir Doppler sinyali örneğini göstermektedir. Şekil 2.3, 64 pencereye ayrılmış Doppler sinyalinin Welch yöntemi sonucu güç spektrum yoğunluğunu göstermektedir. Şekil 2.4, 128 pencereye ayrılmış Doppler sinyalinin Welch yöntemi sonucu güç spektrum yoğunluğunu. Şekil 2.5, 256 pencereye ayrılmış Doppler sinyalinin Welch yöntemi sonucu güç spektrum yoğunluğunu göstermektedir.



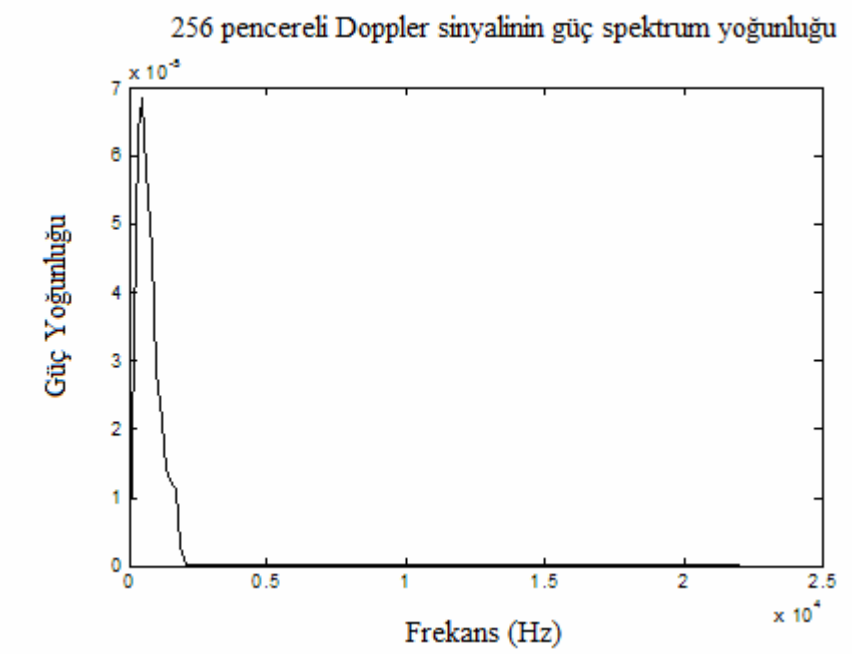
Şekil 2.2 44100 Hz' de örneklenmiş bir Doppler sinyali örneği



Şekil 2.3 64 pencereye ayrılmış Doppler sinyalinin Welch yöntemi sonucu güç spektrum yoğunluğu



Şekil 2.4 128 pencereye ayrılmış Doppler sinyalinin Welch yöntemi sonucu güç spektrum yoğunluğu



Şekil 2.5 256 pencereye ayrılmış Doppler sinyalinin Welch yöntemi sonucu güç spektrum yoğunluğu

Şekil 2.3, Şekil 2.4 ve Şekil 2.5 şekilleri incelendiğinde, her bir güç spektrumundaki frekans bileşenin aynı olduğu gözüküyor ancak bu frekansdaki güç yoğunluğunu farklı olduğu görünmektedir. Bu fark da percere uzunluklarının farklı olmasında yani çözünürlüğün az ya da fazla olmasından kaynaklanmaktadır.

2.2. Özellik Seçme ve Boyut Azaltımı Yöntemleri

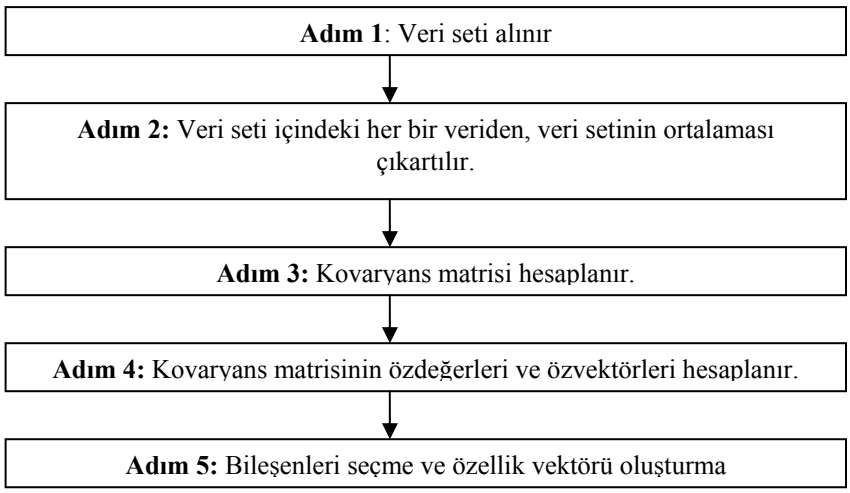
Özellik seçme ve boyut azaltımı yöntemleri; veri kümesinde bulunan ilgisiz veya gereksiz özellikleri uzaklaştırmak ve veri dağılımını daha uygun ve ayrıştırılabilir bir hale getirmek için uygulanır. Bu sayede, sınıflandırıcının hem hesaplama maliyeti azalmaktadır hem de sınıflama performansı artmaktadır. Boyut azaltımı olarak *temel bileşen analizi* kullanıldı. Özellik seçme yöntemi olarak da *Kernel F-skor özellik seçme* ve *bilgi kazancına dayanan özellik seçme* yöntemleri kullanıldı.

2.2.1. Temel bileşen analizi –TBA (Principal component analysis-PCA)

Temel bileşen analizi, bilgi kayıpsız bir veri kümesinde özelliklerin sayısını azaltmak için bir teknik olarak ve daha anlamlı yeni değişkenlerin tanımlanması için uygun bir işlem olarak sıklıkla tercih edilirler. *TBA*, veri azaltımının ilk derecede önemli olduğu görüntü sıkıştırma gibi uygulamalarda çok başarılı oldu. Birçok uygulamada, bir veri kümesinin uygun boyutunda sadece bilgi sağlamak için kullanılır. Yani, eğer bir veri kümesi M özellik içerirse, bilgiyi göstermek için bütün M özelliğe gerçekten ihtiyaç var mı veya önemli bilginin hala büyük bir kısmını içeren daha küçük bir sayıda değişken ile birleştirilebilir mi? Eğer böyleyse, yeni veri kümesinin en uygun boyutu ne olur ? Bu sorulara *TBA* ile cevap aranabilir (Semmlow 2004).

TBA, bağlantılı değişkenler kümesinin, ilişkisiz özellikler olarak adlandırılan *temel bileşenler* ile yeni bir veri kümesine dönüştürülmesiyle çalışır. Eğer bir veri kümesindeki özellikler birbirleriyle ilişkisiz ise (korelasyon=0), TBA hiçbir değere sahip olmaz. İlişkisiz olmanın yanı sıra, temel bileşenler birbirine diktir ve onların değişkenliğine (varyans değerlerine) göre sıralanırlar. Değişimi büyük olandan küçük olana doğru *temel bileşenler* sıralanırlar. Yani, ilk temel bileşen, tek bir boyut için orijinal veri kümesindeki değişkenliğin en büyük miktarını gösterir (Semmlow 2004, Lindsay 2002).

Şekil 2.6' da temel bileşen analizinin blok diyagramı verilmiştir.



Şekil 2.6 Temel bileşen analizinin blok diyagramı

Bu blok diyagram beş adımdan oluşmaktadır. Birinci adımda, veri seti alınır. İkinci adımda, veri seti içindeki her bir veriden, veri setinin ortalaması çıkartılır ve böylece sıfır ortalamalı veri seti elde edilmiş olur. Üçüncü adımda, kovaryans matrisi hesaplanır. Kovaryans, ölçümle belirtilen iki değişken boyutu arasındaki doğrusal ilişkiyi gösterir. Eğer tek boyut arasındaki kovaryans hesaplanırsa bu değer varyansı ifade eder. Elde edilen kovaryans matrisi, p boyutlu bir veri seti için, pxp boyutlu olur. Eğer değişkenler arasında belirgin bir ilişki yoksa kovaryans sıfıra yakın bir değer olur (Smith 2002). Eşitlik 2.3, x ve y değişkenleri arasındaki kovaryansı vermektedir. Dördüncü adımda, kovaryans matrisinin özdeğerleri ve özvektörleri hesaplanır. Özvektörleri olan nxn matrisin n tane öz vektörü vardır.

$$\mathrm{cov}(x,y)=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1} \tag{2.3}$$

Verilen n mertebeli bir A matrisi için λ gibi bir sabit ve X gibi sıfır olmayan bir vektör belirlenir ki bu değerler Eşitlik 2.4 ile verilen denklem sistemini sağlarlar. Yapılan bu işleme özdeğer bulma problemi denir (Smith 2002).

$$AX=\lambda X \tag{2.4}$$

Burada A, n. dereceden bir kare matristir. Matris işlemleri uygulandığı zaman, denklem sistemi, Eşitlik 2.5 şekline dönüşmüş olmaktadır. Eşitlik 2.5 homojen bir denklem sistemidir.

$$(A-\lambda I)X=0 \tag{2.5}$$

X'in sıfırdan farklı olabilmesi için $(A-\lambda I)$ matrisin determinantının 0 olması gerekir. Yani $|A-\lambda I|=0$ olur. Buna göre λ cinsinden polinominal bir karakteristik denklem bulunur. Yani bu denklemin derecesi A matrisinin derecesine eşittir. İkinci dereceden bir A matrisi için iki tane kök (λ_1 ve λ_2) vardır. Bu kökler gerçek yada

karmaşık olabilir. Eğer matris simetrikse kökler gerçek sayılardır. A, pxp boyutlu bir matris ise, bu matrisin karakteristik denkleminin p tane kökü vardır (Smith 2002).

Özvektörler bulunurken, $(A-\lambda I)X=0$'da her bir özdeğer (λ_i) için bir tane özvektör (v_i) vardır. Özvektörleri bulmak için aşağıdaki adımlar izlenebilir. Herbir λ_i için $(A-\lambda_i I)$ matrisi bulunur. Daha sonra, $(A-\lambda_i I)$'nin adjointi (A'nın transpozu alınmış matris) bulunur. $Adj(A-\lambda_i I)$ matrisinin sütunlarının tümünün birbiriyle orantılı olması gerekir (Smith 2002, Semmlov 2004). Her sütunun elemanları, ilgili sütundaki elemanların kareleri toplamının karekköküne bölünür. Elde edilen sayılar v_i'nin elemanları olup $v_i.v=1$'dir.

Beşinci adımda ise, bileşenleri seçme ve özellik vektörü oluşturma işlemi yapılır. Bu adım verinin sıkıştırıldığı ve boyutun azaltıldığı adımdır. Daha önceki bölümde de belirtildiği gibi, özdeğerler birbirlerinden farklı değerler alırlar. En yüksek değerli özdeğer, veri setinin temel bileşenidir. Bu durum bileşenlerdeki sıralamanın önemini gösterir. Bundan sonra daha az önemli bileşenin atılma işlemi gerçekleştirilir. Bu durumda bazı bilgiler kaybolacaktır. Eğer özdeğerler küçükse bilgi kaybı da az olacaktır. Bazı bileşenler atıldıktan sonra elde edilecek olan veri matrisinin boyutu, orijinal veri matris boyutundan daha küçük olur. Bu durum daha açık olarak ifade edilirse, n boyutlu bir veri seti için, n adet özvektör ve özdeğer hesaplanır. Bu özvektörler arasından p tanesi seçilirse, son veri seti sadece p boyutta olur. Bundan sonra ise özellik vektörünün oluşturulması işlemi vardır. Bu özellik vektörü, özvektörler listesinden seçilen özvektörlerin sütun formundan matris formuna getirilmesiyle elde edilir (Semmlov 2004, Smith 2002).

Temel bileşen analizini daha iyi anlayabilmek için aşağıda x ve y' den oluşan iki boyutlu bir veri kümesi verilmiş olsun. Bu veri kümesinin *temel bileşenlerini* bulalım. Tablo 2.1 x ve y değişkeninden oluşan bir veri kümesini göstermektedir.

Tablo 2.1 x ve y değişkeninden oluşan bir veri kümesi

x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

Birinci adımda veri bir sütunun ortlaması alınır ve veri kümesinden her bir değerden çıkarılır ve bu sayede sıfır ortalamalı bir veri kümesi elde edilir. Tablo 2.2, veri kümesinin ortalama değerden çıkarılmış halini göstermektedir.

Tablo 2.2 Ortalama değerden çıkarılmış veri kümesini yeni hali

x	y
0.69	0.49
-1.31	-1.21
0.39	0.99
0.09	0.29
1.29	1.09
0.49	0.79
0.19	-0.31
-0.81	-0.81
-0.31	-0.31
-0.71	-1.01

Sonra kovaryans matrisi hesaplanır ve aşağıdaki matris elde edilir:

$$\text{cov} = \begin{bmatrix} 0.616555 & 0.6154444 \\ 0.615444 & 0.7165555 \end{bmatrix}$$

Daha sonra kovaryans matrisinin özdeğer ve özvektörleri hesaplanır. Hesaplanan özvektörler, varyansı büyük olandan küçük olana doğru sıralanırlar ve en yüksek varyans değerine sahip olan özvektör x ve y' nin yerine kullanılacak yeni boyut olur.

$$oz\deg er = \begin{bmatrix} 0.049083 \\ 1.284027 \end{bmatrix}$$

$$oz\,vektorler = \begin{bmatrix} -0.735178 & 0.677873 \\ 0.677873 & 0.735178 \end{bmatrix}$$

1. ve 2. özvektörlerin varyansları (değişimleri); 0.998 ve 0.0016 dır. Bu değerlere 1. özvektör x ve y değişkenlerini göstermek için kullanılabilir.

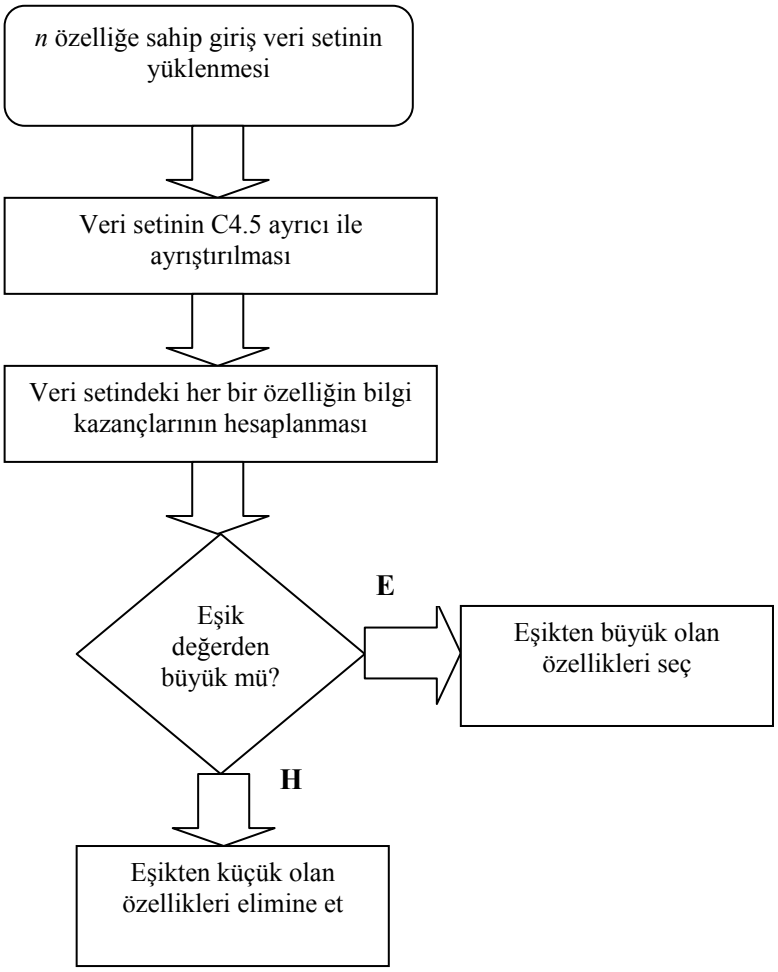
2.2.2. Bilgi kazancına dayanan özellik seçme algoritması

Bilgi kazancına dayanan özellik seçme algoritması, yüksek boyutlu bir veri kümesinden ilgisiz, fazla ve bilgi kazancı düşük olan özellikleri elimine eder. Bu işlem sayesinde, en iyi ve daha çok bilgiye sahip özellikler elde edilir (Liu 1998). Bu özellik seçme işlemi iki aşamadan oluşmaktadır. İlk aşamada, veri ayrışımı yapılmaktadır yani sürekli veriler C4.5 ayrıcı ile ayrık değerli hale dönüştürülmektedir. Bilgi kazancının hesaplanmasında sürekli değerler kullanılmadığından bu işlem gereklidir. Sonra ikinci aşamada ise, ayrışan bu sürekli verilerin her bir özelliğinin bilgi kazancı hesaplanmaktadır. Özelliklerden elde edilen bilgi kazançlarına göre eşik değerden küçük olan özellikler elimine edilmektedir. Bilgi kazançlı özellik seçme algoritmasının blok şeması Şekil 2.7' de gösterilmiştir. Sürekli değerleri verileri ayrık değerleri verilere dönüştürürken C4.5 ayırıcı kullanıldı. C4.5 ayırıcı, karar ağacı metodunda özelliklerin bilgi kazancı hesaplanırken özellik değerlerinin ayrık değerlere dönüştürülmesinde kullanılmaktadır.

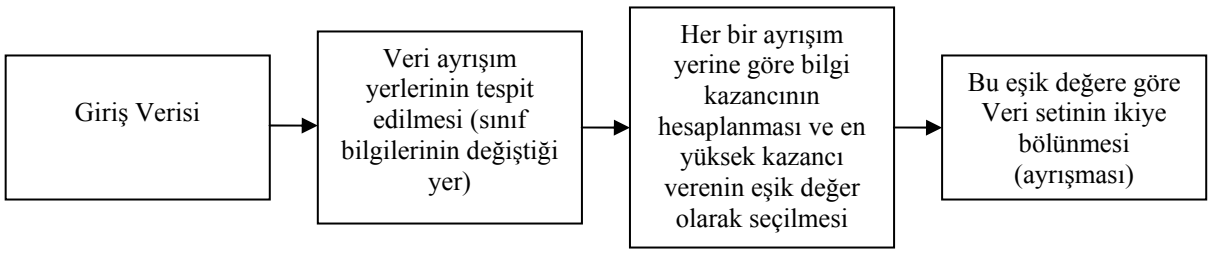
C4.5 ayırıcı sürekli özellikleri bir eşik değeri kullanarak her bir özelliği ikiye ayırarak karar ağacı oluşturmaktadır (Pazzani 1995). Bilgi kazancı teorisi, veriyi bölmek için en büyük kazanç oranını veren eşik değerini belirlemede kullanılmaktadır. Böl ve Fethet algoritması her iterasyonda aralığın daha küçük alt kümelere bölünüp bölünmeyeceğine karar vermek için uygulanmaktadır. C4.5 ayırıcının blok şeması Şekil 2.8' de verilmiştir. Bu durum bir örnekle anlatılabilir:

Bu örnekte, tenis oynayıp oynamama, sürekli değerli sıcaklık değerine göre değişmektedir. Tenis oynama durumu var ise evet, yok ise hayır olarak etiketlenmiştir. Sıcaklık değişkenini ayrık değerli hale getirmek için ilk olarak sınıf

etiketlerinin değiştiği yerler işaretlenir. Sonra kestiği yerlerdeki sıcaklık değerlerini ortalamaları hesaplanır. Her bir alternatif ayırım için bilgi kazançları hesaplanır ve en yüksek bilgi kazancını veren yer ayırım noktası olarak seçilir.



Şekil 2.7 Bilgi kazancı tabanlı özellik seçme algoritması



Şekil 2.8 C4.5 veri ayrışımı (Pazzani 1995)

2.2.2.1. Bilgi kazancı hesabı ve özellik seçme işlemi

Bilgi kazancı, veri kümesini oluşturan özellik ile çıkış etiketi arasındaki karşılıklı bilgiyi hesaplamak için kullanılan bir terimdir. Bilgi kazancı hesaplanmasında Entropi (Entropy) ölçüsünü kullanmaktadır (Quinlan 1986, Mitchell 1997). *Entropi*, örneklerin belirsizliğini ölçmek için kullanılır. Eşitlik 2.6'de *S* kümesinin *Entropi* değeri gösterilmiştir.

$$Entropi(S) = \sum_{i=1}^{c} - p_i log_2 p_i \qquad (2.6)$$

burada p_i, *S* kümesindeki *i.* sınıfa ait örneklerin sayısının *S* kümesindeki bütün örneklerin sayısına oranıdır. *c*, sınıf sayısıdır. *S* kümesindeki bütün örnekler aynı sınıfta ise *Entropi* değeri sıfırdır yani belirsizlik yoktur. *S* kümesindeki örneklerin sınıf sayısı eşit ise *Entropi* değeri maksimum değeri olan bir değerini almaktadır yani belirsizlik maksimumdur.

Bir özelliğin etkinlik ölçüsü bilgi kazancı terimi ile kullanılmaktadır. Eşitlik 2.7'de *A* özelliğinin bilgi kazancı *Kazanç(S,A)* tanımlanmaktadır.

$$Kazanç(S, A) = Entropi(S) - \sum_{v \in Değerler(A)} \frac{|S_v|}{|S|} Entropi(S_v) \qquad (2.7)$$

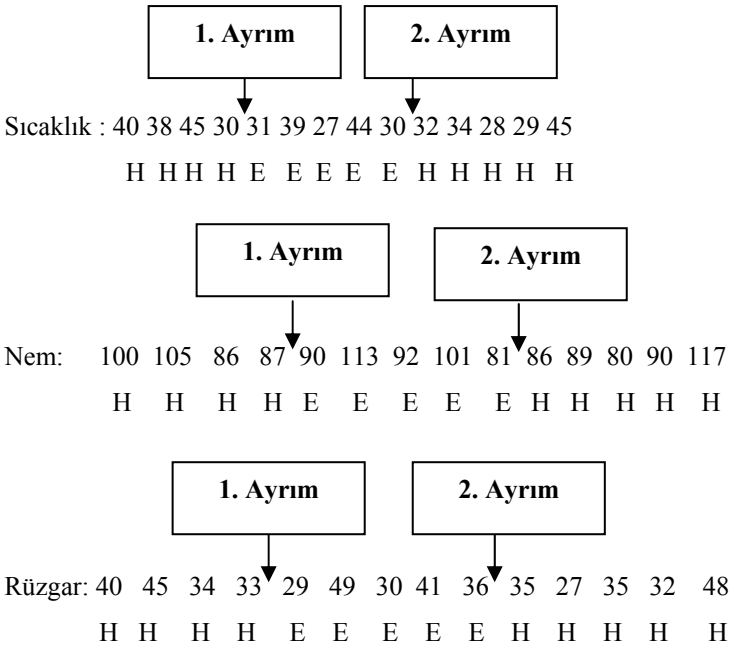
Burada *Değerler(A), A* özelliğinin bütün olası değerlerinin kümesidir; S_v ise *S* kümesindeki *A* özelliğinin *v* değerine sahip örneklerinin kümesidir.

Her bir özelliğin bilgi kazancı hesaplandıktan sonra kullanıcı tarafından girilen eşik değere göre özellikler seçilir. Bilgi kazancına dayanan özellik seçme yöntemini daha iyi anlayabilmek için aşağıda sürekli değerli bir tenis oynama veri kümesi verilmiştir. Tablo 2.3, tenis oynama durumunu gösteren üç özellikten oluşan veri kümesini göstermektedir.

Tablo 2.3 Sürekli değerli üç özelliğe sahip tenis oynama durumu için veri kümesi

Gün	Sıcaklık	Nem	Rüzgar	Tenis Oynar
1	*40*	*100*	*40*	***Hayır***
2	*38*	*105*	*45*	***Hayır***
3	*45*	*86*	*34*	***Hayır***
4	*30*	*87*	*33*	***Hayır***
5	*31*	*90*	*29*	***Evet***
6	*39*	*113*	*49*	***Evet***
7	*27*	*92*	*30*	***Evet***
8	*44*	*101*	*41*	***Evet***
9	*30*	*81*	*36*	***Evet***
10	*32*	*86*	*35*	***Hayır***
11	*34*	*89*	*27*	***Hayır***
12	*28*	*80*	*35*	***Hayır***
13	*29*	*90*	*32*	***Hayır***
14	*45*	*117*	*48*	***Hayır***

İlk olarak, her bir sürekli değerli özellik C4.5 ayırıcı ile ayrık değerlere dönüştürülür.



Her bir özellikteki 2 ayrım noktası için bilgi kazançları hesaplanır ve en yüksek bilgi kazancına sahip olan ayrım noktası ayırıcı nokta olarak seçilir ve daha sonra ayrık hale getirilen özelliklerin ayrı ayrı bilgi kazançları hesaplanır ve ortalama bilgi kazancı hesaplanır. Ortalama bilgi kazancını geçen özellikler seçilir diğerleri elimine edilir.

Sıcaklık özelliği için 1. ayrım noktası bilgi kazancı hesabı:

İlk olarak genel entropi hesaplanır. 14 sonuçtan 9 hayır, 5 evet vardır.

$Entropi(S) = -(9/14)log_2(9/14) - (5/14)log_2(5/14) = 0.939$

Eşik değer: (30+31)/2=30.5



H H H H E E E E E E E E E E

Yukarıdaki durum için entropi hesaplanır. 14 sonuçtan 4 hayır, 10 evet vardır.

$Entropi(S1) = -(4/14)log_2(4/14) - (10/14)log_2(10/14) = 0.862$

1.ayrım noktasının bilgi kazancı: $Entropi(S) - Entropi(S1) = 0.939 - 0.862 = 0.077$

Sıcaklık özelliği için 2. ayrım noktası bilgi kazancı hesabı:

İlk olarak genel entropi hesaplanır. 14 sonuçtan 9 hayır, 5 evet vardır.

$Entropi(S) = -(9/14)log_2(9/14) - (5/14)log_2(5/14) = 0.939$

Eşik değer: (30+32)/2=31



H H H H H H H H H E E E E E

Yukarıdaki durum için entropi hesaplanır. 14 sonuçtan 9 hayır, 5 evet vardır.

$Entropi(S1) = -(9/14)log_2(9/14) - (5/14)log_2(5/14) = 0.939$

2.ayrım noktasının bilgi kazancı: $Entropi(S) - Entropi(S1) = 0.939 - 0.939 = 0.0$

Bu sonuçlara göre sıcaklık özelliğinin ayrım noktası 1. noktadır. Bu ayrım noktasına göre ayrık hale dönüştürülmüş sıcaklık özelliği:

Sıcaklık:

Değer	40	38	45	30	31	39	27	44	30	32	34	28	29	45
Ayrık Değeri	A1	A1	A1	A1	A2	A2	A2	A2	A2	A2	A2	A2	A2	A2
Sınıfı	H	H	H	H	E	E	E	E	E	H	H	H	H	H

Sonra, sıcaklık özelliğinin bilgi kazancı hesaplanır.

Değerler(Sıcaklık)= {A1, A2}

S= [5+, 9-]

S_{A1}*= [0+, 4-]*

S_{A2}*= [5+, 5-]*

$$Kazanç(S, Sıcaklık) = Entropi(S) - \sum_{v \in \{A1,\ A2\}} \frac{|S_v|}{|S|} Entropi(S_v)$$

= 0.939-(4/14)Entropi(S_{A1}*)-(10/14)Entropi(*S_{A2}*)*

= 0.939 – (4/14)0.0 – (10/14)1.0

= 0.511 olarak hesaplanmaktadır.

Sıcaklık özelliğinden sonra diğer geriye kalan özellikler için de aynı işlem yapılır. Burada her bir özelliğin bilgi kazançları aynı olduğu için herhangi bir özellik seçilemedi veya elimine edilmedi.

2.2.3. Kernel F-Skor özellik seçme yöntemi

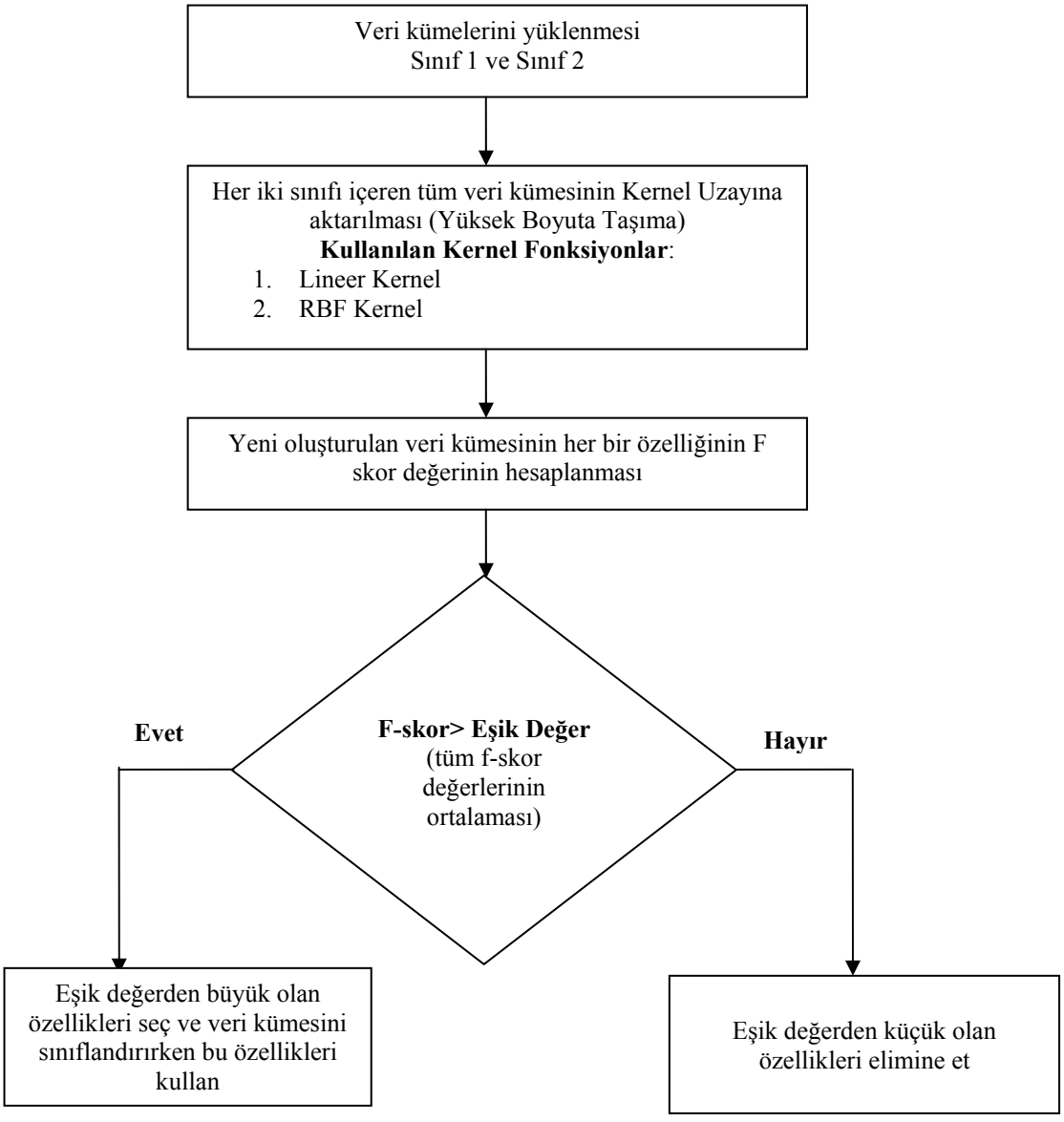
F-skor yöntemi, gerçek değerli iki sınıfın ayırt edilmesini ölçen basit bir tekniktir. F-skor yönteminde, veri kümesindeki her bir özelliğin eşitlik 2.8' e göre f-skor değerleri hesaplanır ve hesaplanan f-skor değerlerinin ortalaması alınarak özellikleri seçmek için eşik değer seçilir. Eşik değerden büyük olan özellikler seçilir, diğerleri ise veri kümesinden uzaklaştırılır. $x_k =, k = 1,....,m$ eğitme vektörleri

verilsin, n_+ ve n_- sırasıyla pozitif ve negatif örneklerin sayısı olsun. *i.* özelliğin f-skor değeri eşitlik 2.8' de hesaplanır: (Chen 2003):

$$F(i) = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}(x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}(x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2} \tag{2.8}$$

Burada, $\bar{x}_i, \bar{x}_i^{(+)}, \bar{x}_i^{(-)}$, sırasıyla tüm veri kümesi, pozitif ve negatif veri kümelerinin *i.* özelliğinin ortalamalarıdır. $x_{k,i}^{(+)}$, *k.* pozitif örneğinin *i.* özelliğidir ve $x_{k,i}^{(-)}$, *k.* negatif örneğinin *i.* özelliğidir. Pay, pozitif ve negatif kümeler arasındaki ayırımı gösterirken, payda ise pozitif ve negatif kümelerin varyanslarını gösterir. F- skor değeri ne kadar büyükse, o özelliğin ayırt edici özelliği de büyüktür. Fakat F-skor' nun bir dezavantajı, değişkenler arasındaki karşılıklı bilgiyi hesaba katmaz (Chen 2003). Bu önerilen özellik seçme yönteminde, non-lineer veri dağılımına sahip olan veri kümelerinin daha kolay sınıflandırılabilir hale getirilmesi ve hesaplama maliyetinin azaltılması için Kernel F-skor yöntemi önerildi. Bu yöntemde, ilk olarak veri kümesi kernel (lineer veya RBF fonksiyonu) fonksiyonu ile lineer olarak ayrılabilir daha yüksek boyutlu bir veri kümesine taşınır. Daha sonra, bu yüksek boyutlu veri kümesinin özelliklerinin F-skor değerleri hesaplanır ve bu hesaplanan değerlerden yüksek olanları (eşik değerden büyük olanlar) seçilir ve diğerleri elimine edilir. Önerilen özellik seçme yöntemi sayesinde, yüksek boyutlu özellik sayısından, ilgisiz veya fazla özellikler uzaklaştırılır. Kernel fonksiyon kullanımının sebebi, lineer olarak ayrılamayan medikal veri kümelerini lineer olarak ayrılabilir özellik uzayına dönüştürmektir. Şekil 2.9, Kernel F-skor özellik seçmenin akış şemasını göstermektedir. Kullanılan kernel fonksiyonları:

- Lineer (Dot product) kernel : K(x,x') = x.x' ;
- RBF kernel: K(x,x') = $\exp(-\|x - x'\|^2 / \sigma^2)$; burada σ pozitif bir gerçek sayısıdır.



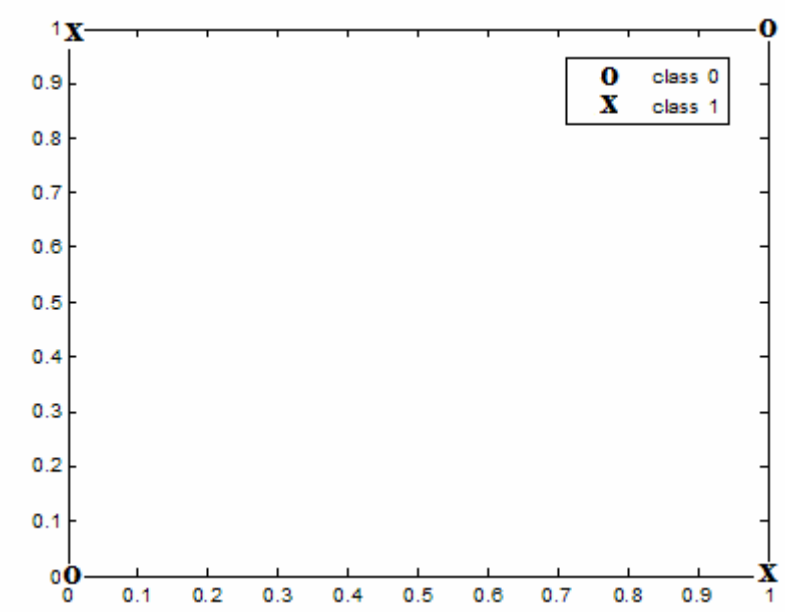
Şekil 2.9 Kernel F-skor özellik seçme yönteminin akış şeması

Kernel F-skor özellik seçme yöntemi, non-lineer bir veri dağılıma sahip XOR problemine uygulandı. Tablo 2.4 XOR probleminin veri dağılımını göstermektedir. Şekil 2.10, XOR problemi veri dağılımını grafik olarak göstermektedir ve bu şekilden anlaşılacağı gibi non-lineer bir dağılıma sahip olduğu görülmektedir. XOR

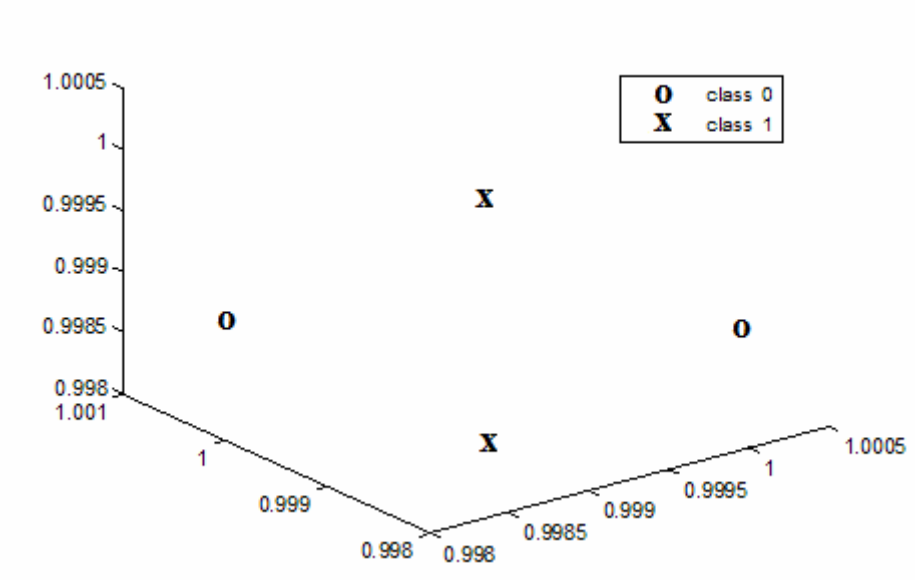
problemi RBF ve lineer Kernel uzayı atıldıktan sonra özellik sayısı veri sayısına arttı yani 4' e yükseldi. Ve daha sonra her bir özelliğin f-skor değerleri hesaplandıktan özellik sayısı 3'e indi. Şekil 2.11, XOR veri kümesinin RBF Kernel uzayına atıldıktan sonraki veri dağılımını göstermektedir. Şekil 2.12, XOR veri kümesinin lineer Kernel uzayına atıldıktan sonraki veri dağılımını göstermektedir. Şekil 2.11 ve Şekil 2.12' ü incelediğimizde XOR problemi, lineer Kernel uzayına atıldıktan sonra lineer bir biçimde ayrılmaktadır.

Tablo 2.4 XOR problemi veri dağılımı

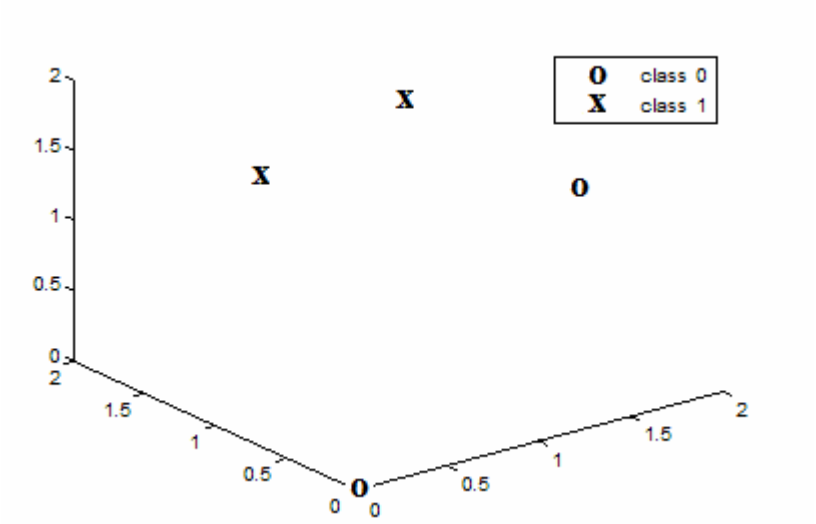
Girişler		**Çıkış**
X1	X2	Y
0	0	0
0	1	1
1	0	1
1	1	0



Şekil 2.10 Orijinal XOR veri dağılımı



Şekil 2.11 RBF Kernel uzayına atıldıktan sonraki XOR veri dağılımı



Şekil 2.12 Lineer Kernel uzayına atıldıktan sonraki XOR veri dağılımı

3. VERİ AĞIRLIKLANDIRMA VE ÖN-İŞLEME YÖNTEMLERİ

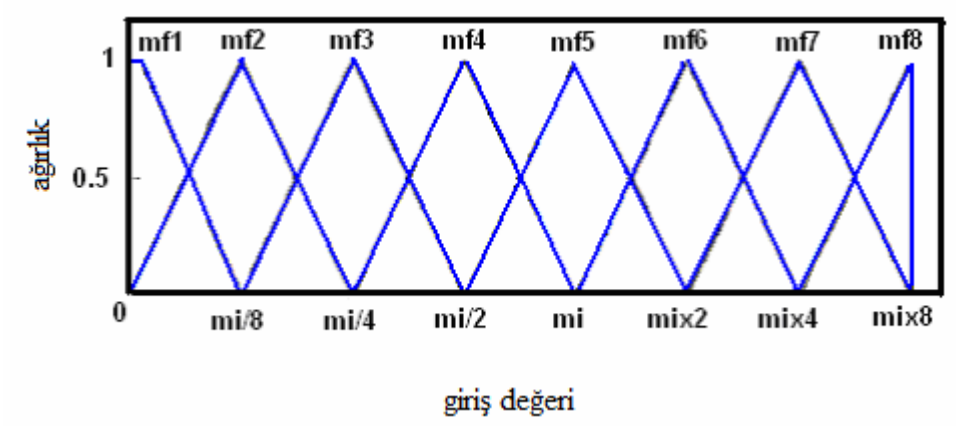
Medikal verilerin sınıflandırılmasında sınıflama algoritmaları, verilerin aykırı değerlere sahip olması, verilerde eksik veri bulunması, aynı veriden iki defa bulunması, verilerin lineer olarak ayrılamayan yapıya sahip olmaları gibi nedenlerden dolayı iyi bir performans elde edemeyebilirler. Bu sorunları ortadan kaldırmak için çeşitli veri ön-işleme algoritmaları geliştirilmiştir. Veri ön-işleme teknikleri; veri temizleme, normalizasyon, veri dönüşümü, veri ağırlıklandırma, özellik çıkarımı, özellik seçme gibi yöntemleri içerir. Bu çalışmada, veri ağırlıklandırma ve veri dönüşümü yöntemleri üzerinde çalışılmıştır.

Sınıflama algoritmaları, makinece okunabilir (sensör) veriden, otomatik olarak bilgiyi çıkartmak için kullanılır. Maalesef, sınıflama algoritmalarını başarısı genellikle işlem yapılan verinin kalitesine bağımlıdır. Eğer veri uygun değilse, sınıflama algoritmaları daha az başarı ve daha az anlaşılır sonuçlar elde edebilir veya onları hiçbir şekilde kullanamaz. Dolayısıyla, veri ön-işleme sınıflama algoritmalarında çok önemli bir adımdır. Ön-işleme, gürültülü veri, fazla veri, eksik veri değerlerini içeren birkaç çeşit problemi çözmek için gereklidir.

Bu tez çalışmasında, altı adet veri ön-işleme yöntemi kullanıldı. Bu yöntemler içinden; *bulanık mantık tabanlı veri ağırlıklandırma işlemi*, *k-NN (k-en yakın) komşu tabanlı veri ağırlıklandırma işlemi* ve *benzerlik tabanlı veri ağırlıklandırma* yöntemleri bu tez çalışmasında geliştirilen algoritmalardır. Ayrıca tez çalışmasında kullanılan diğer bir veri ön-işleme algoritması ise *GAA* yöntemidir. Non-lineer veri dağılımlarına sahip verileri lineer hale getirerek sınıflama başarısını arttıran yeni bir özellik uzayı dönüşüm yöntemidir.

3.1. Bulanık Mantık Tabanlı Veri Ağırlıklandırma İşlemi

Bulanık ağırlıklandırma işleminde, her bir özellik kendi eski değerine göre yeni bir değer alır. Giriş ve çıkış üyelik fonksiyonları, bu ön işlemede tanımlanırlar. Şekil 3.1 ve Şekil 3.2' de gösterildiği gibi üçgen üyelik fonksiyonları olarak seçilirler. Üçgen üyelik fonksiyonu seçilmesinin sebebi; hesaplama işleminin basit olması ve değişiklik yapma olanağının diğer fonksiyonlara göre daha kolay olmasıdır.



Şekil 3.1 Giriş üyelik fonksiyonları



Şekil 3.2 Çıkış üyelik fonksiyonları

İlk olarak, bu üyelik fonksiyonlarının oluşturulması şöyle gösterilir: İlk adım olarak, her bir özelliğin örnek ortalamaları farklı olduğu için, her bir özelliğin ortalama değerleri, eşitlik 3.1' de gösterildiği gibi bütün örneklerin uygun özellik değerleriyle hesaplanır.

$$m_i = \frac{1}{N}\sum_{k=1}^{N} x_{k,i} \qquad (3.1)$$

Burada, $x_{k,i}$, x_k, $(k=1,2,...N)$ örneğinin *i.* özellik değerini gösterir. Her bir özellik için bu ortalama değerler hesaplandıktan sonra, giriş üyelik fonksiyonu Şekil 3.1' de gösterildiği gibi üçgenlerle biçimlendirilir. Bu üçgenlerin destekleri Şekil 3.1' de gösterildiği gibi $m_i/8$, $m_i/4$, $m_i/2$, m_i, $2 \times m_i$, $4 \times m_i$, $8 \times m_i$ ile gösterilir. Giriş üyelik fonksiyonlarının satırları *mf1, mf2,...mf8* olarak isimlendirilir. Çıkış üyelik fonksiyonunun biçimlendirilmesi için, tekrar 8 parça üyelik fonksiyonu oluşturulur (Şekil 3.2). [0,1] aralığı, 8 eşit parçaya bölünür ve uygun satırlar *mf1', mf2',...mf8'* olarak isimlendirilir.

Sonraki adım, ağırlıklandırma işlemini kullanan adımdır. Bir özellik değeri için, örneğin x_k örneğinin *i.* özellik değeri olan $x_{k,i}$ olsun, bu değer, giriş üyelik fonksiyonun *x* eksenindeki gibi alınır ve bu değerin giriş üyelik fonksiyonunu kestiği noktaların *y* değerleri belirlenir. Örnek olarak, eğer bu değer 0 ve $m_i/8$ arasında ise, sonra bu nokta *mf1* ve *mf2* eğrilerinin kesecektir. Bu kesişim noktalarındaki *y* değerleri, y_1 ve y_2, üyelik değerleri (μ) olarak bilinirler ve sonra bunlar bir bulanık kural tabanını oluşturmada kullanılacaktır. İlk olarak, $\mu(i)$ giriş üyelik değeri, yukarıdaki kesişim noktaları kullanılarak belirlenir:

Burada, $\mu_A(x_{k,i})$ and $\mu_B(x_{k,i})$ üyelik değerleri, yukarıda bahsedildiği gibi kesişim noktalarına uygundur. Sistemimiz için hazırlanan kural tabanı Tablo 3.1' de gösterilir. $\mu(i)$ değer, $x_{k,i}$ bizim özellik değeri içim eşitlik 3.2 kullanılarak belirlendikten sonra, çıkış ağırlık değeri, Tablo 3.1' de gösterildiği gibi çıkış üyelik fonksiyonları ve kurallar kullanılarak belirlenir.

$$\mu(i)=\mu_{A\cap B}(x_{k,i}) = \text{MIN}\,(\mu_A(x_{k,i}), \mu_B(x_{k,i})),\ x \in X \qquad (3.2)$$

Son adım olarak ağırlıkların belirlenmesinde, ilk olarak $\mu(i)$ giriş üyelik değeri, bizim orijinal özellik değerinin uygun ağırlıklandırılmış değerini belirlemek için çıkış üyelik fonksiyonuna sunulur. Bu üyelik değeri şimdi çıkış üyelik fonksiyonunun y ekseninde bir nokta olarak alınır. Bu üyelik değeri tarafından kesilen kesişim noktaları belirlenir. Çıkış üyelik fonksiyonundan görülüyor ki, birden

fazla kesişim noktası olacaktır. Kullanılacak kesişim noktaları, Tablo 3.1' deki kurallar vasıtasıyla karar verilir. Örnek olarak, giriş özellik değeri, giriş üyelik fonksiyonunda mf1 ve mf2 eğrilerini keserse, bu özellik için çıkış değeri, çıkış üyelik fonksiyonunda *mf1'* ve *mf2'* eğrilerini kesen iki noktanın ortalaması olacaktır. Ayrıca, bu oluşturulan kurallar her bir durum için kullanılır (Polat 2007).

Tablo 3.1. Kullanılan sistem için bulanık kural tabanı

Kurallar
if Input_value cuts mf(1) and mf(2) then Output_value=(mf(1)'(y)+mf(2)'(y))/2
if Input_value cuts mf(2) and mf(3) then Output_value= (mf(2)'(y)+mf(3)'(y))/2
if Input_value cuts mf(3) and mf(4) then Output_value= (mf(3)'(y)+mf(4)'(y))/2
if Input_value cuts mf(4) and mf(5) then Output_value= (mf(4)'(y)+mf(5)'(y))/2
if Input_value cuts mf(5) and mf(6) then Output_value= (mf(5)'(y)+mf(6)'(y))/2
if Input_value cuts mf(6) and mf(7) then Output_value= (mf(6)'(y)+mf(7)'(y))/2
if Input_value cuts mf(7) and mf(8) then Output_value= (mf(7)'(y)+mf(8)'(y))/2

Bulanık veri ağırlıklandırma ön-işlemeye bir örnek verelim. *x* ve *y* değişkeninden oluşan bir veri kümemiz olsun. Tablo 3.2, *x* ve *y* değişkeninden oluşan veri kümesini göstermektedir. İlk olarak *x* ve *y* değişkenleri [0,1] aralığında normalize edilir. Daha sonra her bir sütunun ortlama değeri hesaplanır ve giriş ve çıkış üyelik fonksiyonlarının kesim yerleri belli olur. Tablo 3.3, *x* ve *y* değişkenin [0,1] aralığındaki normalize edilmiş değerlerini göstermektedir.

Tablo 3.2 *x* ve *y* değişkeninden oluşan veri kümesi

x	***y***
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

Tablo 3.3 *x* ve *y* değişkenin [0,1] aralığındaki normalize değerleri

Veri sayısı	***x***	***y***
1	0,76	0,73
2	0	0
3	0,65	0,95
4	0,53	0,65
5	1	1
6	0,69	0,86
7	0,57	0,39
8	0,19	0,17
9	0,38	0,39
10	0,23	0,08
Ortlama değeri *(m_i)*	**0.503**	**0.526**

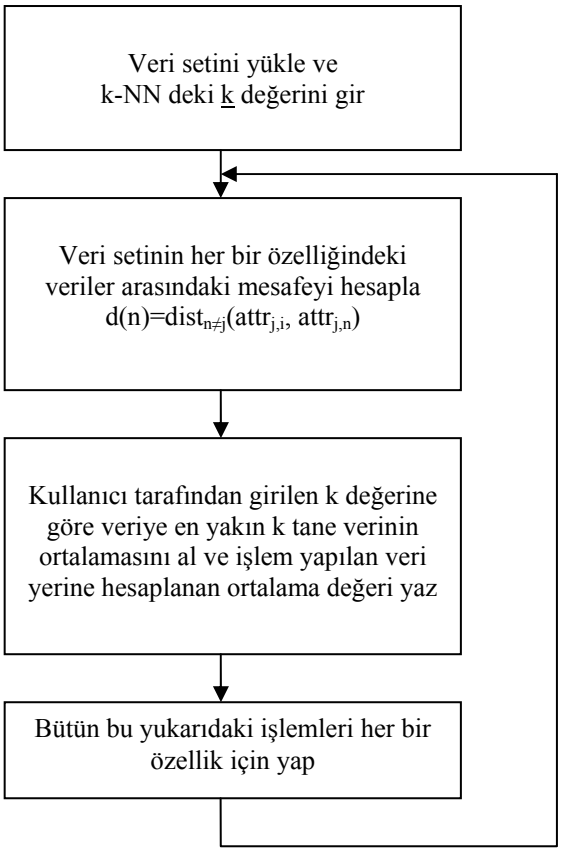
Bulanık veri ağırlıklandırma işlemi her bir sütun için ayrı ayrı yapılmaktadır. Burada, x sütunun 1. verisini yukarıdaki prosedüre göre ağırlıklandıralım. *x* sütunun 1. verisi olan 0.76 verisini giriş üyelik fonksiyonunda hangi aralığı girdiği belirlenir. m_i ve $m_i/2$ (m_i=0.503) aralığına düşmektedir. Kestiği üyelik fonksiyonlar şekil 3.1' den görüleceği gibi *mf5* ve *mf6* eğrilerini kesmektedir. *mf5*' den kestiği değer 0.489, *mf6*' dan kestiği değer ise 0.510 dur. Tablo 3.1' e baktığımızda, eğer *mf5* ve *mf6*' yı keserse bu iki değerden minimum değere sahip olanı al ve çıkışta *mf5'* ve *mf6'* üyelik fonksiyonları keserek bu iki fonksiyondan elde eldilen değerlerin ortalaması alınarak yeni ağırlıklı değer bulunmuş olur. Giriş üyelik fonksiyonunda elde edilen minimum değer 0.489 dur. Bu değerin *mf5'* üstünde kestiği değer 0.631, *mf6'* de kestiği değer de 0.628 ve bunların ortlaması da 0.63 çıkmaktadır. Yani 0.76 değerinin bulanık veri ağırlıklandırma sonrası yeni değeri 0.63 dür. Tablo 3.4, *x* ve *y* değişkenin bulanık veri ağırlıklandırma sonrası yeni halini göstermektedir.

Tablo 3.4 *x* ve *y* değişkenin bulanık ağırlıklandırma sonrası yeni değerleri

Veri sayısı	***x***	***y***
1	0,63	0,63
2	0,21	0,21
3	0,63	0,63
4	0,63	0,63
5	0,63	0,63
6	0,63	0,63
7	0,63	0,63
8	0,49	0,49
9	0,63	0,63
10	0,49	0,49

3.2. k-NN (k-en yakın) Komşu Tabanlı Veri Ağırlıklandırma İşlemi

Bu ön-işleme metodu, *k-NN* (*k*- en yakın komşu) algoritması kullanılarak her bir veri örneğinin özellik değerlerinin yeniden belirlenmesine dayanır. Bu ön-işleme metodunun akış şeması Şekil 3.3' de verilmiştir.



Şekil 3.3 k-NN tabanlı veri ağırlıklandırma ön-işleme metodunun blok diyagramı

Ağırlıklandırma işlemi, ayrı ayrı her bir özellik için yapılır. Örnek olarak, bütün veri setinde ilk özelliğin bütün değerleri hesaba katılır ve *k-NN* şemasına göre değiştirilir ve sonra aynı işlem veri örneklerinin ikinci özelliği için yapılır ve bu işlem bütün özellikler için devam eder. Bir veri örneğinin bir özellik değerinin belirlenmesinde yapılan işlem şu şekilde açıklanabilir: *i* özellik indeksi ve *j* veri örneğidir. Yani, *j*. veri örneğinin yeni değerinin araştırılmasıdır. Şekil 3.3' de görüldüğü gibi, yapılması gereken ilk şey, bütün diğer özellikler için yeni veri

değerlerini bulmaktır. Burada, basit mutlak fark, bir fark ölçümü olarak kullanılır (Eşitlik 3.3):

$$d(x_i(j), x_i(m)) = | x_i(j) - x_i(m) | \tag{3.3}$$

$x_i(j)$, *i.* özelliğin *j.* değeriyken, $x_i(k)$ *i.* özelliğin *k.* değeridir. Farklar hesaplandıktan sonra, en yakın *k* özellik değerleri alınır ve bu değerlerin ortalama değeri hesaplanır (Eşitlik 3.4):

$$mean_value(j) = \frac{1}{k} \sum_{\substack{attr_n \in attr_k, \\ n=1}}^{k} attr_n \tag{3.4}$$

$attr_n$, *k*- en yakın özellik değerlerinde *n.* özelliğin değerini gösterir ve *k*, ilgili *i* özellik değerinde en yakın noktaların sayısıdır. Bu hesaplanan ortalama değer, bu hesaplamaların yapıldığı ilgili özellik değerinin yeni değeri olarak alınır. Yani, *j.* veri örneğinin *i.* değeri bu değerle değişir. Bu işlem, aynı durumda veri örneğinin her bir özellik değeri için yapılır (Polat 2007e). *k-NN tabanlı veri ağırlıklandırma ön-işleme* yöntemindeki amaç, sınıf içindeki değişimi (varyansı) azaltmaktır. Bu sayede sınıflar arasındaki ayrışım daha da arttırmış olur.

k-NN tabanlı veri ağırlıklandırma ön-işlemeye bir örnek verelim. *x* ve *y* değişkeninden oluşan bir veri kümemiz olsun. Tablo 3.5 *x* ve *y* değişkeninden oluşan veri kümesini göstermektedir.

Tablo 3.5 *x* ve *y* değişkeninden oluşan veri kümesi

x	***y***
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3.0
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

İlk olarak veri kümesindeki her bir özellik içinde veriler arasındaki farklar hesaplanır. Sonra kullanıcı tarafından girilen *k* değerine göre ağırlıklandırma işlemini yapacağımız veriye en yakın *k* tane değer alınır ve bunların ortalaması alınarak yeni değeri belirlenmiş olur. Ortalaması alınırken kendi değerini de hesaba katmaktadır.

Örnek olarak *k*' nın 3, 4 ve 5 değerlerini örnek verelim. Tablo 3.6, *k*' nın 3 değeri için Tablo 3.5' in ağırlıklandırılmış halini göstermektedir. Tablo 3.7, *k*' nın 4 değeri için Tablo 3.5' in ağırlıklandırılmış halini göstermektedir. Tablo 3.8 de *k*' nın 5 değeri için Tablo 3.5' in ağırlıklandırılmış halini göstermektedir.

Tablo 3.6 *x* ve *y* değişkenin k-NN tabanlı veri ağırlıklandırma sonrası yeni dağılımı (k=3 için)

x	***y***
2,33	2,43
0,86	0,9
2,16	2,86
2,03	2,43
2,63	2,86
2,33	2,86
2,03	1,43
0,86	0,9
1,5	1,43
1,2	0,9

Tablo 3.7 *x* ve *y* değişkenin k-NN tabanlı veri ağırlıklandırma sonrası yeni dağılımı (k=4 için)

x	***y***
2,25	2,55
1,025	1,075
2,25	2,75
2,1	2,225
2,525	2,75
2,25	2,75
2,1	1,625
1,025	1,075
1,625	1,625
1,025	1,075

Tablo 3.8 *x* ve *y* değişkenin k-NN tabanlı veri ağırlıklandırma sonrası yeni dağılımı (k=5 için)

x	***y***
2,18	2,64
1,2	1,18
2,18	2,64
1,98	2,1
2,42	2,64
2,18	2,64
2,18	1,48
1,2	1,18
1,5	1,48
1,2	1,18

3.3. Genelleştirilmiş Ayrışım Analiz (GAA) Yöntemi

GAA, sıradan lineer ayrışım analizinin *(LAA- Linear Discriminant Analysis* (*LDA*)) non-lineer bir genişlemesidir (G. Baudat 2000, Asl 2008). Giriş eğitme verisi, bir kernel fonksiyon ile nesnelerin farklı sınıflarının lineer olarak ayrıldığının farz edildiği yüksek boyutlu bir özellik uzayında haritalanır. *LAA* sonra, veriyi en iyi açıklayan vektörlerden ziyade sınıflar arasındaki en iyi ayrışımı yapan vektörleri araştıran haritalanmış veriye uygulanır (H.C.Kim 2002). Gerçekte, *LAA*' nın amacı, sınıflar arasındaki saçılmanın sınıflar içindeki saçılmaya oranını maksimize edecek bir dönüşüm bulmaktır. Buna ilaveten, veriyi açıklayan bağımsız özellikler verilirse, *LAA* istenilen sınıfların en büyük ortalama farkını sağlayan bu özelliklerin lineer bir birleşimini üretir (A.M. Martinez 2001). Sonuç olarak, eğer veri setinde N sınıf varsa, özellik uzayının boyutu *N-1*' e azaltılabilir.

Farz edelim ki, *X* eğitme veri seti *N* sınıftan ve *M* özellikten oluşsun. x_{pq}, *p.*sınıfta *q.* özellik vektörünü gösterir, n_p *p.* sınıfın sınıf büyüklüğüdür ve ϕ non-lineer bir haritalama fonksiyonudur. Böylece *X* uzayı, daha yüksek bir boyutlu özellik uzayında haritalanır: $\phi : x_i \in R^f \rightarrow \phi(x_i) \in R^F, F \rangle\rangle f$ ().

$\phi(x_i)$ incelemeleri, *F* uzayında merkez olarak farz edilir. *GAA*, *X* eğitim veri setini *Y* yeni bir veri setinde izdüşümü alınmadan önce, *F* özellik uzayında *B* sınıflar arasındaki saçılma matrisi ve *V* sınıf için saçılma matrisi eşitlik 3.5 ve 3.6 ile tanımlanır:

$$V = \frac{1}{M} \sum_{p=1}^{N} \sum_{q=1}^{n_p} \phi(x_{pq}) \phi^T(x_{pq}) \tag{3.5}$$

$$B = \frac{1}{M} \sum_{p=1}^{N} n_p \left(\sum_{q=1}^{n_p} \phi(x_{pq}) \right) \left(\sum_{r=1}^{n_p} \phi^T(x_{pr}) \right)^T \tag{3.6}$$

GAA' nın amacı, *F* özellik uzayında sınıflar içindeki uyumsuzluğu (intra-classes inertia) minimize eden ve sınıflar arasındaki uyumsuzluğu (inter-classes

inertia) da maksimize eden *v* projeksiyon (izdüşüm) vektörünü bulmaktır. Bu durum eşitlik 3.7' deki maksimizasyon problemine eşittir (Asl 2008):

$$v = \arg\max_{v} \frac{v^T B v}{v^T V v} \tag{3.7}$$

v projeksiyon vektörü, $\lambda = \frac{v^T B v}{v^T V v}$ öz değeriyle birleştirilmiş $V^{-1}B$ matrisinin öz vektörüdür. *v*' nin bütün çözümleri, $\phi(x)$' in etrafında yatar. Bu yüzden, a_i genişleyen katsayılardan oluşsun. *v* projeksiyon vektörünü eşitlik 3.8' de yeniden yazılırsa;

$$v = \sum_{i=1}^{M} a_i \phi(x_i) \tag{3.8}$$

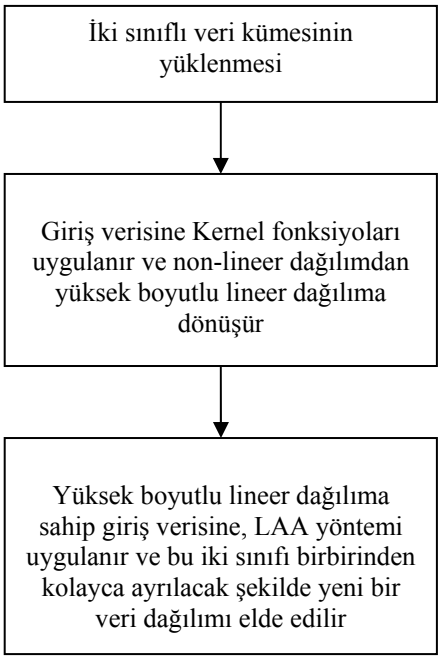
şeklinde olur.

$k(x_i, x_j) = k_{ij} = \phi(x_i)\phi(x_j)$ kernel fonksiyonu kullanılarak ve $K = (k_{ij})_{i=1,...M;\, j=1,...M}$ kernel matrisinde öz vektörlerin ayrışımını gerçekleştirmek için, $a = a/(a^T K a)^{1/2}$ her bir projeksiyon vektörü için *M* normalize edilmiş genişleme katsayıları elde edilir (G. Baudat 2000).

Test veri kümesinden bir *x* özellik vektörü için, v_i *i.* öz vektördeki projeksiyon (izdüşümü) eşitlik 3.9 ile hesaplanabilir:

$$y^i = v^{i^T} \phi(x) = \sum_{j=1}^{M} a_j^i \phi(x_j).\phi(x) = \sum_{j=1}^{M} a_j^i k(x_j, x) \tag{3.9}$$

Bu eşitlikte, a_j^i, *i.* öz vektörün *j.* genişleme katsayısını gösterir (Asl 2008). *GAA*' ın blok diyagramı Şekil 3.4' de verilmiştir.



Şekil 3.4 Genelleştirilmiş ayrışım analizi yönteminin akış şeması

GAA'ya bir örnek verelim. *x* ve *y* değişkeninden oluşan bir veri kümemiz olsun. Tablo 3.9, *x* ve *y* değişkeninden oluşan veri kümesini göstermektedir.

Tablo 3.9 x ve y değişkeninden oluşan veri kümesi

Veri sayısı	**x**	**y**	**Sınıf etiketi**
1	2.5	2.4	0
2	0.5	0.7	0
3	2.2	2.9	0
4	1.9	2.2	0
5	3.1	3.0	0
6	2.3	2.7	1
7	2	1.6	1
8	1	1.1	1
9	1.5	1.6	1
10	1.1	0.9	1

GAA, danışmalı bir veri ön-işleme yöntemidir. Yukarıdaki örnekte iki sınıflı (sınıf 0 ve sınıf 1) 10 veriden oluşan bir veri kümesi verilmiştir. Bu veri kümesini *GAA*' ya verirken sınıflar ayrı ayrı verilir ve bu sınıfları birbirinden maksimum seviyede ayıracak düzlem bulunur. Bulunan düzleme göre veri dağılımı yeniden düzenlenir. Tablo 3.9' de verilen veri kümesi *GAA*' ya uygulandıktan sonra oluşan yeni veri dağılımı Tablo 3.10' da verilmiştir.

Tablo 3.10 x ve y değişkeninden oluşan veri kümesine GAA uygulandıktan sonraki veri dağılımı

Veri sayısı	x	y	Sınıf etiketi
1	0,87	0,57	0
2	-1,66	-1,42	0
3	0,49	1,16	0
4	0,11	0,34	0
5	1,64	1,28	0
6	0,62	0,93	1
7	0,24	-0,36	1
8	-1,03	-0,95	1
9	-0,39	-0,36	1
10	-0,90	-1,19	1

3.4. Benzerlik Tabanlı Veri Ağırlıklandırma Yöntemi

Bu çalışmada, Sun Yi ve arkadaşları (Sun Yi 2005) tarafından önerilen Birliktelik Eğilimine Dayanan Ağırlıklandırma (*Weighting Based on Trend Association-WBTA*) yönteminden esinlenerek *benzerlik tabanlı veri ağırlıklandırma yöntemi* önerildi. Bu yöntem, temel özellik değerleri arasındaki birliktelik eğiliminin benzerlik derecesine dayanarak bir özellik ağırlıklandırma işlemi önerir. Bu yöntemin avantajı, uzman tecrübesinden bağımsızdır. Sun Yi ve arkadaşlarının önerdiği $C_i(k)$ birliktelik katsayısı yerine kosinüs genliği (cosine amplitude), min-mak (minimum-maksimum) metodu, mutlak üstel (absolute exponential) yöntemi ve üstel benzerlik katsayısı (exponential similarity coefficient) yöntemlerini içeren benzerlik ölçümleri kullanıldı (Polat 2008d).

WBTA yöntemi (Sun Yi 2005) şu şekilde açıklanabilir:

$Attr=\{X_i|\ i=1,2,...,n\ \}$ ilk özellik kümesini göstersin ve i, özellik sayısıdır. Her bir X_i, özelliklerle birleştirilmiş değerler serisidir:

$$X_i =\{x_i(k)|\ k=1,2,...,m\ \},\ (i=1,2,...,n) \tag{3.10}$$

Bu eşitlikte, k veri sayısıdır.

$W_i =\{w_i|\ i=1,2,...,m\ \}$, özellik ağırlık kümesidir.

Tanımlama 1: Benzerlik ölçümleri, özellikle *k* veri örneği olan, veri örneklerinin toplamı kullanarak yapılır. Aşağıda açıklandığı gibi dört farklı benzerlik ölçümü kullanılmıştır:

A. Kosinüs Genliği Benzerlik Yöntemi (Cosine Amplitude):

Kosinüs Genliği, kullanışlı bir yöntemdir. Burada, veri örneklerinin her biri, her bir noktanın komple bir açıklama için *n* koordinata gerek duyulduğu *n*-boyutlu bir uzayda bir nokta olarak düşünülür. $r_{i,j}$ ilişkisinin her bir elemanı, x_i and x_j olarak adlandırılan iki veri kümesinin ikili karşılaştırılması ile sonuçlanır. x_i veri örneği ve x_j veri örneği arasındaki ilişkinin güçlülüğü, $r_{i,j} = \mu_R(x_i, x_j)$ güçlülüğü açıklayan bir üyelik değeriyle verilir. İlişki matrisi, *mxm* boyutunda olacaktır. Kosinüs genlik yöntemi aşağıdaki durumda $r_{i,j}$' yi hesaplar ve $0 \le r_{i,j} \le 1$ aralığından olur (R. Setiono 2000). Eşitlik 3.11, Sun Yi ve arkadaşlarının kullandığı birliktelik formülünü göstermektedir (Sun Yi 2005).

$$r_{ij} = \frac{\left| \sum_{j=2}^{n} \sum_{i=1}^{n} x_{ik} x_{jk} \right|}{\sqrt{(\sum_{i=1}^{n} (x_{ik})^2)(\sum_{j=2}^{n} (x_{jk})^2)}} \qquad (3.11)$$

Burada, k , veri sayısıdır ve $k = 1,2,...,m$

B. Min-Mak. Benzerlik Yöntemi:

Kosinüs genlik yönteminden hesapsal olarak daha basit olan diğer bir yöntem, min-mak benzerlik yöntemidir. Min-mak birleşim yöntemine benzer tarafları olmasına rağmen, bu benzerlik yöntemi, birleşmeden farklıdır. Benzerlik, $x_{i,j}$ veri çiftlerinde minimum ve maksimum operatörlerle bulunuşu eşitlik 3.12'de verilmiştir (Sun Yi 2005).

$$r_{ij} = \frac{\sum_{j=2}^{n} \sum_{i=1}^{n} \min(x_{ik} x_{jk})}{\sum_{j=2}^{n} \sum_{i=1}^{n} \max(x_{ik} x_{jk})} \tag{3.12}$$

Burada, k , veri sayısıdır ve $k = 1,2,...,m$

C. Mutlak Üstel Benzerlik Yöntemi (Absolute exponential):

Burada, r_{ij} eşitlik 3.13' de verilmiştir (Sun Yi 2005):

$$r_{ij} = \exp(-\sum_{j=2}^{n} \sum_{i=1}^{n} \left| x_{ik} - x_{jk} \right| \tag{3.13}$$

Burada, k , veri sayısıdır ve $k = 1,2,...,m$

D. Üstel Benzerlik Katsayı Yöntemi (Exponential similarity coefficient method):

Burada, r_{ij} eşitlik 3.14' de verilmiştir (Sun Yi 2005):

$$r_{ij} = (\frac{1}{n}) \sum_{j=2}^{n} \sum_{i=1}^{n} \exp(\frac{-3(x_{ik} - x_{jk})^2}{4(S_k)^2} \tag{3.14}$$

Burada, k , veri sayısıdır ve $k = 1,2,...,m$

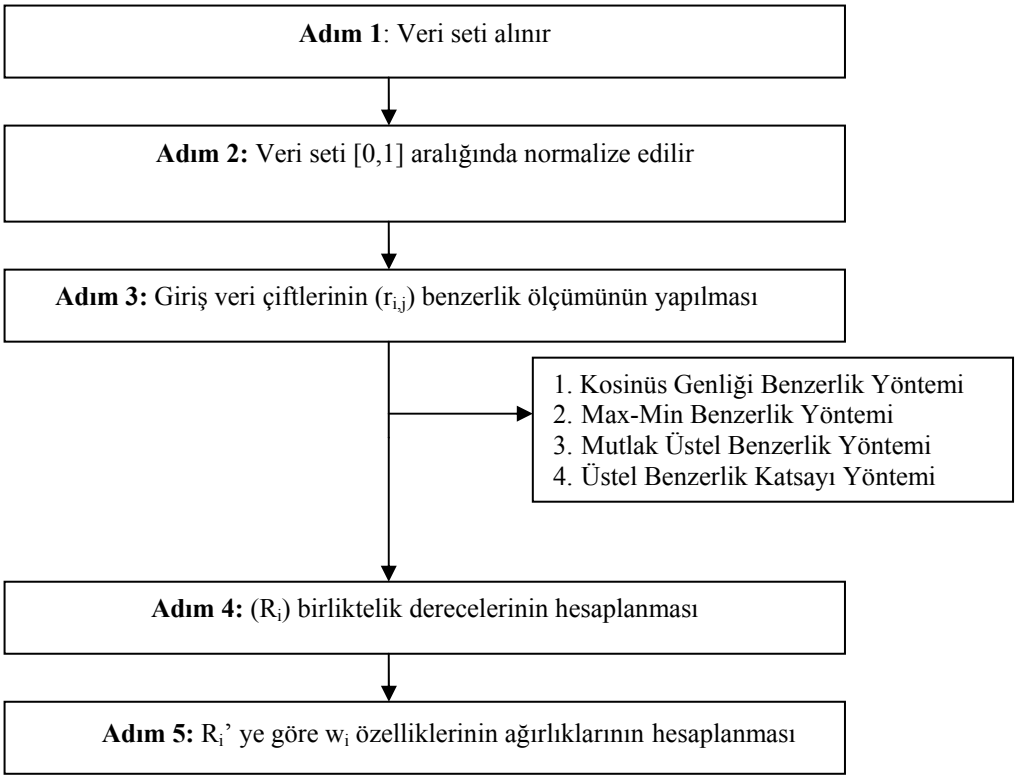
Tanımlama 2: R_i, X_i ve X_j arasındaki birliktelik derecesidir (Sun Yi 2005):

$$R_i = \frac{1}{n} \sum_{k=1}^{m} r_{i,j}(k) \tag{3.15}$$

Tanımlama 3: w_i, X_i özelliğinin ağırlığıdır:

$$w_i = \frac{R_i}{\sum_{i=1}^{n} R_i} \tag{3.16}$$

Eğimlilik dizisinin benzerlik derecesi, X_i ve X_j *((i=1,2,..,n, j=2,3,...,n)* arasındaki R_i birliktelik derecesi ve w_i benzerliği ile açıklanır. *Benzerlik tabanlı veri ağırlıklandırma ön-işlemenin* blok şeması Şekil 3.5' de verilmiştir.



Şekil 3.5 Benzerlik tabablı veri ağırlıklandırma ön-işlemenin blok şeması

Benzerlik tabanlı veri ağırlıklandırma ön-işleme' ye bir örnek verelim. *x* ve *y* değişkeninden oluşan bir veri kümemiz olsun. Tablo 3.11, *x* ve *y* değişkeninden oluşan veri kümesini göstermektedir. Tablo 3.12' de veri kümesinin [0,1] aralığında normalize edilmiş değerlerini göstermektedir. İlk olarak veri seti [0,1] aralığında normalize edilir, daha sonra kullanıcı tarafından seçilecek benzerlik fonksiyonuna göre değişkenler arasındaki benzerlik değerleri hesaplanır. Sonra, hesaplanan benzerlik değerlerine göre birliktelik katsayıları hesaplanır ve bu değere göre veri kümesi ağırlıklandırılır. Tablo 3.13, kosinüs genliği benzerlik fonksiyonuna göre veri kümesinin ağırlıklandırılmış yeni dağılımını gösterir. Tablo 3.14, mutlak üstel benzerlik fonksiyonuna göre veri kümesinin ağırlıklandırılmış yeni dağılımını gösterir. Tablo 3.15 de, üstel benzerlik katsayı fonksiyonuna göre veri kümesinin ağırlıklandırılmış yeni dağılımını gösterir. Max-min benzerlik fonksiyonunda minimum değerler 0 çıktığı için ağırlıklandırma işlemi yapılmadı.

Tablo 3.11 x ve y değişkeninden oluşan veri kümesi

Veri sayısı	**x**	**y**
1	2.5	2.4
2	0.5	0.7
3	2.2	2.9
4	1.9	2.2
5	3.1	3.0
6	2.3	2.7
7	2	1.6
8	1	1.1
9	1.5	1.6
10	1.1	0.9

Tablo 3.12 *x* ve *y* değişkenin [0,1] aralığındaki normalize değerleri

Veri sayısı	***x***	***y***
1	0,76	0,73
2	0	0
3	0,65	0,95
4	0,53	0,65
5	1	1
6	0,69	0,86
7	0,57	0,39
8	0,19	0,17
9	0,38	0,39
10	0,23	0,08

Tablo 3.13 *x* ve *y* değişkenin, kosinüs genliği benzerlik fonksiyonuna göre veri kümesinin ağırlıklandırılmış yeni dağılımı

Veri sayısı	***x***	***y***
1	1	1
2	1	1
3	0,90	0,85
4	0,87	0,90
5	0,85	0,82
6	0,83	0,83
7	0,84	0,85
8	0,86	0,86
9	0,87	0,88
10	0,44	0,44

Tablo 3.14 *x* ve *y* değişkenin, mutlak üstel benzerlik fonksiyonuna göre veri kümesinin ağırlıklandırılmış yeni dağılımı

Veri sayısı	***x***	***y***
1	0,97	1
2	1	1
3	0,74	1
4	0,88	1
5	1	1
6	0,84	1
7	0,83	1
8	0,98	1
9	0,99	1
10	0,43	0,5

Tablo 3.15 *x* ve *y* değişkenin, üstel benzerlik katsayı benzerlik fonksiyonuna göre veri kümesinin ağırlıklandırılmış yeni dağılımı

Veri sayısı	***x***	***y***
1	0,99	0,45
2	1	0,45
3	0,76	0,45
4	0,95	0,45
5	1	0,45
6	0,91	0,45
7	0,90	0,45
8	0,99	0,45
9	0,99	0,45
10	0,46	0,22

4. BİYOMEDİKAL VERİLERİ SINIFLAMADA KULLANILAN YÖNTEMLER

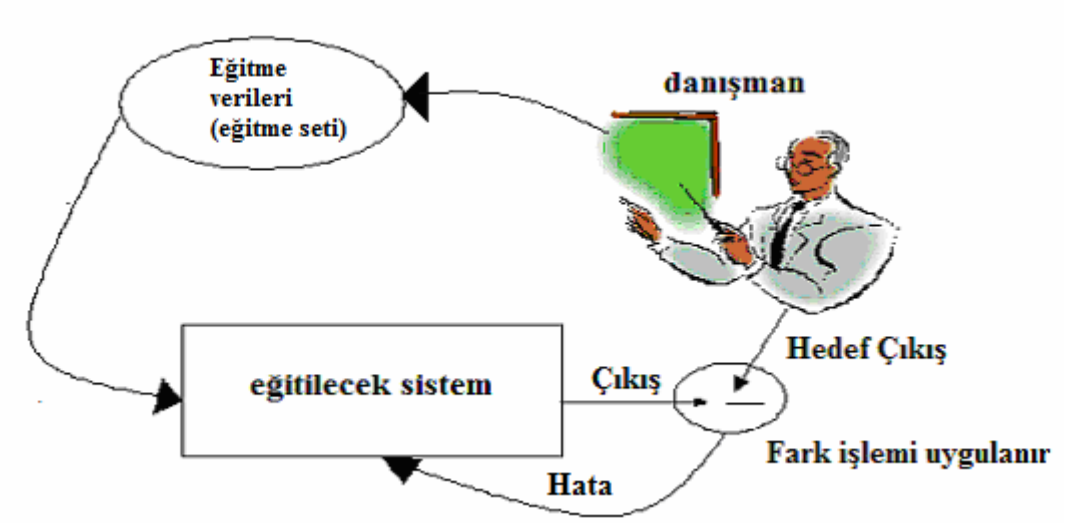
Bu bölümde, UCI makine öğrenmesi veri tabanından alınan medikal veri setleri (özelliği belirlenmiş) ve özelliği çıkarılmış sürekli zamanlı Doppler, PERG ve VEP sinyallerini sınıflamada kullanılan yöntemler tanıtılacaktır. Bu yöntemler, *Yapay Sinir Ağları*, *ANFIS*, *C4.5 karar ağaçları*, *AIRS* (*Artificial Immune Recognition System*) ve *Bulanık Kaynak Dağılımlı-AIRS* sınıflama algoritmalarıdır.

4.1. Sınıflama Algoritmaları

Makine öğrenmesi algoritmalarının amacı, nesneleri, nesneler üzerinde yapılan ölçümlere dayanarak sınıflara tahsis etmektir. Sınıflama işlemi temel olarak iki aşamadan oluşmaktadır. Birinci aşamada sınıflama sistemi, nesnelerden alınan veriler ışığında eğitilir. Eğitme işlemi bittikten sonra, ikinci aşamada herhangi bir nesneden alınan veri, sınıflama sistemine giriş olarak verilerek bunun hangi sınıfa ait olduğuna karar verilir.

Sınıflama algoritmaları genel olarak iki gruba ayrılır: Danışmanlı öğrenme ve Danışmansız öğrenme. Danışmanlı öğrenmede, girişler ve çıkışlar arasında bir model oluştururken eğitme işlemi sırasında giriş verilerinin etiketleri (sınıfları) de sisteme verilir. Böylece hedef çıkış değer ile hesaplanan çıkış arasındaki hata değeri minimize edilmiş olur. Danışmanlı öğrenmeye ait şematik gösterim Şekil 4.1' de verilmiştir. Danışmanlı öğrenmede, eğitme verileri eğitilecek modele veya sisteme verilerek sistem parametreleri ayarlanır ve ayarlanan modele göre test işlemi gerçekleştirilir. Kullanılan sınıflayıcıya göre modeli oluşturmak için kullanılan ve yenilenen parametreler değişmektedir. Örneğin, *ANFIS* algoritması için bulanık çıkarım sistemi parametrelerinin güncellenmesi, *YBTS* ve *Bulanık-YBTS*

algoritmaları için eğitme sonucu hafıza hücrelerinin bulunması, *yapay sinir ağları* için de ağırlık katsayılarının bulunması örnek olarak verebilir.



Şekil 4.1 Danışmanlı öğrenme sisteminin şematik gösterimi

Danışmansız öğrenme de ise eğitilecek modele eğitme verilerinin sınıf etiketleri verilmez, eğitme verileri arasındaki benzerlik, uzaklık gibi fonksiyonlar kullanılarak bir model oluşturulur. Genel olarak danışmansız öğrenme yöntemleri kümeleme algoritmalarıdır.

4.1.1. Çok katmanlı algılayıcı yapay sinir ağları

Bir *Yapay Sinir Ağı* (*YSA*), insan beynine benzer bir yapıya sahip olan birbirleri ile bağlantılı elemanlardan oluşan matematiksel bir modeldir. Bir *yapay sinir ağı* modeli, giriş ve çıkışlar arasındaki ilişkiyi açıklamak için uygun veri örnekleriyle eğitilir. Böylece eğitilen yapay sinir ağına yeni bir giriş yapıldığında uygun ve doğru çıkışlar elde edilebilir (Simpson 1989).

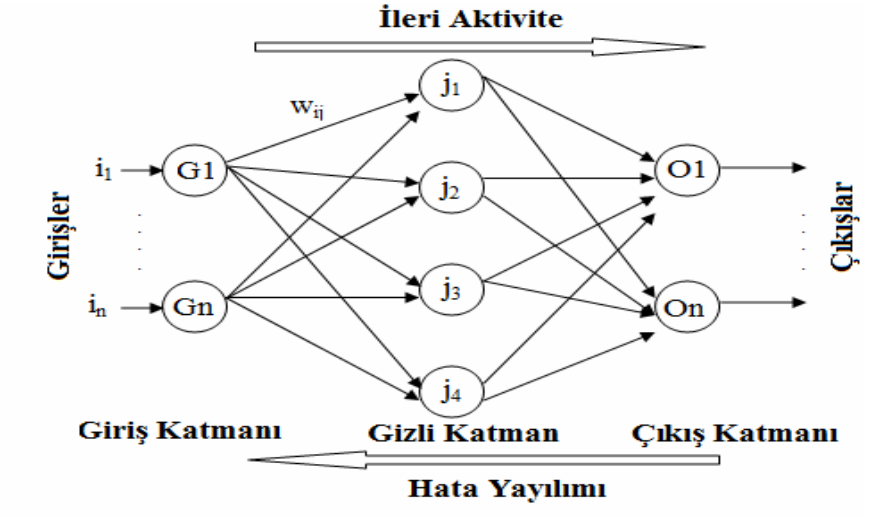
Medikal verileri sınıflamak için çok katmanlı bir ileri beslemeli *YSA* modeli kullanıldı. Bu tip bir *YSA* modelinin avantajı, özellikle sinyal işleme gibi hız gerektiren uygulamalarda eğitilmiş bir ağ sonucu daha hızlı yakınsar. *YSA* eğitimi, genellikle bir nonlineer en az kareler problemine bağlı olarak formülize edilir (Haykin 1994, Kara ve ark. 2006).

4.1.2. Geriye yayılım (Back-propagation) algoritması

Geri yayılım algoritması, çok katmanlı bir perseptronun bağlantı ağırlıklarını ayarlayan geniş olarak kullanılan bir eğitme prosedürüdür (Rumelhart ve ark. 1986). Özellikle, bir Levenberg–Marquardt (*LM)* algoritması, maksimum komşuluk fikrine dayanan en az kareler metodunu kullanan bir tahmin algoritmasıdır. Şekil 4.2'de gösterildiği gibi, çok katmanlı bir perseptron üç katmandan oluşur: giriş katmanı, çıkış katmanı ve gizli katmandır. Her bir katman, önceden öntanımlanmış nöronlardan oluşur. Giriş katmanındaki nöronlar sadece, x_i giriş sinyallerini gizli katmandaki nöronlara dağıtmak için tampon bölge olarak çalışırlar. Gizli katmandaki her bir j düğümü, gizli katmandan w_{ij} bağlantılarının kuvvetleriyle onları ağırlıklandırıldıktan sonra x_i giriş sinyallerini toplar ve bir f toplam fonksiyonu olarak kendi çıkışlarını eşitlik (4.1)' deki gibi hesaplar:

$$y_j = f(\sum w_{ij} x_i) \tag{4.1}$$

Burada, y fonksiyonu, etki eden bütün sinyallerin ağırlıklandırılmış toplamını bir nörona dönüştürmek için gerekli olan bir çıkış fonksiyonudur. f aktivasyon fonksiyonu, basit bir eşik fonksiyonu, sigmoid, hiperbolik, tanjant veya RBF (radial basis function) fonksiyonu olabilir. Çıkış katmanındaki nöronların çıkışı da benzer yolla hesaplanır (Kara ve ark. 2006, Beale 1990, Güler ve ark. 2003, Haykin 1994, Türkoğlu ve ark. 2002, Wright ve ark. 1999).



Şekil 4.2 Çok katmanlı bir perseptron

Bir network' un eğitilmesi, farklı öğrenme algoritmaları kullanılarak network ağırlıklarını ayarlamaktan oluşur. Bir öğrenme algoritması, *t* zamanında *i* ve *j* nöronları arasındaki bir bağlantının ağırlığındaki $\Delta w_{ij}(t)$ değişimini verir. Levenberg-Marquardt öğrenme algoritması için, ağırlıklar eşitlik (4.2) ve (4.3)'e göre yenilenilir (Kara ve ark. 2006): i_1 i_2 i_3 i_n

$$\Delta w_{ij} = [J^T(w)J(w) + \mu I]^{-1} J^T(w)E(w) \tag{4.2}$$

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t) \tag{4.3}$$

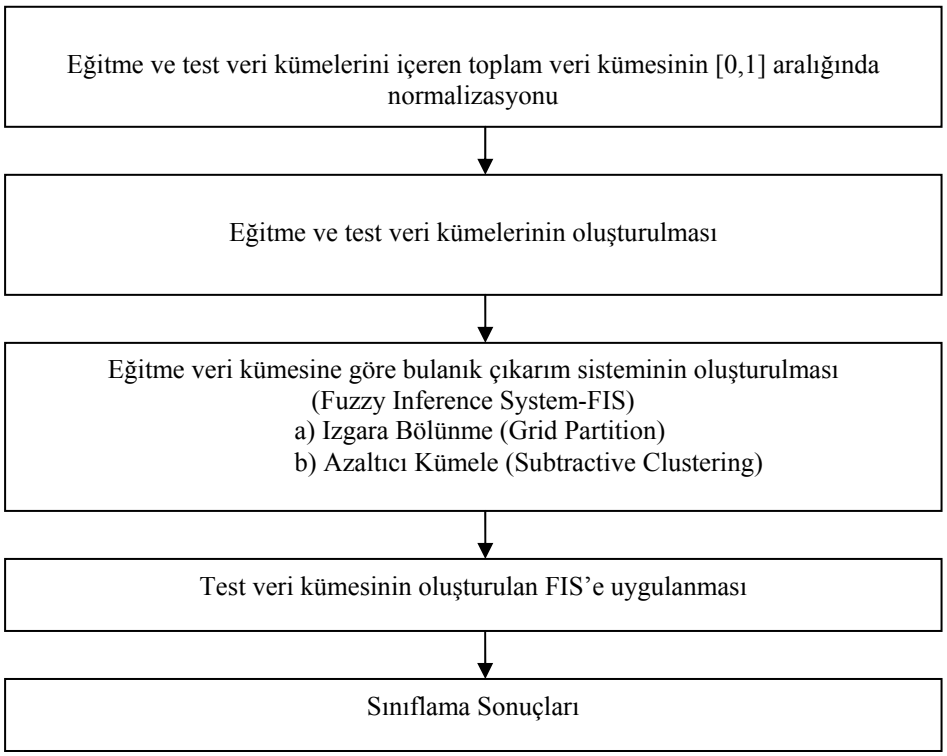
Eşitlik (4.2) ve (4.3)' de, *J*, Jacobian matristir, *m* ise bir sabittir. *I*, birim matrisi ve *E(w)* bir hata fonksiyonudur (Kara ve ark. 2006, Übeyli ve ark. 2003).

4.2. ANFIS (Adaptive Network Based Fuzzy Inference System)

ANFIS, yapay sinir ağlarının adaptiflik özelliği ile bulanık mantığın niteliksel çıkarımını birleştiren hibrid bir öğrenme algoritmasıdır. ANFIS, eğimli azalma ve en küçük kareler yöntemlerini birleştiren etkin ve hibrid bir öğrenme algoritması

kullanır. En küçük kareler yöntemi, hızlı öğrenmeyi sağlarken, eğimli azalma en küçük kareleri yönteminin temel fonksiyonlarını üreten üyelik fonksiyonlarını değiştirir (Jang 1992, Jang 1993).

ANFIS eğitme ve test akış şemaları Şekil 4.3' de verilmiştir.



Şekil 4.3 ANFIS' in öğrenme ve test akış diyagramı

Eğitme işleminde kullanılan adımları kısaca açıklarsak:

a) Veri Normalizasyonu: Veri bu dönüşüm sayesinde [0,1] aralığına dönüşür. Tez çalışmasında normalizasyon yöntemi olarak minimum-maksimum (min.-mak.) normalizasyon yöntemi kullanılmıştır. Min.-mak. normalizasyon eşitliği, eşitlik 4.4' de verilmiştir.

$$v'(i) = \frac{v(i) - \min_A}{\max_A - \min_A} \tag{4.4}$$

Burada, *v(i)* normalizasyonu alınacak *v* matrisinin *i.* elemanıdır. min_A ve max_A, *v* matrisindeki *A*. özelliğin minimum ve maksimum değerleridir. *v'(i)*, normalizasyondan sonraki yeni değerdir.

b) Bulanık Çıkarım Sisteminin Oluşturulması: Bulanık çıkarım sistemlerinin oluşturulmasında iki sistem kullanılmaktadır bunlar; a) Izgara Bölünme (Grid Partition) b) Azaltıcı Kümeleme (Subtactive Clustering) dir. Bulanık çıkarım sistemlerinin yapısı, "*böl-fethet*" modeline benzer, bir bulanık kuralın öncesi, lokal bir bulanık bölgeyi tanımlarken, sonraki ise değişik oluşumlar tarafından bölge içindeki davranışı tanımlar (Jang ve ark. 1997).

1) ***Izgara Bölünme (Grid Partition)***: Bu bölünme metodu kontrolöre giriş olarak sadece birkaç durum değişkenini içeren bir bulanık kontrolörün tasarımında sıklıkla seçilir. Bu bölünme stratejisi, her bir giriş için sadece üyelik fonksiyonlarının küçük bir sayısına ihtiyaç duyar. Her ne kadar böyle olsada, kısmen büyük bir sayıda girişe sahip olan giriş değişkenlerinde, problemlerle karşılaşabilir. Örnek olarak, her bir girişte 2 üyelik fonksiyonu ve 10 girişli bir bulanık model, büyümenin genişlemediği bulanık *if-then* kurallarıyla sonuçlanacaktır. "*The curse of dimensionality (boyut belası)*" olarak tercih edilen bu problem diğer bölümleme stratejileri tarafından önlenebilir (Jang ve ark. 1997).

2) ***Azaltıcı Kümele (Subtactive Clustering):*** Dört değişkenli ve her bir boyutun 10 ızgara satırlığa sahip olduğu bir kümeleme problemi, geliştirilmesi gereken 10^4 ızgara noktalarında sonuçlanacaktır. Bu probleme alternatif bir yaklaşım, veri noktalarının (ızgara olmayan noktalar) küme merkezleri için aday olarak düşünüldüğü Chiu tarafından önerilen azaltıcı kümele (subtractive clustering) dir. Hesaplama, basit olarak veri sayısıyla orantılıdır ve düşünülen problemin boyutundan bağımsızdır (Jang ve ark. 1997).

Bir M boyutlu uzayda $\{x_1,....,x_n\}n$ veri noktasının bir koleksiyonunu düşünelim. Genelleştirme kaybı olmadan, veri noktaları bir hiperküp içerisinde normalize edildiği farz edilir. Her bir veri noktası küme merkezleri aday olduğu için, x_i veri noktasında bir yoğunluk ölçümü eşitlik (4.5) ile tanımlanır (Jang ve ark. 1997):

$$D_i = \sum_{j=1}^{n} \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{(r_a / 2)^2}\right) \tag{4.5}$$

r_a, pozitif bir sabittir. Bu yüzden, bir veri noktası eğer birçok komşu veri noktasına sahip olursa yüksek bir yoğunluk değerine sahip olacaktır. r_a çapı bir komşuyu tanımlar; bu çapın dışında kalan veri noktaları sadece yoğunluk hesabına katkıda bulunurlar.

Her bir veri noktasının yoğunluk ölçümü hesaplandıktan sonra, yüksek yoğunluk ölçümlü veri noktası ilk küme merkezi olarak seçilir. x_{c1} seçilecek nokta olsun ve D_{c1} de onun yoğunluğu olsun. Sonra, her bir x_i veri noktası için yoğunluk ölçümü eşitlik (4.6) ile revize edilir (Jang ve ark. 1997):

$$D_i = D_i - D_{c1} \exp\left(-\frac{\left\|x_i - x_{c1}\right\|^2}{(r_b / 2)^2}\right) \tag{4.6}$$

r_b, pozitif bir sabittir. Bu yüzden, x_{c1} ilk küme noktasına yakın olan veri noktaları, sonraki küme merkezine bağlı olarak muhtemelen seçilmeyen noktaları oluşturmak için büyük ölçüde azaltılmış yoğunluk ölçümlerine sahip olacaktır. r_b sabiti, yoğunluk ölçümünde ölçülebilir azalmalara sahip olan bir komşu tanımlar. r_b sabiti, birbirine yakın uzayda küme merkezlerini korumak için r_a çapından normalde büyüktür; genel olarak r_b, $1.5r_a$ ya eşittir (Jyh-Shing Roger Jang ve ark. 1997).

Her bir veri noktası için yoğunluk ölçümü revize edildikten sonra, x_{c2} sonraki küme merkezi seçilir ve veri noktaları için yoğunluk ölçümlerinin hepsi tekrar revize edilir. Bu işlem, küme merkezlerinin uygun bir sayısı üretilene kadar sürer (Jang 1993, Jang ve ark. 1997).

c) Oluşturulan Bulanık Çıkarım Sisteminin Hibrid Öğrenme Algoritması ile Optimizasyonu: ANFIS ağına ait parametrelerin güncellenmesinde, iki aşamalı bir öğrenme algoritması olan, hibrid öğrenme algoritması kullanılmıştır. Hibrid öğrenme algoritmasında, ANFIS ağ yapısına ait parametreler, giriş ve çıkış parametreleri olarak iki kısımda ele alınır. Toplam parametre kümesi; $S = S1 + S2$ şeklinde ifade edilecek olursa, $S1$ giriş parametrelerine (non-lineer parametreler), $S2$ de çıkış

parametrelerine (lineer parametreler) karşılık gelir. Şekil 4.4, *S1* ve *S2* parametrelerini göstermektedir. Hibrid algoritmanın ilk aşaması ya da ileri yön geçişi ile ANFIS ağına eşitlik (4.7) şeklinde ifade edilebilir (Jang 1993, Jang ve ark. 1997).

$$A\theta = B \tag{4.7}$$

Matrissel eşitlikte yer alan θ vektörü, *S2* çıkış parametrelerinin elemanlarından oluşan, bilinmeyen vektördür. Bu eşitlik, standart lineer en küçük-kareler (*least-squares*) problemini gösterir ve θ için en iyi çözüm, $\|A\theta - B\|^2$ nın minimum değeri olan, en küçük kareler tahmini (LSE) θ^* dır (Jang ve ark. 1997). Eşitlik (4.8), θ^*' nün değerini vermektedir (Jang 1993, Jang ve ark. 1997).

$$\theta^* = (A^T A)^{-1} A^T B \tag{4.8}$$

Burada, A^T, A matrisinin transpozu ise ve eğer $A^T A$ singular değilse, $(A^T, A)^{-1}$ A^T, A' nın sahte inversidir. Bunun yerine, yinelemeli olarak (*LSE*) formülü de kullanılabilir. Özel olarak eşitlik (4.9)' deki A matrisinin *i.* satırı a_i^T ile B matrisinin *i.* elemanı b_i^T ile gösterilsin. Bu takdirde θ vektörü, iteratif olarak eşitlik (4.9)' daki gibi hesaplanabilir (Jang ve ark. 1997).

$$\theta_{i+1} = \theta_i + S_{i+1} a_{i+1} (b_{i+1}^T - a_{i+1}^T \theta_i)$$

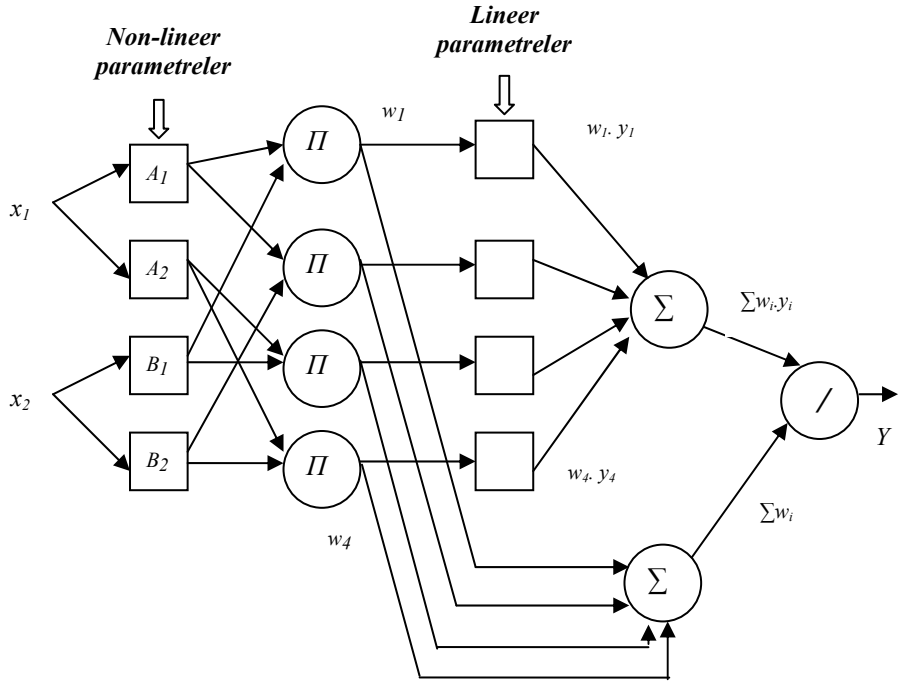
$$S_{i+1} = S_i - \frac{S_i a_{i+1} a_{i+1}^T S_i}{1 + a_{i+1}^T S_i a_{i+1}}, \quad i = 0,1,...,P-1 \tag{4.9}$$

Burada en küçük kareler tahmini θ^*, θ_P' ye eşittir. Eşitlik (4.8)' de ihtiyaç duyulan başlangıç şartları şunlardır; $\theta_o = 0$ ve $S_0 = \gamma I$. Burada γ, pozitif büyük bir sayı, I ise MxM boyutunda birim matristir (Jang ve ark. 1997).

Hibrid öğrenme algoritmasının geri yön geçişinde, *S2* çıkış parametreleri sabit tutulur ve ağın çıkışındaki hata sinyali geriye doğru yayılarak, eğim düşümü (gradient descent) yöntemiyle giriş parametreleri güncellenir. Geri yön geçişinde giriş parametreleri için güncelleme formülü eşitlik 4.10' daki gibidir (Jang ve ark. 1997);

$$\Delta\alpha = -\eta \frac{\partial E}{\partial \alpha} \tag{4.10}$$

Burada α herhangi bir giriş parametresini, η öğrenme oranını, E ise ağın çıkışındaki hata değerini göstermektedir. Tablo 4.1 her bir geçişteki olayları özetler (Güler ve ark. 2005, Jang ve ark. 1997).



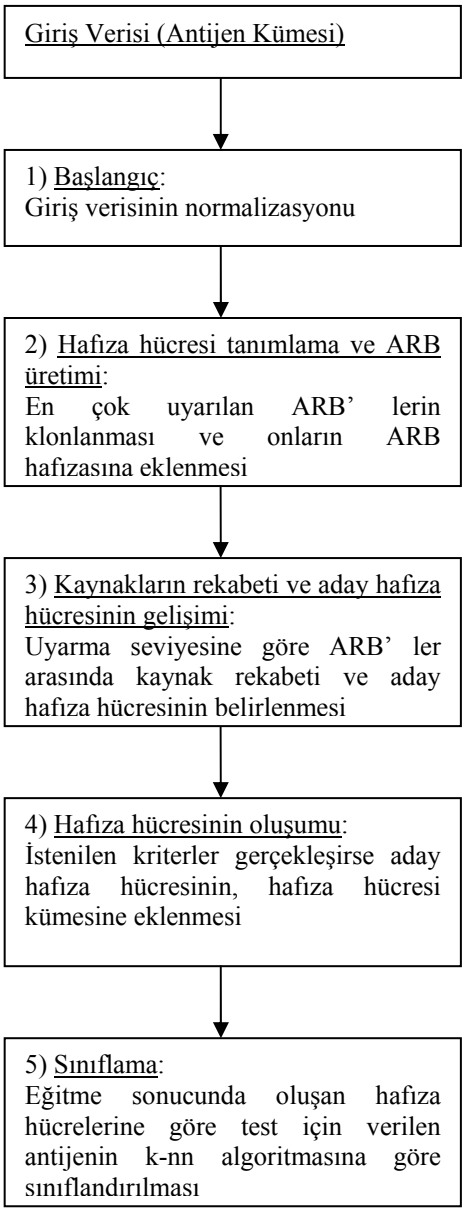
Şekil 4.4 İki girişi bir ANFIS mimarisi

Tablo 4.1. ANFIS için hibrid öğrenme prosedüründeki iki geçiş

	İleri Geçiş	**Geriye Geçiş**
Non-lineer parametreler	Sabit	Eğim iniş metodu
Lineer Parametreler	En az kareler tahmini	Sabit
Sinyaller	Düğüm çıkışları	Hata sinyalleri

4.3. Yapay Bağışıklık Tanıma Sistemi (Artificial Immune Recognition System) Sınıflama Algoritması

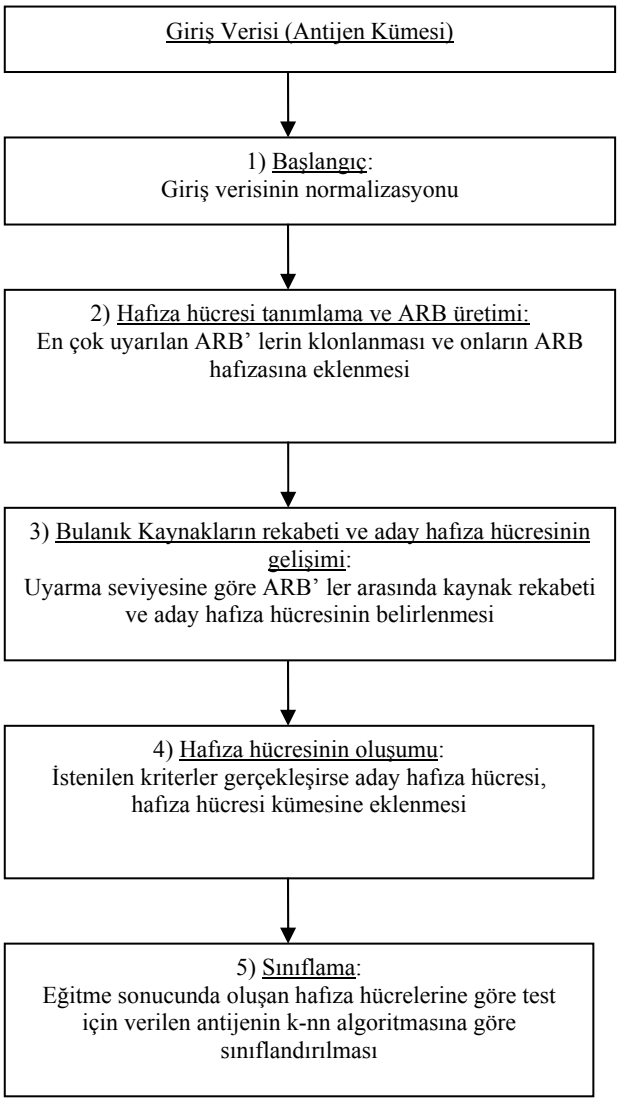
AIRS, doğal ve yapay bağışıklık sistemlerinde kullanılan mekanizmalar tarafından önerilen kaynak sınırlamalı bir yapay bağışıklık sınıflayıcısı olan yeni bir danışmalı öğrenme sistemidir. Şekil 4.5, *yapay bağışıklık tanıma sisteminin* (*YBTS*) blok şemasını göstermektedir. *AIRS* algoritması temel olarak dört aşamadan oluşmaktadır. İlk aşama veri normalizasyonu ve başlangıç aşamasıdır. Bir veri ön işleme aşaması olarak düşünülebilir. Başlangıç aşaması boyunca öncelikle veri kümesindeki herhangi iki veri arasındaki bütün nitelikler öklit uzaklığı [0,1] aralığında olacak şekilde normalize edilir. Normalizasyondan sonra duyarlılık eşiği hesaplanır. Eğitim kümeleri için duyarlılık eşiği bütün eğitim verilerinin özellik vektörleri üzerindeki ortalama duyarlılıktır. İkinci aşama bellek hücresi belirleme ve *ARB (Artificial Recognition Ball- Yapay Tanıma Topu)* oluşumudur. Bu aşamada, veri kümesindeki antijen sayısı kadar hafıza hücresi üretilir ve sistemdeki antijenler hafıza hücrelerine tek tek verilerek bu antijenlere en yakın olan hafıza hücreleri seçilerek klonlanırlar. Bu klonlanma sonucu ARB hücreleri oluşur. Üçüncü aşama aday bellek hücresinin gelişmesinde kaynaklar için rekabettir. Bu aşama, antijenleri en iyi tanıyan ARB hücrelerini bulmak ve sistemden zayıf ARB' leri uzaklaştırmak için yapılır. ARB' ler arasındaki kaynak dağılım mekanizması burada gerçekleştirilir. Eğitim algoritmasının son aşamasında, yerleşmiş bellek hücrelerinin kümesi içerisine aday bellek hücrelerinin potansiyel sunumu gerçekleşmektedir (Watkins 2001, Watkins 2005, Kodaz 2007).



Şekil 4.5 YBTS' nin eğitme ve test akış şeması

4.4. Bulanık Kaynak Dağılımlı Yapay Bağışıklık Tanıma Sistemi (Fuzzy-AIRS)

Şekil 4.6, *Fuzzy-AIRS* sınıflama algoritmasının blok diyagramını göstermektedir (Polat ve ark. 2006b).



Şekil 4.6 Fuzzy-AIRS sınıflama algoritmasının akış şeması

Bu sınıflama algoritmasında, *AIRS* sınıflama algoritmasında kullanılan *B* hücrelerini antikorlara dağıtmak için kullanılan kaynak dağılım mekanizması, bulanık mantık ile birleştirelerek hafızada kullanılan *B* hücresi sayısını uygun değerde ayarlanmış oldu. Bu mekanizma bulanık kaynak dağılım mekanizması olarak adlandırıldı. Bu mekanizmanın avantajı, uyarılma seviyesi düşük olan *ARB*' ye [0 ile 0.5 arasında olan] daha az kaynak, uyarılma seviyesi yüksek olan *ARB*' ye [0 ile 0.5 arasında olan] daha çok kaynak tahsis yapılarak kaynak paylaşımı optimum hale getirilmiştir. Bu sayede, *AIRS*' de daha az kaynak sayısı ile hem *AIRS* algoritmasının sınıflama performansını arttırabilir hem de eğitme süresini azaltabilir. Aşağıdaki bölümde bulanık kaynak dağılım mekanizması açıklandı (Polat ve ark. 2006b).

4.4.1. Bulanık kaynak dağılım mekanizması

AIRS' de kaynakların rekabeti, *AIRS*' nin performansını arttırmak için yüksek duyarlılıklı *ARB*' lere izin verir. Bu kaynak dağılımı mekanizmasına göre, kaynakların yarısı, antijenin sınıfındaki *ARB*' lere tahsis edilirken, kalan yarısı diğer sınıflara dağıtılır. *AIRS*' deki kaynak dağılımı, klonal oranla uyarılma oranının çarpılmasıyla bulunan bir sayıya göre yapılır. Baurav Marwah ve Lois Boggess' in çalışmalarında (Baurav Marwah ve Lois Boggess 2002), farklı bir kaynak mekanizması denediler. Onların mekanizmalarında, çok sıklıkla meydana gelen antijen sınıfları daha fazla kaynak edinirler. Hem klasik *AIRS* hem de Baurav Marwah ve Lois Boggess' in çalışmalarında, kaynak dağılımı, duyarlılıkla orantılı olarak yapılır. Bu lineerlik, daha fazla hafıza hücresini ve daha uzun süreli sınıflama zamanıyla sonuçlanan sistemde daha fazla kaynak kullanımına gerek duyar.

AIRS' de kaynak dağılımını optimum bir şekilde yapmak için, kaynak dağılım mekanizması, bulanık mantık ile yapıldı. Böylece, bulanık kurallar' dan dolayı kaynak dağılımında bir non-lineerlik oluştu. Yüksek duyarlılıklı *ARB*' ler ve düşük duyarlılıklı *ARB*' ler arasındaki kaynak dağılımındaki fark, klasik kaynak dağılımı mekanizmasından daha büyüktür.

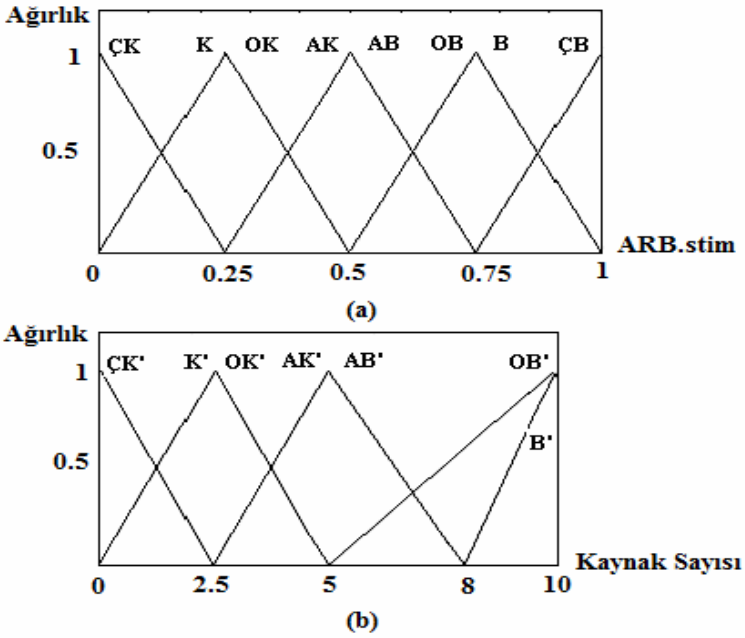
Bulanık kurallar kullanarak üyelik değerinin belirlenmesi ve çıkış değerinin elde edilmesi önemli olurken, diğer bir önemli nokta ise Tablo 4.2' de gösterilen giriş ve çıkış üyelik fonksiyonlarında kullanılan dilsel değerlerin belirlenmesidir.

Bulanık kaynak dağılım mekanizmasının giriş değişkeni, *ARB*' nin uyarılma seviyesidir böylece çıkış değişkeni, ilgili *ARB*' ye tahsis edilecek kaynakların sayısıdır. Diğer bulanık sistemlerde olduğu gibi, giriş ve çıkış üyelik fonksiyonları oluşturuldu. Giriş üyelik fonksiyonu, Şekil 4.7.a' da gösterilmiştir. Giriş değişkeni, *ARB.stim*, *0* ila *1* arasında değişir. Bir üyelik değeri, giriş üyelik fonksiyonu kullanarak elde edilen değere göre hesaplanır. Kural tabanını oluşturmak için kullanılan hesaplamada, giriş değeriyle (*ARB.stim*) üçgen üyelik fonksiyonun kestiği iki nokta elde edilir. Ayrıca bu noktalar, ilgili üyelik fonksiyonu için giriş değişkenin üyelik değerleri olarak adlandırılır. Bu noktaların minimumu, *x* giriş değişkenin üyelik değeri olarak alınır ve bu durum eşitlik 4.15' de gösterilmiştir:

$$\mu_{A\cap B}(x) = MIN(\mu_A(x), \mu_B(x)), x \in x \qquad (4.15)$$

Burada $\mu_A(x)$, A' da x' in üyelik değeridir ve $\mu_B(x)$, B'de x' in üyelik değeridir, A ve B, x uzayında iki kümedir. Hesaplanan giriş üyelik değeri, Şekil 4.7.a' da gösterilen çıkış üyelik fonksiyonuyla çıkış değeri elde etmek için kullanılır.

Şekil 4.7.b' nin x ekseninde, *0* ila *10* arasında değişen ve *ARB* için üyelik fonksiyonları kullanılarak hesaplanacak tahsis edilen kaynak sayısı gösterilir. Yukarıda açıklandığı gibi elde edilen giriş üyelik değeri olan y eksenindeki ağırlık, birkaç noktada üyelik üçgenlerinin kesişimidir. Bulanık kaynak dağılımı için kural tabanı, Tablo 4.2' de gösterilmiştir.



Şekil 4.7 (a) Giriş Üyelik Fonksiyonu, (b) Çıkış Üyelik Fonksiyonu

Burada, *ÇK*, *K*, *OK* gibi değerler, giriş üyelik üçgenlerinin etiketleridir ve *ÇK'*, *K'*, *OK'* gibi değerler de çıkış üyelik değerlerinin dilsel etiketleridir. Tablo 4.2' deki kurallar, hangi noktaların ortalaması alınacağını tanımlar. Örnek olarak, giriş değeri, giriş üyelik fonksiyonları arasında *ÇK* ve *K* üçgenlerini keserse, ortalaması alınacak noktalar, sadece çıkış üyelik fonksiyonlarında *ÇK'* ve *K'* üçgenleri olacaktır.

Tablo 4.3' te giriş ve çıkış üyelik fonksiyonları için dilsel değerler verilmiştir. Bu dilsel değerler, *0* ve *0.50* arasında uyarılma değerine sahip *ARB*' ler için tahsis

edilecek kaynak sayısı daha az olurken, *0.50* ve *1.00* arasında uyarılma seviyesine sahip *ARB*' ler için daha fazla olunacağı anlamına gelen bir durumda belirlenir. *AIRS*' deki kaynak dağılımı ile bulanık kaynak dağılımı arasındaki kaynak dağılımı farkını göstermek için aşağıda örnek bir uygulama verilmiştir. Uygulamada klonal oran *10* olarak alınmıştır. Bu örnekte, üç adet *ARB* kullanılmış ve uyarılma seviyeleri verilmiştir. *AIRS*' deki kaynak sayısı ile bulanık kaynak tahsisinden elde edilen kaynak sayıları ayrı ayrı sütunlara yazılmıştır. Bu örnekten görüleceği gibi, *AIRS*' de kullanılan normal kaynak tahsisi ile ARB' lere tahsis edilen toplam kaynak sayısı *13* iken, bulanık kaynak tahsisi ile elde edilen toplam kaynak sayısı ise *11. 1375*' tir. Bu durum Tablo 4.4' de verilmiştir (Polat ve ark. 2006b).

Tablo 4.2 Bulanık kaynak dağılımı için kural tabanı

Kurallar
Eğer ARB.stim=ÇK ve K ise Çıkış=(ÇK'+K')/2
Eğer ARB.stim=K ve OK ise Çıkış=(K'+OK')/2
Eğer ARB.stim=OK ve AK ise Çıkış=(OK'+AK')/2
Eğer ARB.stim=AK ve AB ise Çıkış=(AK'+AB')/2
Eğer ARB.stim=AB ve OB ise Çıkış=(AB'+OB')/2
Eğer ARB.stim=OB ve B ise Çıkış=(OB'+B')/2
Eğer ARB.stim=B ve ÇB ise Çıkış=(B'+ÇB')/2

Tablo 4.3 Giriş ve çıkış üyelik fonksiyonları için dilsel değerler

Giriş	**Çıkış**
ÇK- Çok Küçük	**ÇK'**- Çok Küçük
K- Küçük	**K'**- Küçük
OK- Orta Küçük	**OK'**- Orta Küçük
AK- Az Küçük	**AK'**- Az Küçük
AZ- Az Büyük	**AZ'**- Az Büyük
OB- Orta Büyük	**OB'**- Orta Büyük
B- Büyük	**B'**- Büyük
ÇK- Çok Büyük	

Tablo 4.4 Fuzzy kaynak dağılımı için örnek bir uygulama

ARB	**ARB uyarılma seviyeleri**	**AIRS' deki kaynak denklemi (ARB.stim*clonal rate) clonal rate=10**	**AIRS' deki kaynak sayısı**	**Fuzzy-AIRS' deki kaynak sayısı**
1. ARB	0.2	0.2*10	2	1.25
2. ARB	0.4	0.4*10	4	3.1875
3. ARB	0.7	0.7*10	7	6.7
Toplam	1.3	13	13	11.1375

4.5. C4.5 Karar Ağaçları

Karar ağaçları tümevarımsal sonuç çıkarmaya dayanan genel bir sınıflama yöntemidir. Ayrık değerli fonksiyonları yakınsayan, gürültülü verilere karşı dayanıklı ve ayıran ifadeleri öğrenme yeteneği olan bir yöntemdir. Tamamen ifade edilmiş hipotez uzayında arama yapar. Karar ağacı oluşturmada, genel olarak küçük ağaçlar büyüklere tercih edilir. Öğrenilen fonksiyon karar ağacı şeklinde gösterilir (Mitchell 1997).

Karar ağacındaki her bir düğüm, eğitim örneğine ait bir niteliğin (özelliğin) test edilmesini sağlar ve bu düğümden çıkan her dal özeliğin bir değerine karşılık gelir. Bir örnek, ağacın kök düğümünden başlayarak sınıflandırılmaya başlanır ve her düğümde ilgili özellik için test edilir (Mitchell 1997).

Karar ağaçları, kesişimlerin birleşimi olarak düşünülebilir. Ağacın kökünden bir yaprağa uzanan yol, özellik testlerinin kesişimi ve ağacın kendisi bu kesişimlerin birleşimidir. Karar ağacı algoritması, ağacı oluşturmak için eğitim verileri kümesini kullanır (Mitchell 1997, Quilan 1986).

Karar ağaçlarında, eğitim örneklerine uyan bir hipotez uzayı aranır. Basitten karmaşığa doğru tepe tırmanma araması yapılır. Boş ağaçtan başlanır ve eğitim verisini doğru biçimde sınıflandıracak karar ağacının bulunması için ağaç oluşturur. Aramada fonksiyon olarak *Bilgi Kazancı* (*Information Gain*) kullanılır. Hipotez uzayı, eldeki verilere bağlı olarak ayrık değerli fonksiyonların eksiksiz uzayıdır. Her fonksiyon, ayrı bir karar ağacı ile gösterildiği için, eksik hipotez uzayı kullanma

riskinden kurtulunmuş olur. Hipotez uzayı hedef fonksiyonu kesin olarak içerir (Mitchell 1997).

Eğitim verilerinde gürültü varsa veya eldeki veri çok az ise, karar ağaçlarında eğitim verilerini aşırı öğrenme (overfitting) durumu ortaya çıkabilir. Bu sorunu önlemek için ağacın büyümesi önceden durdurulabilir, fakat buna doğru karar vermek zordur. Pratikte daha kullanışlı diğer bir yaklaşım ise, aşırı öğrenme ortaya çıkacak şekilde ağacın büyümesine izin verip daha sonra budama yapılır (Kodaz 2007, Quinlan 1993, Mitchell 1997, Akdemir ve ark. 2007). C4.5 karar ağaçları ile sürekli değerli özelliklerle çalışabilme ve eğitme işlemi sırasında aşırı eğitme problemi ortadan kalkmıştır (Quinlan 1993, Kodaz 2007).

5. KULLANILAN BİYOMEDİKAL VERİ KÜMELERİ

Tezde kullanılan biyomedikal veri kümeleri olarak özelliği önceden belirlenmiş ve özelliği çıkarılmış olmak üzere veri kümeleri iki şekilde incelendi. Özelliği belirlenmiş medikal veri kümelerinde özellikler, birbirinden bağımsız olan özelliklerdir. Bu özellikler, hastanın laboratuvar sonuçları, cinsiyeti, yaşı, ağırlığı gibi özelliklerden oluşmaktadır. Özelliği çıkarılmış medikal veri kümelerinde ise özellikler, hastadan veya sağlıklı kişiden alınan bir biyopotansiyel sinyalin (EEG, Doppler, ECG, PERG, VEP gibi) spektral analiz algoritmaları ile işlenip biyopotansiyel sinyalden özellik çıkartmaya dayanır.

Veri kümelerinde; hastalığın temeli, veritabanı bilgisi ve makine öğrenmesi teknikleri, sinir ağları ve sinyal işleme tekniklerindeki geçmiş tahmin performansını içerir. Hastalığın temeli, hastalık hakkındaki biyolojik ve medikal bilgiye önem verir. Hastalığın özelliğini tespit etmede, en iyi öğrenen sistemi bulmak veya oluşturmak için hastalığın kökenini bilmek gerekir. Veritabanı bilgisi, veritabanını büyüklüğü ve eksik özelliğe sahip örneklerin veritabanında olup olmadığını içerir. Sonra, veritabanı özellikleri, eğer gerekliyse kendi ölçtüğü birimi ve kendine uygun değerlerden oluşan listeden oluşur.

5.1. Özelliği Belirlenmiş Medikal Veri Kümeleri

Kullanılan medikal veri kümeleri; kalp hastalığı veri kümesi, SPECT görüntüleri ile kalp hastalığı veri kümesi ve E.coli promoter gen dizileri tahmini veri kümesidir. Bu medikal veritabanları, Irvine' deki California Üniversitesi (UCI), Bilgisayar Bilimi ve Bilgi Bölümünden, makine öğrenmesi repertuarından elde edildi.

5.1.1. Kalp hastalığı veri kümesi

Kalp hastalığı, kalbin kan odacıklarını veya kalp kasını etkiler. Kalp rahatsızlığı, genellikle, kan odacık problemidir, örnek olarak kan odacığının darlaşması veya kan akışının blokajı gibi problemlerdir (Yan 2006).

Kalp hastalığının belirtileri, göğüs ağrısı, aşırı yorgunluk, yüksek kan basıncı ve kan şekeridir. İlave olarak, kalp hastalığının birkaç biçimi vardır: kalp atağı, kalp ağrısı ve kalp tümörü v.b (Yan 2006).

Kalp hastalığı veritabanı 270 örnekten oluşmaktadır. Orijinal olarak, veritabanı 76 ham özellikten oluşmaktadır. Bununla birlikte, bütün yayınlanmış deneylerde sadece bu özelliklerden 13 özellik kullanılmıştır. Kullanılan özellikler Tablo 5.1' de listelenmiştir. Bu özellikler ayrık değerli yani evet ve hayır şeklinde olan ve sürekli değerlerden oluşmaktadır. Kalp hastalığı veritabanını iki değişken sınıfı vardır: hastalık yok (% 55.56) ve hastalık var (%44.44) (UCI veritabanı 2008).

Bu veritabanındaki 270 örnekten 150 tanesi "hastalık yok" sınıfına, 120 tanesi de "hastalık var" sınıfına aittir. Kalp hastalığı veritabanında her bir durum için çıkış değerlerinin birkaçı Tablo 5.2' de verilmiştir. Tablo 5.3' de ise kalp hastalığı veri kümesindeki özelliklerin istatistiki değerleri verilmiştir.

Tablo 5.1 Kalp hastalığı veri kümesindeki özellikleri dağılımı (UCI veritabanı 2008)

• Yaş – 20' den 77' ye yaş arası
• Cinsiyet – erkek veya kız, sırayla 0 ve 1 olarak gösterilir.
• Göğüs ağrısı tipi- göğüs ağrısının dört tipi
o Değer 1: angina tipi
o Değer 2: atypical angina
o Değer 3: non-anginal ağrısı
o Değer 4: asymptomatic
• Geriye kalan kan basıncı- mm Hg' de ölçülmüş.
• Serum kolesterolü- mg/dl' de ölçülmüş.
• Hızla artan kan şekeri > 120 mg/dl- doğru:1 veya yanlış:0;
• Geriye kalan elektrokardiyografik sonuçları
o Değer 0: Normal
o Değer 1: Anormal olan ST-T dalga oluşması

o Değer 2: Sol ventrikular hypertropy olasığının gösterilmesi
• Maksimum kalp oranının elde edilmesi
• Angina' ya neden olan uygulama
o Değer 1: Evet;
o Değer 0: Hayır;
• Geriye kalan uygulama tarafından oluşan ST depresyon.
• Zirve (doruk) ST uygulama parçasının eğimi
o Değer 1: Eğimin artması
o Değer 2: Düz
o Değer 3: Eğimin azalması
• Fluoroscopy tarafından renklenen odacıkların sayısı – 0' dan 3' e değerlidir.
• Thal – Thalasemias
o Değer 3: Normal
o Değer 6: Değişmez hasar
o Değer 7: Eski hasar

Tablo 5.2 Kalp hastalığı veritabanında her bir durum için çıkış değerlerine bağlı özellikler

Durum	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	**Teşhis**
1	44	1	3	120	226	0	0	169	0	0	1	0	3	Hastalık yok
2	44	1	3	140	235	0	2	180	0	0	1	0	3	Hastalık yok
...	.	.	.	.	.	.	.	.	.	.	.	.	.	...
...	.	.	.	.	.	.	.	.	.	.	.	.	.	...
270	63	0	4	108	269	0	0	169	1	1.8	2	2	3	Hastalık var

Tablo 5.3 Kalp hastalığı veri kümesinde bulunan özelliklerin istatistiki değerleri

Özelllik Numarası	**Özelliğin Adı**	**Minimum Değeri**	**Maksimum Değeri**	**Ortalama Değeri**	**Standart Sapması**
1	Yaş	29	77	54.433	9.0922
2	Cinsiyet	0	1	0.67778	0.46733
3	Göğüs ağrısı tipi	1	4	3.1741	0.94833
4	Geriye kalan kan basıncı	94	200	131.34	17.829
5	Serum kolesterolü	126	564	249.66	51.59
6	Hızla artan kan şekeri	0	1	0.14815	0.35525
7	Geriye kalan elektrokardiyografik sonuçları	0	2	1.0222	0.99604
8	Maksimum kalp oranının elde edilmesi	71	202	149.68	23.123
9	Angina' ya neden olan uygulama	0	1	0.32963	0.47008
10	Geriye kalan uygulama tarafından oluşan ST depresyon	0	6.2	1.05	1.1431
11	Zirve (doruk) ST uygulama parçasının eğimi	1	3	1.5852	0.61325
12	Fluoroscopy tarafından renklenen odacıkların sayısı	0	3	0.67037	0.94215
13	Thal – Thalasemias	3	7	4.6963	1.9371

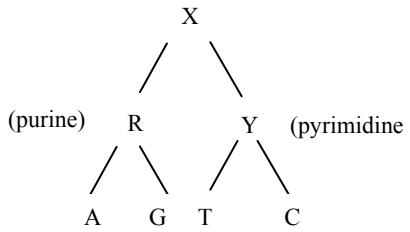
5.1.2. E. Coli promoter gen dizileri veri kümesi

Promoterlar, *RNA* polimerleriyle etkileşim vasıtasıyla transkripsiyon başlangıcının yerini ve frekansını etkileyen *DNA* dizileridir. Transkripsiyon başlangıcından *35* ve *10* taban çiftleri (*base pair-bp*) yukarısında olanlar hakkındaki iki korunmuş bölge, ilgili birkaç promoterlerin karşılaştırılmasıyla tanımlanır. E. Coli genleri için promoterların karşılaştırılması ve daha geniş derlemeleri, kendi bakteriyofaj ve plazmasını destekler ve sonuç promoter dizisinin içeriğini genişletir: *a-35* (*TTGACA*) ve *-10* (*TATAAT*) bölgeleri, *-10* bölgesinin *3* sonundan yaklaşık *7 bp* aşağıda olan purinde başlayan transkripsiyonla *17 bp* ile ayrılırlar. *-35* ve *-10* bölgeleri, promoterlara karşı en büyük korumayı gösterirken ve aynı zamanda transkripsiyon gücünü etkileyen yakın olarak bütün mutasyonal bölgelerken, başlangıç noktasına ilaveten *-35* ve *-10* bölgelerinin yanındaki diğer tabanlar rastgele frekanslardan daha fazla meydana gelirler ve bazen promoter aktivitesini etkilerler. İlave olarak, *-35* ve *-10* bölgeleri arasındaki boşluğun değişimi, promoter gücünde bir rol oynar (Harley ve ark.1987, Geoffrey ve ark.1990).

E.coli promoter gen dizileri veri kümesi, 57 özellik ve 53'ü promoter ve 53' ü promoter olmayan toplam 106 örnekten oluşur. Giriş özellikleri 57 dizili DNA nükleoitleridir. Özel bir notasyon, DNA dizisindeki özelleşmiş yerleri basitleştirmek için kullanılır. Biyolojik literatür, transkripsiyonun başladığı bölgeyle ilgili yerleri hesaba katar. 50 nükleoitten önce ve takip eden 6 bir örneği oluşturur. Bir kuralın öncekileri, giriş özelliklerini tercih ederken, onlar ilk olarak başlangıç yerini söyler ve sonra takip edilmesi gereken diziyi listelerler. Bu açıklamalarda, "*x*", herhangi bir nükleoitin yeterli olacağını gösterir. Bu yüzden, doğrulama için ilk kural, transkripsiyonun başladığı bölgeden önce *"a" 45* nükleotin olması gerektiğini söyler. Diğer bir *"a" -44* olmalıdır ve sonra herhangi iki nükleoit görülebilir(Geoffrey ve ark. 1990, UCI veritabanı 2008). Şekil 5.1 DNA dizilerini gösterir. E. Coli promoter gen dizileri veri kümesinde her bir durum için çıkış değerlerinin birkaçı Tablo 5.5' de verilmiştir. DNA dizilerindeki nükloitlerin görülme sıklıkları Tablo 5.6'da görülmektedir. Tablo 5.7' de ise E. Coli promoter gen dizileri veri kümesindeki özelliklerin istatistiki değerleri verilmiştir.

Table 5.4 Promoterlar için bir alan teorisi (UCI veritabanı 2008)

promoter	-contact (kontak), conformation (doğrulama).
contact	- minus_35, minus_10.
minus_35	-@_37 “cttgac”.
minus_35	-@_36 “ttgxca”.
minus_35	-@_36 “ttgaca”.
minus_35	-@_36 “ttgac”.
minus_10	-@_14 “tataat”.
minus_10	-@_13 “taxaxt”.
minus_10	-@_13 “tataat”.
minus_10	-@_12 “taxxxt”.
conformation	-@_45 “aaxxa”.
conformation	-@_45 “axxxa”, @_4 “t”, @_28 “txxxtxaaxxtx”.
conformation	@_49 “axxxxt”, @_1 “a”, @_27 “txxxxaxxtxtg”.
conformation	@_47 “caaxttxac”, @_22 “gxxxtxc”, @_8 “gcgccxcc”.



Şekil 5.1 DNA nükleoitleri (UCI veritabanı 2008)

Tablo 5.5 E. Coli promoter gen dizileri veritabanında her bir durum için çıkış değerlerine bağlı özellikler

Durum	x_1	x_2	x_3	x_4	x_5	**…**	**…**	**…**	x_{53}	x_{54}	x_{55}	x_{56}	x_{57}	**Teşhis**
1	2	0	3	2	0	…	…	…	1	2	3	1	2	Promoter
2	2	1	3	2	0	…	…	…	1	3	3	0	0	Promoter
...						…	…	…						...
...						…	…	…						...
106	0	2	0	2	1	…	…	…	1	2	3	2	0	Promoter Olmayan

Tablo 5.6 DNA dizilerindeki nükleoitlerin görülme sıklıkları

Nükloit	**Promoter**	**Promoter olmayan**
A :	%27.7	%24.4
G :	%20.0	%25.4
T :	%30.2	%26.5
C :	%22.1	%23.7

Tablo 5.7 E. Coli promoter gen dizileri veri kümesinde bulunan özelliklerin istatistiki değerleri

Özellik Numarası	Minimum Değeri	Maksimum Değeri	Ortalama Değeri	Standart Sapması
1	0	3	1.6226	1.1113
2	0	3	1.3396	1.1318
3	0	3	1.3679	1.0931
4	0	3	1.6038	1.1047
5	0	3	1.2264	1.1013
6	0	3	1.1698	1.1283
7	0	3	1.3208	1.1537
8	0	3	1.3585	1.1175
9	0	3	1.4057	1.1309
10	0	3	1.3396	1.1483
11	0	3	1.3491	1.1659
12	0	3	1.6604	1.0979
13	0	3	1.3113	1.0128
14	0	3	1.6887	1.1765
15	0	3	1.5377	0.96323
16	0	3	1.8019	0.96563
17	0	3	1.3491	0.87999
18	0	3	1.3396	1.2354
19	0	3	1.7925	1.2264
20	0	3	1.3868	1.1291
21	0	3	1.4717	1.0921
22	0	3	1.3302	1.0879
23	0	3	1.5566	1.0823
24	0	3	1.4717	1.048
25	0	3	1.4245	1.0898
26	0	3	1.283	0.99768
27	0	3	1.434	1.1161
28	0	3	1.7075	1.0458
29	0	3	1.5377	0.98263
30	0	3	1.3396	1.1727
31	0	3	1.2642	1.1015
32	0	3	1.3585	1.109
33	0	3	1.6132	1.069
34	0	3	1.6038	1.0064
35	0	3	1.5094	1.1264
36	0	3	1.566	1.0553
37	0	3	1.5377	0.97298
38	0	3	1.4434	1.0735
39	0	3	1.3868	1.1207
40	0	3	1.0755	1.0789
41	0	3	1.2453	1.1559
42	0	3	1.6415	1.0478
43	0	3	1.5094	0.99285
44	0	3	1.7358	1.0755
45	0	3	1.2925	1.0091
46	0	3	1.6321	1.1682
47	0	3	1.5566	1.1581
48	0	3	1.6226	1.1447
49	0	3	1.6604	1.1565
50	0	3	1.7736	1.1918
51	0	3	1.5377	1.167
52	0	3	1.6604	1.1565

53	0	3	1.4906	1.0573
54	0	3	1.6792	1.1121
55	0	3	1.5189	1.1346
56	0	3	1.5943	1.114
57	0	3	1.3868	1.0331

5.1.3. SPECT (Single Photon Emission Computed Tomography) görütüleri kalp hastalığı teşhisi veri kümesi

SPECT görüntüleme, miyokardiyal perfüzyon için bir teşhis aracı olarak kullanılır. Hastaya radyoaktif bir izleyici enjekte edilir. Sonra, iki çalışma gerçekleşir, bunlardan biri maksimum stres esnasında enjeksiyondan 10-15 dakika sonra gerçekleşir ve bir diğeri rahatlama çalışması esnasında enjeksiyondan 2-5 saat sonra arasında gerçekleşir. Çalışmalar, üç boyutlu görüntülerin iki kümesi olarak toplanır. Bütün görüntüler, miyokardiyum içinde radyoaktif maddelerin dağılımıyla orantılı olan *LV* (*Left Ventricular-sol karıncık*) kas perfüzyünonu (büzülme) gösterir. Lukasz ve arkadaşları, *LV* perfüzyonunda anormal durumları teşhis etmek için stres ve rahat çalışmaları karşılaştırmışlardır (Lukasz ve ark. 2001).

Kardiak *SPECT* görüntüleri için diğer görselleştirme yöntemleri vardır. Bunlardan birisi, 3 boyutlu bir görüntüyü, küresel ve silindirik koordinatların birleşiminde veya küresel koordinatlarında radyal izdüşümü ile 2 boyutlu yüzeye izdüşümüne dayanan bull's eye (Boğanın gözü) metotudur. Diğer yöntem aileleri, *LV*' nin 3 boyutlu yüzey dönüşümüyle bağlatılıdır. Kalp kasının hareketini görselleştirmek için kan havuzu *SPECT* görüntülerinin kullanırlar. Kalp ile ilgili dinamik görüntüleri yorumlamayla ilgili çalışmalar da vardır. Genel olarak kalp ile ilgili hareket analizi, koroner arter çevresi bozuklukları veya miyokardinal anormalliklerle ilgili patolojileri tanımlamaya izin verir. Bu tez çalışmasında kullanılan *SPECT* görüntüleri veri kümesi UCI (University of California, Irvine) makine öğrenmesi veritabanından alınmıştır (UCI veritabanı 2008).

Bölgesel teşhis için yedi sınıf (Normal, Reversible (Terslenebilir), Partially Reversible (Kısmi olarak terslenebilir), Artifact (İnsan veya dış etki ile oluşturulan bozukluklar), Fixed (Sabit), Equivocal (Şüpheli) ve Reverse Redistribution (Ters tekrar dağılım) tanımlanır ve sonra sekiz sınıf (Normal, Ischemia, Infarct, Infarct and

Ischemia, Artifact, Equivocal, Reverse Redistribution ve *LV* Dysfunction) tam teşhis için belirlenir. *SPECT* görüntüleri veri kümesinde, her bir hasta iki kategoride sınıflandırılır: normal ve anormal. 267 *SPECT* görüntü kümesi veritabanı, orijinal *SPECT* görüntüleri özellikleri belirlemek için işlenir. Sonuç olarak, 44 sürekli özellik örüntüleri, her bir hasta için oluşturuldu. Örüntü, 22 ikili özellik örüntüleri elde etmek için tekrar işlendi (Lukasz ve ark. 2001, Bakırcı ve ark. 2004, Kurgan ve ark. 2002). Kalp ile ilgili *SPECT* görüntüleri veri kümesinde 55 normal ve 212 anormal örüntüler vardır. Bu veri kümesi dengesiz bir veri dağılımına sahiptir. *SPECT* görüntüleri veri kümesinde her bir durum için çıkış değerlerinin birkaçı Tablo 5.8' de verilmiştir. Tablo 5.9' de ise *SPECT* görüntüleri veri kümesindeki özelliklerin istatistiki değerleri verilmiştir (UCI veritabanı 2008, Polat 2007b).

Tablo 5.8 Kalp ile ilgili SPECT görüntüleri veritabanında her bir durum için çıkış değerlerine bağlı özellikler

Durum	x_1	x_2	x_3	x_4	x_5	…	…	…	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	**Teşhis**
1	0	0	0	0	0	…	…	…	0	0	0	0	1	Normal
2	0	0	0	0	0	…	…	…	0	0	0	0	0	Normal
...						…	…	…						...
...						…	…	…						...
267	1	0	0	0	1	…	…	…	0	0	1	1	0	Anormal

Tablo 5.9 Kalp ile ilgili SPECT görüntüleri veri kümesinde bulunan özelliklerin istatistiki değerleri

Özellik Numarası	**Minimum Değeri**	**Maksimum Değeri**	**Ortalama Değeri**	**Standart Sapması**
1	0	1	0.44569	0.49704
2	0	1	0.24719	0.43138
3	0	1	0.39326	0.48847
4	0	1	0.28464	0.45124
5	0	1	0.40449	0.49079
6	0	1	0.23596	0.42459
7	0	1	0.28464	0.45124
8	0	1	0.42697	0.49464
9	0	1	0.31086	0.46285
10	0	1	0.37828	0.48496
11	0	1	0.24345	0.42916
12	0	1	0.29588	0.45644
13	0	1	0.49438	0.49997
14	0	1	0.30337	0.45971
15	0	1	0.17603	0.38085
16	0	1	0.31086	0.46285
17	0	1	0.14232	0.34938
18	0	1	0.13109	0.33749
19	0	1	0.24719	0.43138
20	0	1	0.3221	0.46728
21	0	1	0.3633	0.48095
22	0	1	0.41199	0.49219

5.2. Özelliği Çıkarılmış Medikal Veri Kümeleri

Özelliği çıkarılmış medikal veri kümeleri, Doppler sinyali ile damar sertliği veri kümesi, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile Macular hastalığı veri kümeleridir. Bu veri kümeleri içinde, spektral analiz algoritmaları ile özelliği çıkarılmış veri kümesi Doppler sinyali ile damar sertliği veri kümesidir. Diğer veri kümelerinde özellik değerleri ise, hastadan veya sağlıklı kişiden alınan zaman eksenli PERG ve VEP sinyallerden alınan genlik değerleridir. Bu tez çalışmasında kullanılan özelliği çıkarılmış olan medikal veri kümeleri Fatih Üniversitesi, Müh. Fak., Elektrik-Elektronik Mühendisliği Bölümü Öğretim Üyesi Prof. Dr. Sadık Kara ve ekibinden alınmıştır.

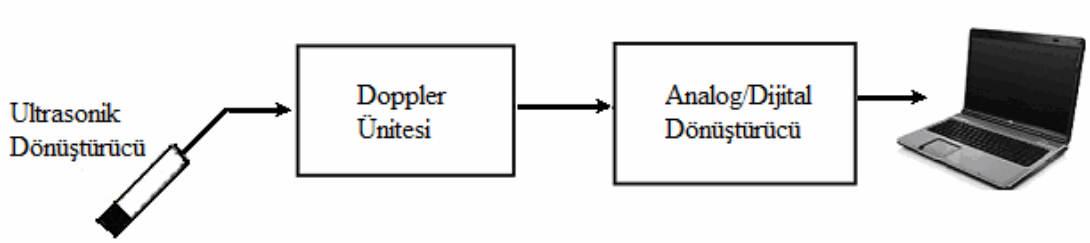
5.2.1. Doppler sinyali ile damar sertliği veri kümesi uygulamaları

Damar sertliği (Atheroclerosis) yaygın bir atardamar hastalığıdır ve hücre dışındaki lipidler, lifli doku, düz kas hücreleri ve kalsiyum toplanması ile ortaya çıkmaktadır (Dirgenali 2003, Hirai ve ark. 1989).

Doppler sonogramları bir zaman serisinin gücünü açıklamakta ve böylece Doppler sinyalinin spektral analizi atardamar içerisindeki hız dağılımı hakkında bilgi sağlamaktadır (Dirgenali 2003, Evans ve ark. 1989, Vaitkus ve ark.1988, Evans 2000).

Damar sertliği veri kümesi 114 veriden oluşmaktadır. 114 verinin 60'ı hasta, 54'ü sağlıklı kişilerden oluşmaktadır. Veriler Prof. Dr. Sadık Kara ve Dr. Fatma Latioğlu tarafından Erciyes Üniversitesi Tıp Fak. Radyoloji bölümünde Toshiba PowerVision 6000 Doppler Ultrason ünitesi kullanılarak elde edilmiştir. 8 MHz lineer ultrason sondası sol karotis atardamara darbeli ultrason sinyali iletmek için kullanılmıştır. Karotis atardamarlarda Doppler tepkilerini elde etmek için ultrason cihazının ses çıkış portu kullanılmıştır. Ultrason cihazının çıkışındaki ses sinyalleri 44100 Hz'de örneklenmiştir ve bir giriş/çıkış kartı aracılığı ile bilgisayara

gönderilmiştir (Dirgenali 2003, Kara 1995). Elde edilen veri bir ADC (Analog to Digital Converter) kullanılarak sayısal hale getirilmiş ve bir giriş/çıkış kartı ile bilgisayara aktarılmıştır. Doppler verisine güç spektrum tahmini için Welch metodu uygulanmıştır (Kara 2006, Özşen ve ark. 2007, Latifoglu 2007a, Latifoglu 2007b). Şekil 5.2, Doppler verisi elde etmek için kullanılan donanımın blok diyagramını göstermektedir (Kara 2006).



Şekil 5.2 Doppler verisi elde etmek için kullanılan donanımın blok diyagramı (Dirgenali 2003)

Zaman ekseninde karotis atardamardan yansıyan Doppler sinyalleri hastalığın varlığı hakkında net bir bilgi verememektedir. Bu sebepten bu sinyaller sağlıklı kişi ile damar sertliği olan kişi arasındaki farkları ortaya çıkarmak için frekans ekseninde analiz edilmektedir. Bu analiz sonucunda bir periyot için sonogram 61 özellikle ifade edilmektedir. Tablo 5.10' da Doppler sinyali ile damar sertliği veri kümesinde her bir durum için çıkış değerlerine bağlı özellikler verilmiştir. Tablo 5.11' de ise Doppler sinyali ile damar sertliği veri kümesindeki özelliklerin istatistiki değerleri verilmiştir.

Tablo 5.10 Doppler sinyali ile damar sertliği veri kümesinde her bir durum için çıkış değerlerine bağlı özellikler

Durum	x_1	x_2	x_3	x_4	x_5	…	…	…	x_{57}	x_{58}	x_{59}	x_{60}	x_{61}	**Teşhis**
1	8.4	8.8	9	8.8	9.6	…	…	…	7.2	8	8.6	9.2	9.8	Sağlıklı
2	8.6	9	9.4	9.6	10	…	…	…	7.2	8	8.6	9	9.6	Sağlıklı
...						…	…	…						...
...						…	…	…						...
114	6.2	6.2	6.2	6.2	6.4	…	…	…	4	4.6	4.8	5.2	5.4	Hasta

Tablo 5.11 Doppler sinyali ile damar sertliği veri kümesinde bulunan özelliklerin istatistiki değerleri

Özellik Numarası	Minimum Değeri	Maksimum Değeri	Ortalama Değeri	Standart Sapması
1	2.2	14.6	6.186	2.9268
2	2	12.4	6.076	2.891
3	2	12.4	6.1579	3.1377
4	2	14.4	6.6825	3.8223
5	2	17.4	7.6544	4.752
6	2	23.2	8.9088	5.9601
7	3.2	26.4	11.014	7.1908
8	2	32.6	13.386	8.4182
9	4.2	33.6	15.726	9.1075
10	5.4	33.6	17.853	9.5438
11	0.6	34.4	19.858	10.308
12	2.6	34.2	21.407	9.9058
13	3.8	34.4	22.411	9.5953
14	12.2	35.4	23.446	8.9424
15	13	35.4	23.756	8.7901
16	12.4	34.8	23.37	8.7959
17	11.6	34.4	22.751	8.7659
18	11	33.6	21.902	8.6337
19	10.2	33	20.825	8.3927
20	9.2	33	19.484	8.0474
21	8.4	32.8	18.23	7.5631
22	7.4	32.4	16.93	6.9479
23	7	30.6	15.633	6.3621
24	6.2	28	14.421	5.6959
25	6	25.2	13.479	5.056
26	5.6	23	12.667	4.4836
27	5.6	20.4	12.032	4.0545
28	5.6	18.6	11.609	3.7561
29	5.2	17.8	11.323	3.6099
30	5.2	17.6	11.198	3.5655
31	5.2	17.8	11.163	3.6605
32	4.8	18.2	11.172	3.8013
33	4.8	19.2	11.258	4.0695
34	4.6	19.4	11.332	4.3075
35	4.4	19.6	11.309	4.635
36	4.4	19.4	11.212	4.8994
37	3.8	19.6	11.098	5.2293
38	3.8	20	10.83	5.4307
39	4	20	10.509	5.6197
40	3.6	19.4	10.174	5.689
41	3.4	19	9.8421	5.6619
42	3.2	18.6	9.5158	5.5571
43	3	18.4	9.2579	5.3654
44	3	17.8	9.0596	5.0286
45	3.2	17	8.8316	4.6647
46	3.4	16.4	8.5772	4.2555
47	3.4	16.2	8.2719	3.8244
48	3.4	16.4	7.8754	3.4624
49	3.2	16	7.386	3.2172
50	3.4	16	6.9333	3.0734
51	3.4	15.6	6.5544	3.1098
52	3.4	15.6	6.3193	3.1727

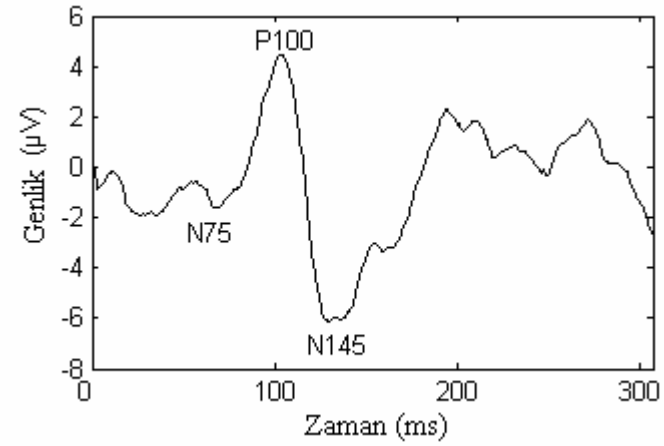
53	3.2	16.8	6.2386	3.2616
54	3.2	17	6.307	3.4685
55	3	17.6	6.507	3.6908
56	2.8	18.4	6.8298	3.8832
57	3.2	18.2	7.1368	4.1039
58	3	18	7.4825	4.3031
59	3	18.4	7.8614	4.4873
60	2.8	18.6	8.1386	4.6768
61	3	18.2	8.3456	4.8173

5.2.2. VEP (Visual Evoked Potential- Görsel Uyarılmış Potansiyel) sinyali ile optik sinir hastalığı veri kümesi

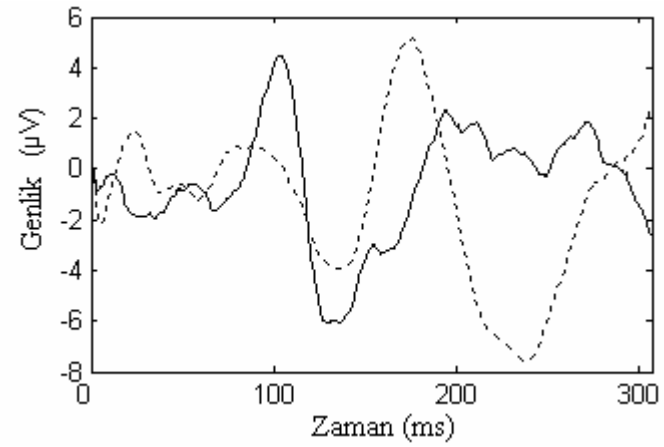
Optik sinir, göz ve beyin arasında önemli bir bağlantıdır. Eğer optik sinir, hastalıktan ciddi şekilde etkilenirse veya sarsıntı veya tümör vasıtasıyla hasar görürse, görme kaybı veya körlük meydana gelebilir (Heckenlively ve ark. 1991, Tasman ve ark. 1992, Stoll ve ark. 1988, Ruther ve ark. 1998, Chamlin ve ark. 1955). Optik sinir hastalığından Macular veya retinanın farklı teşhisinin oluşturulmasında birkaç farklı prosedür kullanılabilir. Bu prosedürlerden birkaçının gerçekleştirilmesi kolaydır ve Amsler grid, color vision testing (renkli gösterim test etme), pupillary reflexes, light-brightness comparison (açık renklilik karşılaştırma) ve Macular dazzle (Macular göz kamaşması) içerir. Diğer prosedürler çok fazla bilgiye gerek duyarlar ve fluorescein anjiyografi, Visual Evoked Potential ve Pattern Electroretinogram (PERG) içerirler (Vernon Odom ve ark. 2004). Görsel elektrofizyoloji testleri (PERG, Electroretinogram-ERG, Electrooculogram-EOG ve VEP içeren), retina ve optik sinirin nasıl çalıştığını anlatırlar (Güven ve ark. 2008).

VEP, optik sinir patolojilerinin hassas bir ölçümü olarak tanımlanır (Atilla ve ark. 2006). Kompleks bir dalga, görsel uyarımdan sonra tahmin edilebilir gecikme zamanlarında meydana gelen ayırt edilebilir pozitif ve negatif tepelerle üretilir (Tasman ve ark. 1992). VEP, görsel sistemin fonksiyonel birleşimi ile ilgili önemli teşhis bilgisi sağlayabilir ve klinik ve araştırma laboratuvarlarında kullanılır. VEP, optik sinir veya görsel kabuk hastalıklarının yanı sıra kalıtsal Macular bozulması gibi dış retina hastalıklarında anormal olabilir (Jancic-Stefanovic ve ark. 2007, William ve ark. 1993, Güven ve ark. 2008).

Genellikle, VEP' in klinik kullanımı, N75, P100 ve N145' in gecikmelerine ve tepe genliğine dayanır (Tasman ve ark. 1992). Şekil 5.3, bileşenlerle etiketlenmiş normal VEP sinyalini gösterir. Şekil 5.4 ise normal ve hastalıklı iki gözün VEP sinyallerini gösterir (Güven ve ark. 2008).



Şekil 5.3 Bileşenlerle etiketlenmiş normal VEP sinyali, N75, P100, N145 ise gecikmelerdir (Güven 2005)

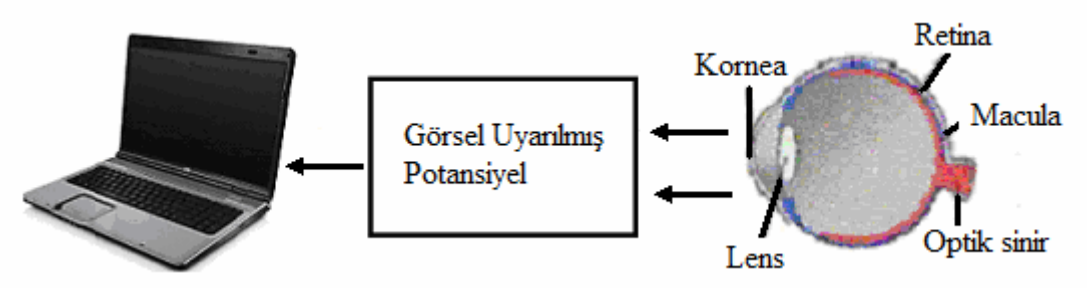


_____ Sağlıklı ------- Hastalıklı

Şekil 5.4 İki farklı insan gözünün VEP cevap örnekleri (Güven 2005)

VEP sinyalli deneyler 129 kişi üzerinde oluşmuştur. Çalışma grubu, 33' den 49'a kadar değişen yaş aralığıyla ve 43.5 yaş ortalamasıyla 55 bayan ve 74 erkekten oluşmaktadır. Gruptaki, 129 kişiden 61' i sağlıklı kişi, geriye kalan ise optik sinir hastalıklı kişilere ait VEP sinyalleridir. Optik sinir hastalığına sahip olan grup, 47 yaş ortalamasıyla 39 bayan ve 29 erkekten oluşmaktadır ve sağlıklı kişiler ise 40 yaş ortalamasıyla 16 bayan ve 45 erkekten oluşmaktadır (Güven ve ark. 2008).

Tez çalışmasında kullanılan VEP sinyalleri Prof. Dr. Sadık Kara ve Dr. Ayşegül Güven tarafından Erciyes Üniversitesi Tıp Fak. Oftalmoloji (Göz Bilimi) Bölümünde alınmıştır. VEP, kafatasına elektrotların yerleştirilmesiyle kaydedilir. Toprak ve referans elektrotları, kulak kırpma elektrotları olarak tercih edilir. Kayıttan önce, her iki kulak memesisin her iki tarafı alkolle temizlenir. VEP kayıtları, International 10-20 System of electrode placement göre yapılmıştır (Güven ve ark. 2008). Şekil 5.5, VEP sinyalini elde etmek için kullanılan sistemin blok diyagramını göstermektedir (Kara 2007b).



Şekil 5.5 VEP sinyalini elde etmek için kullanılan sistemin blok diyagramı (Güven 2005)

Kaydedilen veri, X-Y bileşenlerine dönüştürülmüştür ve VEP sinyallerinden (512 nokta) örnek veri bloğu elde etmek için interpolasyona tabi tutulmuştur. Bu sayede 512 noktalı VEP sinyali 63 noktaya azaltılmıştır. Bu sayede hem işlem maliyeti azalmış hem de birbirine benzeyen noktalar VEP sinyalinden uzaklaştırılmıştır. Tablo 5.12' de VEP sinyali ile optik sinir hastalığı veri kümesinde her bir durum için çıkış değerlerine bağlı özellikler verilmiştir. Tablo 5.13' de ise VEP sinyali ile optik sinir hastalığı veri kümesindeki özelliklerin istatistiki değerleri verilmiştir.

Tablo 5.12 VEP sinyali ile optik sinir hastalığı veri kümesinde her bir durum için çıkış değerlerine bağlı özellikler

Durum	**x_1**	**x_2**	**x_3**	**…**	**…**	**x_{59}**	**x_{60}**	**x_{61}**	**x_{62}**	**x_{63}**	**Teşhis**
1	-6.4	-5.54	-4.68	…	…	-7.65	-8.69	-10.59	-11.74	-11.27	Sağlıklı
2	-4.37	-3.56	-2.05	…	…	-12.24	-12.64	-13.66	-15.18	-15.79	Sağlıklı
...	…	…	…	…	…	…	…	…	…	…	...
...	…	…	…	…	…	…	…	…	…	…	...
129	-0.25	-1.04	-1.24	…	…	-5.15	-6.06	-5.79	-6.26	-7.10	Hasta

Tablo 5.13 VEP sinyali ile optik sinir hastalığı veri kümesinde bulunan özelliklerin istatistiki değerleri

Özellik Numarası	Minimum Değeri	Maksimum Değeri	Ortalama Değeri	Standart Sapma
1	-10.559	18.066	-0.82167	3.8012
2	-10.01	20.386	-0.8617	3.6893
3	-5.9082	24.78	-0.67599	3.8687
4	-7.3242	22.949	-0.55112	3.6535
5	-12.329	18.921	-0.65298	3.8326
6	-13.977	11.963	-1.0524	3.2918
7	-12.268	5.9814	-1.3026	2.8127
8	-16.663	3.7354	-1.2861	3.0135
9	-20.081	7.4219	-1.0798	3.4383
10	-17.822	3.5278	-1.0954	2.9742
11	-9.5215	5.5542	-1.0548	2.7334
12	-10.083	11.536	-1.085	3.2219
13	-11.816	11.841	-1.4488	3.5101
14	-10.754	7.019	-2.0826	3.7501
15	-14.648	7.1289	-1.7272	4.3401
16	-17.456	8.252	-0.33519	4.5787
17	-19.775	9.8022	1.2351	5.0441
18	-21.729	16.223	2.6195	6.0573
19	-21.729	21.313	4.0065	6.7353
20	-21.24	23.34	5.4419	7.4791
21	-22.034	23.682	6.4977	8.5128
22	-25.696	22.449	6.7335	9.1991
23	-24.963	18.762	6.3981	8.9118
24	-27.588	17.09	5.7385	8.3993
25	-30.334	15.601	4.5496	8.1079
26	-32.654	15.784	2.8929	7.9245
27	-34.241	14.026	1.3082	7.7615
28	-34.424	14.404	0.16129	7.8775
29	-37.598	14.795	-0.84669	8.5759
30	-41.016	17.383	-2.0166	8.9682
31	-40.771	18.445	-2.6936	8.2737
32	-33.203	16.65	-2.7306	7.174
33	-23.56	12.012	-2.6477	6.5354
34	-18.152	14.038	-2.7688	6.3704
35	-14.294	13.062	-2.8548	6.0137
36	-13.147	14.526	-2.6807	5.786
37	-13.306	14.282	-2.5749	5.4916
38	-13.428	15.503	-2.9671	5.6309
39	-14.893	14.099	-3.4844	5.9548
40	-17.456	13.489	-3.526	6.2636
41	-21.362	10.315	-3.544	6.2405
42	-21.973	7.1777	-3.9028	6.2788
43	-21.973	9.8999	-4.2707	6.2409
44	-19.556	11.145	-4.3233	6.1921
45	-18.799	8.4106	-4.3483	5.9207
46	-16.919	9.5215	-4.6089	5.7337
47	-17.505	7.8857	-4.7145	5.6739
48	-25.403	7.4585	-4.451	5.9788
49	-18.713	10.291	-4.161	5.7309
50	-17.871	11.218	-4.0894	5.4089
51	-19.446	7.8979	-4.1058	5.6006
52	-26.575	10.742	-3.6127	6.2036

53	-17.761	14.648	-2.8182	5.7198
54	-14.27	20.02	-2.3391	5.3367
55	-14.111	23.804	-2.085	5.5588
56	-16.736	29.541	-1.4422	6.5282
57	-14.417	31.006	-0.85509	6.6854
58	-14.233	32.227	-0.72522	6.8185
59	-13.245	30.884	-0.87971	7.0095
60	-12.646	32.898	-0.77625	7.7911
61	-13.66	32.654	-0.57033	7.8639
62	-15.186	29.236	-0.77425	7.7137
63	-17.883	28.198	-1.1889	7.3541

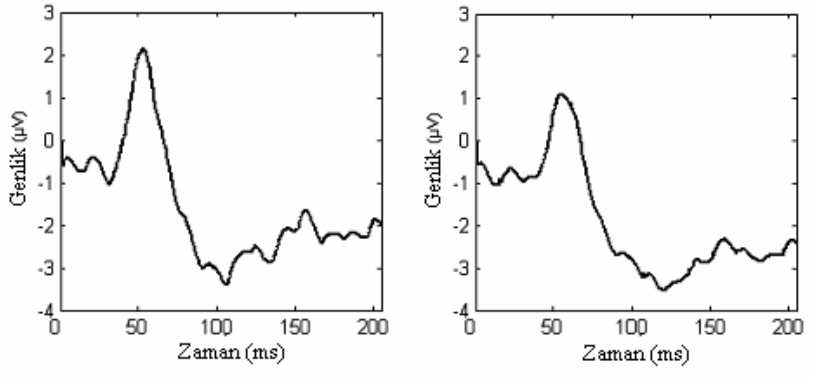
5.2.3. PERG (Pattern Electroretinography) sinyali ile Macular hastalığı veri kümesi

Macular, özel ışığa duyarlı hücreleri içeren merkezi retinada küçük bir alandır ve açık olarak detayları göstermeye izin verir. Macular, travma, iltihap, bozulma, damarlı ve arttırıcı problemleri içeren birçok pataloji tarafından etkilenebilir (Atilla ve ark. 2006, Holder ve ark. 1987, W. Tasman 1992).

Toplam 106 PERG sinyali, sağlıklı ve Macular hastalıklı kişilerden elde edildi. 106 kişilik grup, 30’ dan 55’ e değişen yaş aralığı ve 43 yaş ortalamasıyla 61 bayan ve 45 erkekten oluşmaktadır. Çalışma grubunun 56’ sı Macular hastalıklı geriye kalan ise sağlıklı kişilerden oluşmaktadır. Macular hastalığına sahip grup, 44.6 yaş ortalamasıyla 23 bayan ve 33 erkekten oluşmaktadır ve 50 sağlıklı grup ise 41 yaş ortalamasıyla 38 bayan ve 12 erkekten oluşur. Şekil 5.6, bileşenlerle etiketlenmiş normal ve hastalıklı iki gözün PERG sinyallerini gösterilmektedir (Güven 2005).

Elektrofizyolojik test cihazları, deneyler sırasında kullanılmış ve cihazdan elde edilen sinyaller hesaba katılmıştır. PERG sinyallerinin kaydı, Erciyes Üniversitesi Tıp Fakültesi Oftalmoloji (Göz Bilimi) Bölümünde Tomey Primus 2.5 electrophysiology birimi kullanılarak yapılmıştır. Her bir grup için PERG sinyalleri Şekil 5.6’ da gösterilmektedir (Güven 2005).

Ham PERG sinyali 512 noktadan oluşmaktadır ama bu nokta sayısı fazla olduğu için PERG sinyallerinin interpolasyanı alındıktan sonra 63 noktaya azaltılmıştır (Güven ve ark. 2006).



a. Sağlıklı b. Hastalıklı

Şekil 5.6 İki farklı kişiye ait PERG cevabı sinyalleri (Güven 2005)

Tablo 5.14' de PERG sinyali ile Macular hastalığı veri kümesinde her bir durum için çıkış değerlerine bağlı özellikler verilmiştir. Tablo 5.15' de ise PERG sinyali ile Macular hastalığı veri kümesindeki özelliklerin istatistiki değerleri verilmiştir.

Tablo 5.14 PERG sinyali ile Macular hastalığı veri kümesinde her bir durum için çıkış değerlerine bağlı özellikler

Durum	x_1	x_2	x_3	**…**	**…**	x_{59}	x_{60}	x_{61}	x_{62}	x_{63}	**Teşhis**
1	3.97	3.97	3.88	…	…	2.20	2.57	2.78	2.81	2.85	Sağlıklı
2	-0.439	-0.488	-0.671	…	…	-2.20	-2.28	-2.29	-1.97	-1.86	Sağlıklı
...	…	…	…	…	…	…	…	…	…	…	...
...	…	…	…	…	…	…	…	…	…	…	...
106	2.514	3.2471	3.0273	…	…		1.30	1.90	2.64	2.58	Hasta

Tablo 5.15 PERG sinyali ile Macular hastalığı veri kümesinde bulunan özelliklerin istatistiki değerleri

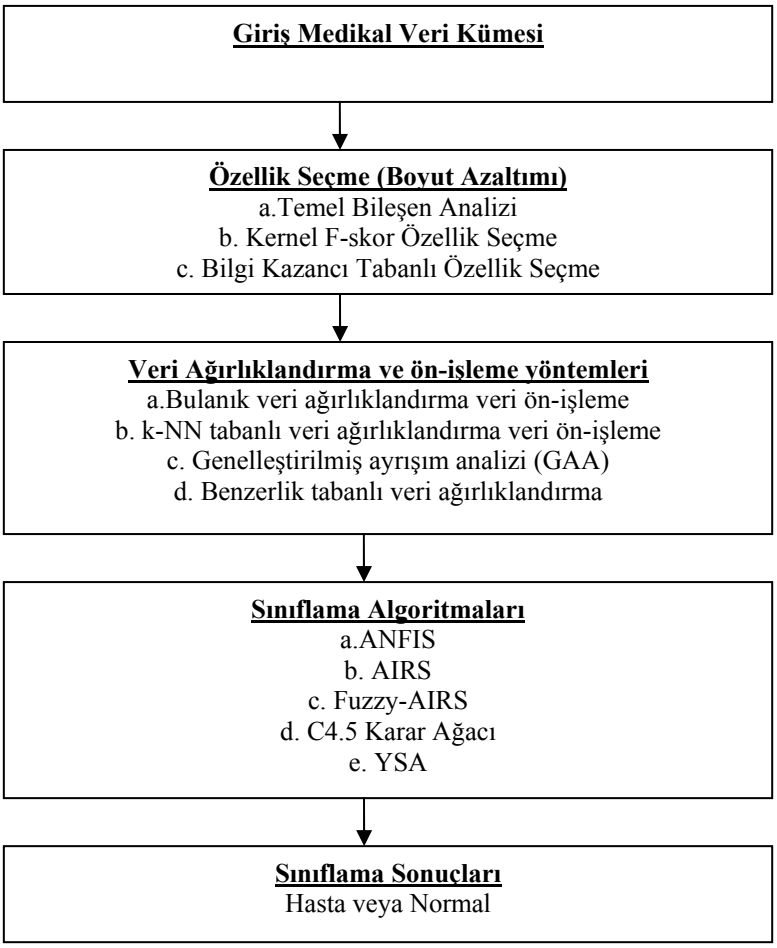
Özellik Numarası	**Minimum Değeri**	**Maksimum Değeri**	**Ortalama Değeri**	**Standart Sapma**
1	-6.5918	5.2856	-0.27178	2.0212
2	-6.3477	5.6519	-0.10376	2.003
3	-6.1646	5.3711	-0.19312	1.9167
4	-6.2134	4.7974	-0.44072	1.8551
5	-6.1646	4.5288	-0.58974	1.8639
6	-5.8594	5.1758	-0.45396	1.8745
7	-5.7617	4.8584	-0.44268	1.8234
8	-6.0913	4.5654	-0.56383	1.8095
9	-6.4575	4.1992	-0.72632	1.8029
10	-6.5186	4.5654	-0.73208	1.8095
11	-6.4697	5.0659	-0.52444	1.891
12	-6.4453	4.8218	-0.27212	1.9345
13	-6.0791	4.9072	0.17263	1.9722
14	-5.6274	5.7373	0.7429	2.1315

15	-4.9438	7.0068	1.449	2.2969
16	-4.6387	7.4829	1.7089	2.3561
17	-4.7607	8.0811	1.3899	2.3586
18	-5.4932	7.9346	0.78263	2.2982
19	-6.1035	7.2876	0.22733	2.1165
20	-6.7139	5.8594	-0.158	2.0031
21	-7.2021	4.5776	-0.65745	1.8731
22	-7.6904	3.6743	-1.3197	1.7261
23	-8.1787	2.9785	-2.016	1.699
24	-8.5449	2.3438	-2.4862	1.7509
25	-8.8379	1.5991	-2.6606	1.7873
26	-9.3872	1.0254	-2.925	1.802
27	-9.3628	1.1353	-3.1825	1.7886
28	-9.3994	0.65918	-3.4339	1.7418
29	-9.2773	0.72022	-3.4681	1.7345
30	-9.5215	0.90332	-3.446	1.8134
31	-9.6436	0.56152	-3.4996	1.7979
32	-9.8877	0.37842	-3.5421	1.7994
33	-9.8877	1.3794	-3.4912	1.837
34	-9.7656	0.74463	-3.2176	1.8296
35	-9.5215	0.81787	-3.2122	1.8258
36	-9.2773	0.40283	-3.3805	1.8543
37	-9.2773	0.37842	-3.4668	1.8275
38	-9.0332	0.5127	-3.3073	1.776
39	-9.0332	2.6245	-2.8737	1.8403
40	-8.7891	1.3672	-2.7523	1.7735
41	-8.5449	0.59815	-2.8604	1.6639
42	-8.4229	0.46387	-2.9819	1.6757
43	-8.0566	1.0376	-2.7943	1.6922
44	-8.1787	1.8799	-2.4743	1.7921
45	-8.0566	3.54	-2.3258	1.871
46	-8.0933	3.0884	-2.3461	1.8433
47	-8.1177	1.4526	-2.4514	1.7491
48	-7.8979	1.3428	-2.3137	1.7432
49	-7.7393	1.5015	-2.2128	1.7912
50	-7.6294	1.7944	-2.1867	1.8275
51	-7.666	3.4058	-2.1552	1.9128
52	-7.6416	3.2593	-2.0772	1.9265
53	-7.3853	2.0874	-1.8995	1.8802
54	-7.5684	2.1362	-2.0043	1.8512
55	-7.7881	2.1118	-2.2109	1.8283
56	-7.7271	2.0996	-2.3337	1.8189
57	-7.5684	2.7466	-2.1946	1.9187
58	-7.4097	3.7476	-1.9547	2.0109
59	-7.4219	2.2095	-2.0322	1.9222
60	-7.5562	2.5757	-2.23	1.8627
61	-7.6294	2.7832	-2.3349	1.8794
62	-7.8735	2.8198	-2.0637	1.9652
63	-7.9956	2.8564	-1.8205	2.0725

6. VERİ ÖN-İŞLEME YÖNTEMLERİNİN MEDİKAL VERİ KÜMELERİNİ SINIFLAMADAKİ PERFORMANSI VE UYGULAMA SONUÇLARI

Bu bölümde, önerilen ve literatürde var olan veri ön-işleme tekniklerinin biyomedikal veri kümelerini sınıflama performasları karşılaştırılarak ve bu yöntemler arasından verileri sınıflayan en iyi uzman sistem seçilmiştir. Veri ön-işleme kısmı iki aşamada incelenmiştir. İlk aşamada veri ağırlıklandırma yöntemlerinin ve daha sonra da özellik seçme yöntemlerinin etkisi incelenmiştir. Şekil 6.1, medikal verileri sınıflamak için kullanılan sistemin blok diyagramını göstermektedir. Bu blok diyagramda, veri ağırlıklandırma ön-işleme olarak, *bulanık mantık tabanlı veri ağırlıklandırma*, *k-NN (k-en yakın komşu) tabanlı veri ağırlıklandırma*, *genelleştirilmiş ayrışım analizi* (*Kernel Fisher Ayrışım Analizi*) ve *benzerlik tabanlı veri ağırlıklandırma* yöntemleri kullanılmış ve önerilmiştir. Boyut azaltımı ve özellik seçme algoritması olarak, *temel bileşen analizi*, *bilgi kazancına dayanan özellik seçme* algoritması ve *Kernel F-skor özellik seçme* yöntemleri özelik seçme algoritmaları olarak kullanılmıştır. Sınıflama algoritması olarak da *ANFIS*, *C4.5 karar ağacı*, *AIRS*, *Bulanık-AIRS* ve *YSA* sınıflama algoritmaları kullanılmıştır.

Biyomedikal veri kümelerini sınıflamada, veri kümeleri, 5-kat çaprazlama kullanılarak eğitme ve test kümelerine ayrılmıştır. Sınıflama performansını görmek için sınıflandırıcının sınıflama doğruluğu, duyarlılık ve seçicilik değerleri ve ROC eğrisi altında kalan değerleri (AUC-Area Under the ROC curve) kullanılmıştır. Ayrıca yöntemler arasında karşılaştırma yaparken de bu değerlerden yararlanılmıştır.



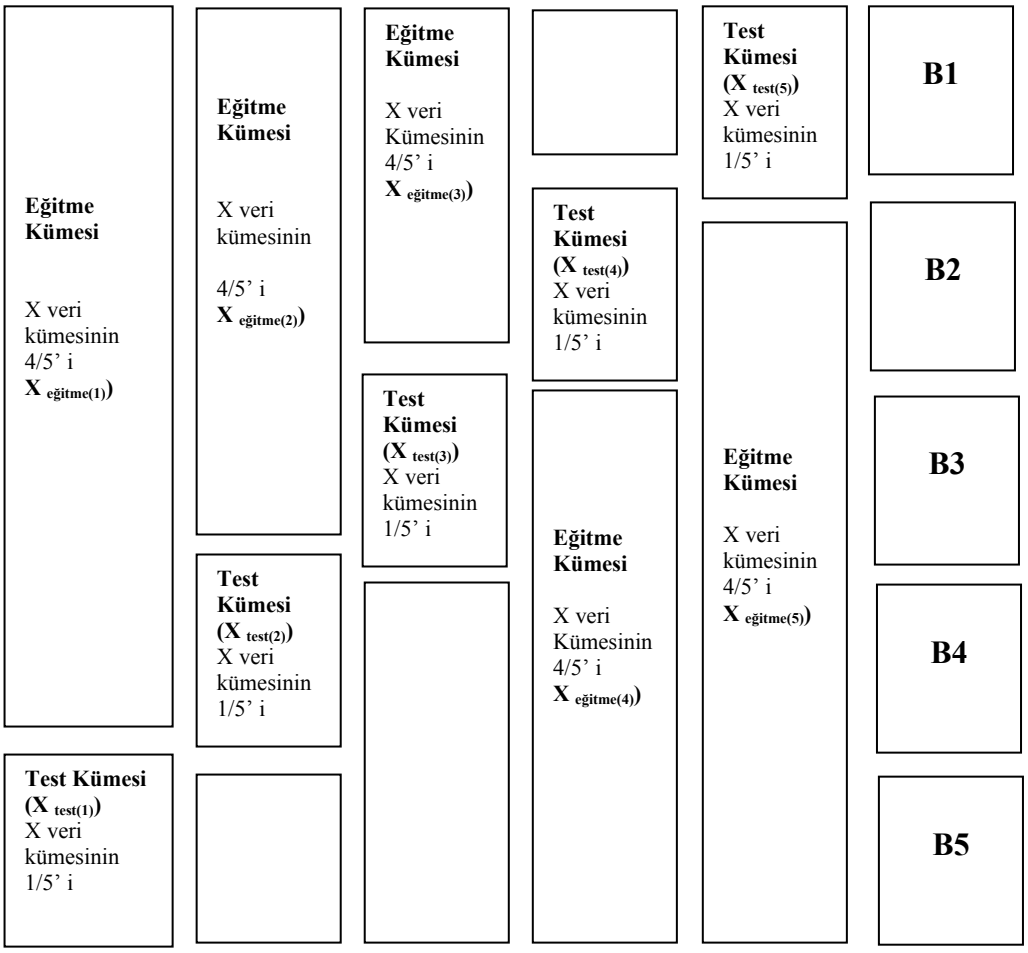
Şekil 6.1 Medikal verileri sınıflamak için kullanılan sistemin blok diyagramı

6.1. Performans Ölçüm Kriterleri

Medikal veri kümelerini sınıflamada sınıflandırıcıların ve kullanılan veri ön-işleme yöntemlerinin performansları görmek için çeşitli istatiksel ölçümler yapılır. Bu ölçümler içinden performansları karşılaşırmak amacıyla 5- kat çaprazlama ile veri kümeleri ayırma, sınıflama doğruluğu, duyarlılık ve seçicilik değerleri ve ROC eğrisi altında kalan değerleri (AUC-Area Under the ROC curve) kullanılmıştır.

6.1.1. 5-kat kez çaprazlama ile veri kümeleri ayırma

Şekil 6.2, 5 kat çaprazlama ile ilgili şematik bir diyagramı göstermektadir. Bu diyagramda, B_1' den B_5' e kadar test parçası oluşmaktadır. İlk olarak X veri kümesi 5 parçaya ayrılır bunun 4 parçası eğitim, 1 parçası da test veri kümesi olarak kullanılır. Daha sonra test kümesi X veri kümesinin sondan ikinci parçası olur geriye kalan da eğitme kümesi olarak seçilir. Bu işlem, son parçanın test veri kümesi olmasına kadar devam eder.



Şekil 6.2 Beş kat çaprazlamayı gösteren şematik diyagram

6.1.2. Sınıflama doğruluğu ve duyarlılık-seçicilik değerleri

Bir sınıflama sistemi tarafından yapılan gerçek ve tahmin edilmiş sınıflamalar hakkındaki bilgi, karışıklık matrisi ile verilir. Tablo 6.1, bir karışıklık matrisini gösterir.

Table 6.1 Karışıklık matrisinin gösterimi

Gerçek	**Tahmin Edilen**	
	Negatif	Pozitif
Negatif	DN	YN
Pozitif	YP	DP

Bu matrisde kullanılan terimler şu şekildedir:

- DN, bir örneğin negatif olduğu doğru tahminlerin sayısıdır.
- YP, bir örneğin pozitif olduğu yanlış tahminlerin sayısıdır.
- YN, bir örneğin negatif olduğu yanlış tahminlerin sayısıdır.
- DP, bir örneğin pozitif olduğu doğru tahminlerin sayısıdır.

Sınıflama doğruluğu eşitlik (6.1) ile gösterildiği gibi hesaplanmaktadır:

$$Sınıflama\ Doğruluğu = (\%)\frac{DP + DN}{DP + DN + YN + YP} x100 \qquad (6.1)$$

Duyarlılık ve seçicilik, bir medikal teşhis testinin sonuçlarını yorumlamak ve açıklamak için sıklıkla kullanılan bir istatistiksel yöntemdir. Duyarlılık, test edilen bütün pozitif insanların test edilen pozitif insanların sayısına oranıdır; yani Doğru Pozitifler / (Doğru Pozitifler + Yanlış Negatifler) dir. Seçicilik, test edilen bütün negatif insanların test edilen negatif insanların sayısına oranıdır; yani Doğru Negatifler / (Doğru Negatifler + Yanlış Pozitifler) dir. Seçiciliğe göre, olasılık, test kişinin hasta olmadığını veren negatifliğidir (Lavrac 1999). Duyarlılık ve seçicilik eşitlik (6.2) ve (6.3)' den hesaplanır.

$$Duyarlılık = \textbf{(\%)} \frac{DP}{DP + YN} x100 \tag{6.2}$$

$$Seçicilik = \textbf{(\%)} \frac{DN}{YP + DN} x100 \tag{6.3}$$

Sınıflama doğruluğunu, duyarlılık ve seçicilik değerlerini hesaplamak için bir örnek verelim. 50 hasta ve 50 normal kişiden bir veri kümemiz olsun. Eğitme ve test kümelerini, tüm veri kümesi üzerinden % (50-50) olarak ayıralım. Sınıflama sonuçlarında, 25 hasta test verisinden 4 yanlış sınıflama, 25 normal test verisinden de 3 hatalı sınıflama yapıldığı farz edilerek. Bu değerler Tablo 6.2' ye aktarılmıştır.

Table 6.2 Sınıflama sonuç tablosu

Gerçek	**Tahmin Edilen**	
	Normal (negatif)	Hasta (pozitif)
Normal (negatif)	22	3
Hasta (pozitif)	4	21

Yukarıdaki değerlere göre sınıflama doğruluğu eşitlik (6.4)' de verilmiştir. Duyarlılık ve seçicilik değerleri ise eşitlik (6.5) ve (6.6)' da verilmiştir.

$$Sınıflama\ Doğruluğu = \textbf{(\%)} \frac{DP+DN}{DP+DN+YN+YP} x100 = \frac{22+21}{22+21+3+4} x100 = \%86.00 \tag{6.4}$$

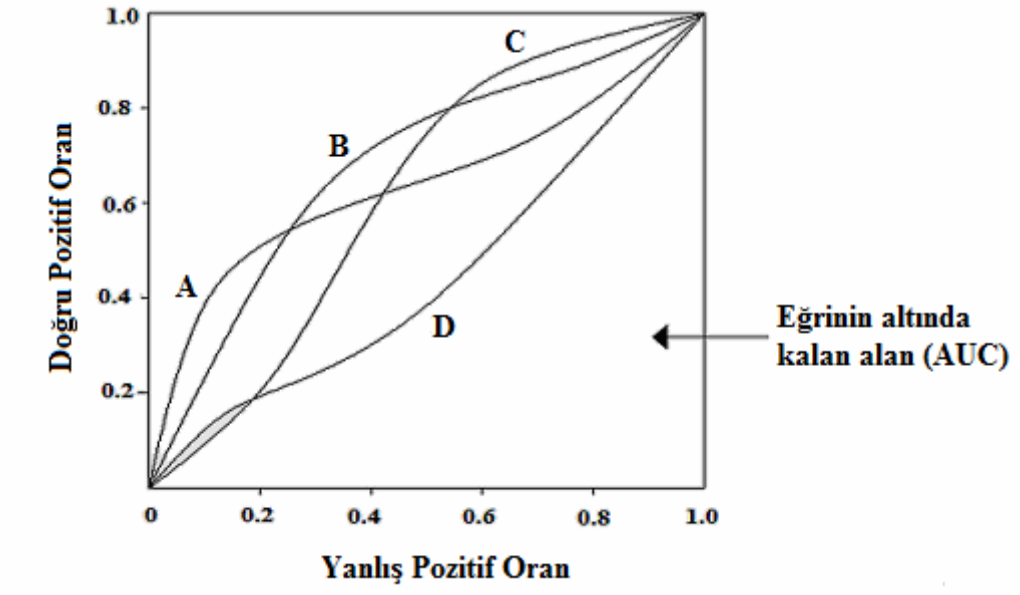
$$Duyarlılık = \textbf{(\%)} \frac{DP}{DP + YN} x100 = \frac{21}{21+4} x100 = \%84.00 \tag{6.5}$$

$$Seçicilik = \textbf{(\%)} \frac{DN}{YP + DN} x100 = \frac{22}{22+3} x100 = \%88.00 \tag{6.6}$$

6.1.3. ROC (Receiver Operating Characterictics-Alıcı İşlem Karakteristiği) eğrileri ve eğrinin altında kalan alan (AUC)

Bir alıcı işlem karakteristiği eğrisi (*ROC*), sınıflandırıcıların performanslarına dayanarak sınıflandırıcıların seçilmesi, organize edilmesi ve görselleştirmesi için kullanılan bir tekniktir. *ROC* eğrileri, genellikle medikal karar verme aşamasında kullanılır, son zamanlarda makine öğrenmesi ve veri madenciliği araştırmalarında da kullanımı gittikçe arttırmıştır (Fawcett 2005).

ROC eğrileri, doğru pozitifin *Y* ekseninde ve yanlış pozitifin *X* ekseninde çizildiği iki boyutlu grafiklerdir. Bir *ROC* eğrisi, doğru pozitifler ve yanlış pozitifler arasındaki görsel ilişkiyi açıklar. Şekil 6.3, farklı *AUC* değerlerine sahip *ROC* eğrilerini göstermektedir. *ROC* eğrisi altında kalan alan (*AUC-area under ROC curve*) ne kadar büyükse sınıflandırıcının performansı da o derece iyidir (Fawcett 2005).



Şekil 6.3 Farklı AUC değerlerine sahip ROC eğrileri

Yukarıdaki şekilde A, B, C ve D sınıflayıcılarına ait ROC eğrileri verilmiştir. Şekil 6.3' den görüleceği üzere eğrinin altında en yüksek alan B sınıflayıcısına, en düşük alan ise D sınıflayıcısına aittir. Yani bu dört sınıflayıcı içerisinden en iyi sınıflayıcı B, en kötü sınıflayıcı ise D' dir.

ROC eğrileri ve *AUC* değerinin etkisini görmek için bir örnek verilmiştir. Bu örnekte 5 +, 5 – olmak üzere 2 sınıf ve 10 örnek vardır. Bu veriyi sınıflamak için iki sınıflama algoritması kullanılmıştır. Sınıflandırıcıların elde ettikleri sonuçlar:

İdeal sınıflama durumda: – – – – – + + + + +
olmaktadır.

1. Sınıflayıcı: – – – – + | – + + + +
2. Sınıflayıcı: + – – – – | + + + + –

sonuçlarını elde etmiştir. Her iki sınıflayıcı için sınıflama doğruluğu %80' dir. Her iki sınıflayıcının *AUC* değerleri eşitlik (6.7)' ye göre hesaplanır.

$$\hat{A} = \frac{\sum_{i=1}^{n_T} r_i - n_+(n_+ + 1)/2}{n_+ n_-} \qquad (6.7)$$

Burada, r_i, (+) sınıftaki örneklerin sırasını göstermektedir. n_+ ve n_-, sırasıyla pozitif ve negatif kümelerin sayısını göstermektedir. $\hat{A}$ ise ROC eğrisi altında kalan alanıdır. Yukarıdaki örnekte 1. sınıflayıcı için AUC değeri:

r_i= 5, 7, 8, 9, 10 n_+=5 n_-=5 dir.

$$\hat{A} = \frac{\sum_{i=1}^{n_T} r_i - n_+(n_+ + 1)/2}{n_+ n_-} = \frac{(5+7+8+9+10) - (5x6)/2}{5x5} = \frac{14}{25} = 0.56 \qquad (6.8)$$

dir. 2. sınıflayıcı için AUC değeri eşitlik (6.9)' da verilmiştir.

r_i= 1, 6, 7, 8, 9 n_+=5 n_-=5 olması durumunda,

$$\hat{A} = \frac{\sum_{i=1}^{n_T} r_i - n_+(n_+ + 1)/2}{n_+ n_-} = \frac{(1+6+7+8+9) - (5x6)/2}{5x5} = \frac{16}{25} = 0.64 \quad (6.9)$$

dir. Yukarıdaki sınıflayıcıların sınıflama doğrulukları aynı olmamakla birlikte AUC değerleri farklı çıkmıştır. Sınıflayıcıları karşılaştırken sadece sınıflama doğruluğunun kullanılması başlı başına güvenli olmayabilir, bunun yanında AUC değerlerinin de kullanılması gerekir.

6.2. Veri Ağırlıklandırma Yöntemlerinin Etkisi

Bu tez çalışmasında önerilen ve kullanılan veri ağırlıklandırma ve ön-işleme yöntemleri; *bulanık ağırlıklandırma ön-işleme*, *k-NN (k- en yakın komşu) tabanlı veri ağırlıklandırma ön-işleme*, *genelleştirilmiş ayrışım analizi* (*Kernel Fisher Ayrışım Analizi*) ve *benzerlik tabanlı veri ağırlıklandırma ön-işleme* yöntemleridir.

İlk olarak medikal veri kümelerine herhangi bir ön-işleme uygulaması yapılmadan sınıflandırıcıların performansları incelenmiştir. Tablo 6.3 kalp hastalığı, Tablo 6.4 E.coli promoter gen dizileri, Tablo 6.5 SPECT görüntüleri, Tablo 6.6 Doppler sinyali ile damar sertliği, Tablo 6.7 VEP sinyali ile optik sinir hastalığı Tablo 6.8 ise PERG sinyali ile Macular hastalığı veri kümesi için gerçekleştirilen sınıflandırıcıların elde ettiği sonuçları göstermektedir.

Tablo 6.3 Kalp hastalığı veri kümesinin sınıflandırılmasında ön-işlemesiz sınıflandırıcılardan elde edilen sonuçlar

ÖN-İŞLEMESİZ SONUÇLAR *Sınıflandırıcılar*	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	54.81	49.59	85.00	0.588
C4.5 Karar Ağacı	79.25	77.58	80.51	0.788
AIRS	55.56	0	100	0.50
Fuzzy-AIRS	51.85	41.93	54.80	0.488
YSA	71.11	67.11	74.32	0.708

Tablo 6.4 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında ön-işlemesiz sınıflandırıcılardan elde edilen sonuçlar

ÖN-İŞLEMESİZ SONUÇLAR *Sınıflandırıcı*	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	63.46	62.96	64	0.634
C4.5 Karar Ağacı	82.09	81.48	84	0.826
AIRS	50	50	0	0.50
Fuzzy-AIRS	50	50	0	0.50
YSA	59.61	55.17	61.90	0.596

Tablo 6.5 SPECT görüntüleri veri kümesinin sınıflandırılmasında ön-işlemesiz sınıflandırıcılardan elde edilen sonuçlar

ÖN-İŞLEMESİZ SONUÇLAR *Sınıflandırıcı*	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	77.44	88	45.45	0.692
C4.5 Karar Ağacı	79.69	89.89	50	0.734
AIRS	81.20	88.57	53.57	0.716
Fuzzy-AIRS	80.45	85.08	52.63	0.642
YSA	77.44	87.25	45.16	0.679

Tablo 6.6 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında ön-işlemesiz sınıflandırıcılardan elde edilen sonuçlar

ÖN-İŞLEMESİZ SONUÇLAR *Sınıflandırıcı*	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	98.24	96.77	100	0.981
C4.5 Karar Ağacı	91.22	90.32	92.30	0.911
AIRS	98.25	96.77	100	0.981
Fuzzy-AIRS	98.25	96.77	100	0.981
YSA	98.25	96.77	100	0.981

Tablo 6.7 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında ön-işlemesiz sınıflandırıcılardan elde edilen sonuçlar

ÖN-İŞLEMESİZ SONUÇLAR *Sınıflandırıcı*	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	73.43	70.73	78.26	0.726
C4.5 Karar Ağacı	79.68	81.81	77.41	0.797
AIRS	48.44	54.54	47.16	0.504
Fuzzy-AIRS	48.44	100	47.61	0.514
YSA	75	76.47	73.34	0.749

Tablo 6.8 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında ön-işlemesiz sınıflandırıcılardan elde edilen sonuçlar

ÖN-İŞLEMESİZ SONUCLAR ***Sınıflandırıcı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	100	100	100	1.00
C4.5 Karar Ağacı	84.90	85.71	84	0.848
AIRS	54.72	53.84	100	0.52
Fuzzy-AIRS	54.72	53.84	100	0.52
YSA	94.33	93.10	95.84	0.942

6.2.1. Bulanık mantık veri ağırlıklandırma yöntemi sonuçları

Bulanık mantık veri ağırlıklandırma ön-işleme ile medikal veri kümeleri [0,1] aralığında ağırlıklandırılmıştır. Bu ağırlıklandırma yöntemindeki amaç, veri kümesini oluşturan özelliklerde birbirine yakın verilere aynı ağırlık katsayısı verilerek o özellik içirisindeki değişimi azaltmaktır. Bu sayede, medikal veri kümesini oluşturan hasta ve normal sınıflardaki verileri bir araya toplayarak sınıflar arasındaki ayrışımı daha da ayrıştırılabilir hale getirilmektedir.

Tablo 6.9 kalp hastalığı, Tablo 6.10 E.coli promoter gen dizileri, Tablo 6.11 SPECT görüntüleri, Tablo 6.12 Doppler sinyali ile damar sertliği, Tablo 6.13 VEP sinyali ile optik sinir hastalığı, Tablo 6.14 ise PERG sinyali ile Macular hastalığı veri kümesi için bulanık mantık veri ağırlıklandırma ve veri ön-işleme ile sınıflandırıcıların birleşiminden elde edilen sonuçları göstermektedir.

Tablo 6.9 Kalp hastalığı veri kümesinin sınıflandırılmasında, bulanık mantık veri ağırlıklandırma yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	74.81	69.69	79.71	0.75
C4.5 Karar Ağacı	91.11	87.50	94.36	0.913
AIRS	80.74	78.33	82.66	0.805
Fuzzy-AIRS	82.22	80.00	84.00	0.82
YSA	87.40	84.12	90.27	0.875

Tablo 6.10 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, bulanık mantık veri ağırlıklandırma yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	55.76	55.17	56.52	0.557
C4.5 Karar Ağacı	88.46	88.46	88.46	0.884
AIRS	55.77	53.19	80.00	0.557
Fuzzy-AIRS	51.92	51.28	50	0.519
YSA	59.61	61.29	66.67	0.596

Tablo 6.11 SPECT görüntüleri veri kümesinin sınıflandırılmasında, bulanık mantık veri ağırlıklandırma yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	98.49	99.05	96.29	0.976
C4.5 Karar Ağacı	100	100	100	1.00
AIRS	86.47	93.13	64.51	0.818
Fuzzy-AIRS	83.46	95.65	56.09	0.841
YSA	100	100	100	1.00

Tablo 6.12 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, bulanık mantık veri ağırlıklandırma yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	73.68	80	68.75	0.740
C4.5 Karar Ağacı	89.47	90	88.88	0.894
AIRS	52.63	71.42	50	0.546
Fuzzy-AIRS	47.37	0	47.37	0.50
YSA	66.67	68.96	64.28	0.667

Tablo 6.13 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, bulanık mantık veri ağırlıklandırma yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	87.50	80.95	100	0.866
C4.5 Karar Ağacı	85.93	87.87	83.87	0.859
AIRS	100	100	100	1.00
Fuzzy-AIRS	100	100	100	1.00
YSA	84.37	77.27	100	0.833

Tablo 6.14 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, bulanık mantık veri ağırlıklandırma yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	100	100	100	1.00
C4.5 Karar Ağacı	60.37	62.96	57.69	0.603
AIRS	50.94	100	49.01	0.535
Fuzzy-AIRS	47.17	0	47.20	0.50
YSA	73.58	69.45	82.35	0.726

Yukarıda elde edilen sonuçlara göre *bulanık veri ağırlıklandırma* ile sınıflandırıcı birleşimi hibrid sistem, 6 medikal veri kümesi içinden kalp hastalığı, E.coli promoter gen dizileri, SPECT görüntüleri ve VEP sinyali ile optik sinir hastalığı veri kümelerini sınflamada ön-işlemesiz elde edilen sonuçlara göre üstün sonuçlar elde etmiştir.

6.2.2. k-NN (k-en yakın komşu) tabanlı veri ağırlıklandırma yöntemi sonuçları

k-NN tabanlı veri ağırlıklandırma yöntemi, medikal veri kümelerini oluşturan her bir özellik için ayrı ayrı uygulanmıştır. Bu yöntemde, her bir özellik sütunundaki verilerin yine aynı sütunda kendilerine en yakın 5, 10, 15 ve 20 veri değerlerinin ortalamaları alınarak veri kümesi yeniden oluşturulur. Bu yöntemin avantajı, özellik içindeki değişimi (varyansı) azaltarak o özelliğin sınıflama yeteneğindeki performansını arttırmaktır. Ağırlıklandırma işleminde kullanılan *k* değeri kullanılan veri kümesine göre değişmektedir.

Sonuç tabloları *k* değerinin yine 5, 10, 15 ve 20 değerlerine göre oluşturulmuştur. Tablo 6.15 kalp hastalığı, Tablo 6.16 E.coli promoter gen dizileri, Tablo 6.17 SPECT görüntüleri, Tablo 7.18 Doppler sinyali ile damar sertliği, Tablo 6.19 VEP sinyali ile optik sinir hastalığı, Tablo 6.20 ise PERG sinyali ile Macular hastalığı veri kümesinde, *k-NN tabanlı veri ağırlıklandırma veri ön-işleme* ile sınıflandırıcıların birleşiminden elde edilen sonuçları göstermektedir.

Tablo 6.15 Kalp hastalığı veri kümesinin sınıflandırılmasında, k-NN tabanlı veri ağırlıklandırma yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

k-NN tabanlı veri ağırlıklandırma ön-işlemede kullanılan k değeri	***Sınıflandırıcı***	***ANFIS***	***C4.5 Karar Ağacı***	***AIRS***	***Fuzzy-AIRS***	***YSA***
5	*Sınıflama Doğruluğu (%)*	73.34	82.22	73.33	74.81	80
	Duyarlılık (%)	70.68	77.27	68.75	80.95	80
	Seçicilik (%)	75.32	86.95	77.46	72.04	80
	ROC eğrisi altında kalan alan	0.728	0.825	0.733	0.73	0.793
10	*Sınıflama Doğruluğu (%)*	73.34	82.96	74.07	79.26	77.03
	Duyarlılık (%)	70.68	78.46	71.18	83.34	74.58
	Seçicilik (%)	75.32	86.95	76.31	77.01	78.95
	ROC eğrisi altında kalan alan	0.728	0.831	0.736	0.78	0.766
15	*Sınıflama Doğruluğu (%)*	71.85	82.22	80.74	74.81	77.03
	Duyarlılık (%)	68.34	77.27	82.69	80.95	71.64
	Seçicilik (%)	74.67	86.95	79.52	72.04	82.35
	ROC eğrisi altında kalan alan	0.715	0.825	0.798	74.81	0.773
20	*Sınıflama Doğruluğu (%)*	72.59	79.25	79.26	74.81	77.03
	Duyarlılık (%)	66.67	77.58	77.58	80.95	74.58
	Seçicilik (%)	78.78	80.51	80.52	72.04	78.95
	ROC eğrisi altında kalan alan	0.73	0.788	0.788	0.73	0.766

Tablo 6.16 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, k-NN tabanlı veri ağırlıklandırma yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

k-NN tabanlı veri ağırlıklandırma ön-işlemede kullanılan k değeri	***Sınıflandırıcı***	***ANFIS***	***C4.5 Karar Ağacı***	***AIRS***	***Fuzzy-AIRS***	***YSA***
5	*Sınıflama Doğruluğu (%)*	61.53	82.69	51.92	51.92	59.61
	Duyarlılık (%)	60.72	84.00	51.28	51.28	55.17
	Seçicilik (%)	62.50	81.48	50	50	61.90
	ROC eğrisi altında kalan alan	0.615	0.826	0.519	0.519	0.516
10	*Sınıflama Doğruluğu (%)*	61.53	82.69	51.92	51.92	59.61
	Duyarlılık (%)	60.72	84.00	51.28	51.28	55.17
	Seçicilik (%)	62.50	81.48	50	50	61.90
	ROC eğrisi altında kalan alan	0.615	0.826	0.519	0.519	0.516
15	*Sınıflama Doğruluğu (%)*	61.53	82.69	53.85	51.92	59.61
	Duyarlılık (%)	60.72	84	52	51.28	55.17
	Seçicilik (%)	62.50	81.48	100	50	61.90
	ROC eğrisi altında kalan alan	0.615	0.826	0.538	0.519	0.516
20	*Sınıflama Doğruluğu (%)*	65.38	82.69	53.85	51.92	59.61
	Duyarlılık (%)	64.28	84	52	51.28	55.17
	Seçicilik (%)	66.67	81.48	100	50	61.90
	ROC eğrisi altında kalan alan	0.653	0.826	0.538	0.519	0.516

Tablo 6.17 SPECT görüntüleri veri kümesinin sınıflandırılmasında, k-NN tabanlı veri ağırlıklandırma yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

k-NN tabanlı veri ağırlıklandırma ön-işlemede kullanılan k değeri	***Sınıflandırıcı***	***ANFIS***	***C4.5 Karar Ağacı***	***AIRS***	***Fuzzy-AIRS***	***YSA***
5	*Sınıflama Doğruluğu (%)*	77.44	79.69	82.71	83.46	78.19
	Duyarlılık (%)	88.00	89.89	86.08	86.20	87.37
	Seçicilik (%)	45.46	50	61.12	64.70	46.67
	ROC eğrisi altında kalan alan	0.692	0.734	0.72	0.841	0.683
10	*Sınıflama Doğruluğu (%)*	77.44	79.69	81.20	81.95	79.69
	Duyarlılık (%)	88.00	89.89	84.61	85.34	86.23
	Seçicilik (%)	45.46	50	53.34	56.25	50
	ROC eğrisi altında kalan alan	0.692	0.734	0.716	0.652	0.666
15	*Sınıflama Doğruluğu (%)*	77.44	79.69	81.20	81.95	79.69
	Duyarlılık (%)	88.00	89.89	84.61	85.34	86.23
	Seçicilik (%)	45.46	50	53.34	56.25	50
	ROC eğrisi altında kalan alan	0.692	0.734	0.716	0.652	0.666
20	*Sınıflama Doğruluğu (%)*	77.44	79.69	82.71	82.71	79.69
	Duyarlılık (%)	88.00	89.89	85.47	85.47	86.23
	Seçicilik (%)	45.46	50	60	60	50
	ROC eğrisi altında kalan alan	0.692	0.734	0.656	0.656	0.666

Tablo 6.18 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, k-NN tabanlı veri ağırlıklandırma yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

k-NN tabanlı veri ağırlıklandırma ön-işlemede kullanılan k değeri	***Sınıflandırıcı***	***ANFIS***	***C4.5 Karar Ağacı***	***AIRS***	***Fuzzy-AIRS***	***YSA***
5	*Sınıflama Doğruluğu (%)*	85.96	91.22	84.21	84.21	85.96
	Duyarlılık (%)	100	90	95.65	95.65	100
	Seçicilik (%)	7714	88.89	76.47	76.47	77.14
	ROC eğrisi altında kalan alan	0.866	0.894	0.848	0.848	0.866
10	*Sınıflama Doğruluğu (%)*	85.96	92.98	84.21	84.21	85.96
	Duyarlılık (%)	100	93.34	95.65	95.65	100
	Seçicilik (%)	7714	92.59	76.47	76.47	77.14
	ROC eğrisi altında kalan alan	0.866	0.929	0.848	0.848	0.866
15	*Sınıflama Doğruluğu (%)*	85.96	96.49	85.96	84.21	85.96
	Duyarlılık (%)	100	96.67	100	95.65	100
	Seçicilik (%)	7714	96.29	77.14	76.47	77.14
	ROC eğrisi altında kalan alan	0.866	0.964	0.866	0.848	0.866
20	*Sınıflama Doğruluğu (%)*	85.96	92.98	85.96	85.96	85.96
	Duyarlılık (%)	100	93.34	100	100	100
	Seçicilik (%)	7714	92.59	77.14	77.14	77.14
	ROC eğrisi altında kalan alan	0.866	0.929	0.866	0.866	0.866

Tablo 6.19 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, k-NN tabanlı veri ağırlıklandırma yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

k-NN tabanlı veri ağırlıklandırma ön-işlemede kullanılan k değeri	***Sınıflandırıcı***	***ANFIS***	***C4.5 Karar Ağacı***	***AIRS***	***Fuzzy-AIRS***	***YSA***
5	*Sınıflama Doğruluğu (%)*	95.31	87.50	79.69	76.56	82.81
	Duyarlılık (%)	94.28	88.23	92	85.18	84.84
	Seçicilik (%)	96.55	86.67	71.80	70.27	80.64
	ROC eğrisi altında kalan alan	0.951	0.874	0.804	0.771	0.828
10	*Sınıflama Doğruluğu (%)*	75	87.50	79.69	81.25	82.81
	Duyarlılık (%)	75	88.23	92	86.67	79.48
	Seçicilik (%)	75	86.67	71.80	76.47	88
	ROC eğrisi altında kalan alan	0.747	0.874	0.804	0.815	0.822
15	*Sınıflama Doğruluğu (%)*	73.43	81.25	71.87	76.56	81.25
	Duyarlılık (%)	72.98	82.53	83.34	95.23	80.56
	Seçicilik (%)	74.07	80	65	67.45	82.14
	ROC eğrisi altında kalan alan	0.730	0.811	0.727	0.771	0.809
20	*Sınıflama Doğruluğu (%)*	78.12	96.87	73.43	82.81	87.50
	Duyarlılık (%)	76.31	97.05	79.31	96	84.21
	Seçicilik (%)	80.76	96.67	68.57	74.35	92.30
	ROC eğrisi altında kalan alan	0.776	0.968	0.738	0.836	0.870

Tablo 6.20 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, k-NN tabanlı veri ağırlıklandırma yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

k-NN tabanlı veri ağırlıklandırma ön-işlemede kullanılan k değeri	***Sınıflandırıcı***	***ANFIS***	***C4.5 Karar Ağacı***	***AIRS***	***Fuzzy-AIRS***	***YSA***
5	*Sınıflama Doğruluğu (%)*	94.33	92.45	60.38	67.92	96.22
	Duyarlılık (%)	96.29	92.85	62.96	92.85	96.42
	Seçicilik (%)	92.30	92	57.69	61.53	96
	ROC eğrisi altında kalan alan	0.944	0.924	0.596	0.702	0.962
10	*Sınıflama Doğruluğu (%)*	92.45	90.56	60.38	60.38	100
	Duyarlılık (%)	92.85	92.59	62.96	62.96	100
	Seçicilik (%)	92	88.46	57.69	57.69	100
	ROC eğrisi altında kalan alan	0.924	0.894	0.596	0.596	1.00
15	*Sınıflama Doğruluğu (%)*	92.45	92.45	64.15	64.15	100
	Duyarlılık (%)	96.15	92.85	100	100	100
	Seçicilik (%)	88.89	92	56.81	56.81	100
	ROC eğrisi altında kalan alan	0.926	0.924	0.648	0.648	1.00
20	*Sınıflama Doğruluğu (%)*	96.22	94.33	64.15	64.15	100
	Duyarlılık (%)	100	96.29	100	100	100
	Seçicilik (%)	92.60	92.30	56.81	56.81	100
	ROC eğrisi altında kalan alan	0.964	0.944	0.648	0.648	1.00

Yukarıda elde edilen sonuçlara göre *k-NN tabanlı veri ağırlıklandırma* ile sınıflandırıcı birleşimi hibrid sistem, k' nın 5 ve 20 değeri için 6 medikal veri kümesi içinden kalp hastalığı, E.coli promoter gen dizileri, SPECT görüntüleri, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile macular hastalığı veri kümelerini sınflamada ön-işlemesiz elde edilen sonuçlara göre üstün sonuçlar elde etmiştir. k' nın 10 değeri için, kalp hastalığı, E.coli promoter gen dizileri, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile macular hastalığı veri kümelerini sınflamada ön-işlemesiz elde edilen sonuçlara göre üstün sonuçlar elde etmiştir. k'nın 15 değeri için ise kalp hastalığı, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile macular hastalığı veri kümelerinde üstün sonuçlar elde etmiştir.

6.2.3. Genelleştirilmiş ayrışım analizi (Generalized Discriminant Analysis) sonuçları

Genelleştirilmiş Ayrışım Analizinin (*GAA*) diğer adı *Kernel LDA* (Linear Discriminant Analysis)'dır. Bu yöndem iki aşamadan oluşur. İlk aşamada lineer bir ayırıma sahip olmayan medikal veri kümesi kernel fonksiyonları ile daha yüksek boyutlu lineer bir uzaya dönüştürülür. Daha sonra, kernel uzaya dönüştürülen medikal veri kümesi *LDA* işlemi ile lineer olarak ayrılabilir bir veri setine dönüştürülür bu sayede daha yüksek bir sınıflama performansı elde edilebilir.

Tablo 6.21 kalp hastalığı, Tablo 6.22 E.coli promoter gen dizileri, Tablo 6.23 SPECT görüntüleri, Tablo 6.24 Doppler sinyali ile damar sertliği, Tablo 6.25 VEP sinyali ile optik sinir hastalığı, Tablo 6.26 ise PERG sinyali ile Macular hastalığı veri kümesi için *genelleştirilmiş ayrışım analizi* ile sınıflandırıcıların birleşiminden elde edilen sonuçları göstermektedir.

Tablo 6.21 Kalp hastalığı veri kümesinin sınıflandırılmasında, genelleştirilmiş ayrışım analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	97.78	98.30	97.36	0.976
C4.5 Karar Ağacı	97.03	96.67	97.34	0.970
AIRS	75.56	74.54	76.25	0.748
Fuzzy-AIRS	77.04	72.30	81.42	0.771
YSA	95.56	100	92.59	0.95

Tablo 6.22 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, genelleştirilmiş ayrışım analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	75	68.58	88.23	0.75
C4.5 Karar Ağacı	84.61	84.61	84.61	0.846
AIRS	53.85	52.17	66.67	0.538
Fuzzy-AIRS	57.69	53.34	71.42	0.557
YSA	59.61	60	59.25	0.596

Tablo 6.23 SPECT görüntüleri veri kümesinin sınıflandırılmasında, genelleştirilmiş ayrışım analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	90.97	94.33	77.28	0.860
C4.5 Karar Ağacı	98.49	99.05	96.29	0.976
AIRS	84.96	94.79	59.45	0.836
Fuzzy-AIRS	84.21	95.69	57.50	0.845
YSA	80.45	85.71	50	0.656

Tablo 6.24 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, genelleştirilmiş ayrışım analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	84.21	95.65	76.47	0.848
C4.5 Karar Ağacı	85.96	95.83	78.79	0.864
AIRS	85.96	95.83	78.79	0.864
Fuzzy-AIRS	89.47	96.15	83.87	0.898
YSA	84.21	95.65	76.47	0.848

Tablo 6.25 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, genelleştirilmiş ayrışım analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	73.68	85.18	70.27	0.771
C4.5 Karar Ağacı	90.62	91.17	90	0.905
AIRS	71.87	73.52	70	0.717
Fuzzy-AIRS	65.63	83.34	58.70	0.670
YSA	71.92	80	70.58	0.752

Tablo 6.26 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, genelleştirilmiş ayrışım analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	86.79	87.09	87.50	0.866
C4.5 Karar Ağacı	73.58	75	73.91	0.732
AIRS	54.72	53.84	100	0.52
Fuzzy-AIRS	54.72	53.84	100	0.52
YSA	98.11	96.55	100	0.982

Yukarıda elde edilen sonuçlara göre *genelleştirilmiş ayrışım analizi* ile sınıflandırıcı birleşimi hibrid sistem, kalp hastalığı, E.coli promoter gen dizileri, SPECT görüntüleri ve VEP sinyali ile optik sinir hastalığı veri kümelerini sınflamada ön-işlemesiz elde edilen sonuçlara göre üstün sonuçlar elde etmiştir.

6.2.4. Benzerlik tabanlı veri ağırlıklandırma veri ön-işleme yöntemi sonuçları

Bu veri ön-işleme yönteminde, özellikler arasındaki benzerliğe dayalı bir veri ağırlıklandırma yöntemi önerildi. Benzerlik fonksiyonları olarak dört çeşit fonksiyon kullanıldı ve veri birlikteliğiyle birleştirildi. Kullanılan benzerlik fonksiyonları: kosinüs genliği (cosine amplitude), minimum-maksimum metodu, mutlak üstel (absolute exponential) yöntemi ve üstel benzerlik katsayısı (exponential similarity coefficient) fonksiyonları olup bu fonksiyonlar birbirleriyle karşılaştırılmıştır.

Tablo 6.27 kalp hastalığı, Tablo 6.28 E.coli promoter gen dizileri, Tablo 6.29 SPECT görüntüleri, Tablo 6.30 Doppler sinyali ile damar sertliği, Tablo 6.31 VEP sinyali ile optik sinir hastalığı, Tablo 6.32 ise PERG sinyali ile Macular hastalığı veri

kümesi için benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçları göstermektedir.

Tablo 6.27 Kalp hastalığı veri kümesinin sınıflandırılmasında, benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Kullanılan Benzerlik Fonksiyonu	***Sınıflandırıcı***	***ANFIS***	***C4.5 Karar Ağacı***	***AIRS***	***Fuzzy-AIRS***	***YSA***
Kosinüs genliği (cosine amplitude)	*Sınıflama Doğruluğu (%)*	100	99.25	99.25	55.56	99.25
	Duyarlılık (%)	100	98.36	98.36	0	98.36
	Seçicilik (%)	100	100	100	55.56	100
	ROC eğrisi altında kalan alan	1.00	0.993	0.993	0.50	0.993
Min-Mak. metodu	*Sınıflama Doğruluğu (%)*	94.07	99.25	65.19	55.56	98.51
	Duyarlılık (%)	89.39	98.36	57.83	0	98.34
	Seçicilik (%)	98.55	100	76.92	55.56	98.67
	ROC eğrisi altında kalan alan	0.945	0.993	0.666	0.50	0.985
Mutlak üstel (absolute exponential)	*Sınıflama Doğruluğu (%)*	97.78	100	79.26	75.56	99.25
	Duyarlılık (%)	96.72	100	75.80	72.58	98.36
	Seçicilik (%)	98.64	100	82.19	79.45	100
	ROC eğrisi altında kalan alan	0.978	1.00	0.791	0.761	0.993
Üstel benzerlik katsayısı (exponential similarity coefficient)	*Sınıflama Doğruluğu (%)*	98.51	56.29	62.96	55.56	99.25
	Duyarlılık (%)	98.34	50.72	91.67	0	98.36
	Seçicilik (%)	98.67	62.12	60.16	55.56	100
	ROC eğrisi altında kalan alan	0.985	0.565	0.585	0.50	0.993

Tablo 6.28 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Kullanılan Benzerlik Fonksiyonu	***Sınıflandırıcı***	***ANFIS***	***C4.5 Karar Ağacı***	***AIRS***	***Fuzzy-AIRS***	***YSA***
Kosinüs genliği (cosine amplitude)	*Sınıflama Doğruluğu (%)*	51.92	76.92	51.52	51.52	63.46
	Duyarlılık (%)	100	76.92	100	100	58.13
	Seçicilik (%)	50.98	76.92	50.98	50.98	88.89
	ROC eğrisi altında kalan alan	0.519	0.769	0.519	0.519	0.634
Min-Mak. metodu	*Sınıflama Doğruluğu (%)*	65.38	26.92	51.52	55.77	69.23
	Duyarlılık (%)	66.67	26.92	100	66.67	67.85
	Seçicilik (%)	64.28	26.92	50.98	53.48	70.84
	ROC eğrisi altında kalan alan	0.653	0.269	0.519	0.50	0.692
Mutlak üstel (absolute exponential)	*Sınıflama Doğruluğu (%)*	50	9.61	50	50	59.61
	Duyarlılık (%)	50	6.45	50	50	59.25
	Seçicilik (%)	0.0	0.0	0.0	0.0	60
	ROC eğrisi altında kalan alan	0.50	0.10	0.50	0.50	0.596
Üstel benzerlik katsayısı (exponential similarity coefficient)	*Sınıflama Doğruluğu (%)*	100	94.23	82.69	80.77	84.61
	Duyarlılık (%)	100	92.59	100	80.77	82.60
	Seçicilik (%)	100	96	74.28	80.77	78.12
	ROC eğrisi altında kalan alan	1.00	0.942	0.826	0.810	0.846

Tablo 6.29 SPECT görüntüleri veri kümesinin sınıflandırılmasında, benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Kullanılan Benzerlik Fonksiyonu	***Sınıflandırıcı***	***ANFIS***	***C4.5 Karar Ağacı***	***AIRS***	***Fuzzy-AIRS***	***YSA***
Kosinüs genliği (cosine amplitude)	*Sınıflama Doğruluğu (%)*	∞	∞	∞	∞	∞
	Duyarlılık (%)	∞	∞	∞	∞	∞
	Seçicilik (%)	∞	∞	∞	∞	∞
	ROC eğrisi altında kalan alan	∞	∞	∞	∞	∞
Min-Mak. metodu	*Sınıflama Doğruluğu (%)*	∞	∞	∞	∞	∞
	Duyarlılık (%)	∞	∞	∞	∞	∞
	Seçicilik (%)	∞	∞	∞	∞	∞
	ROC eğrisi altında kalan alan	∞	∞	∞	∞	∞
Mutlak üstel (absolute exponential)	*Sınıflama Doğruluğu (%)*	99.24	99.24	78.95	78.95	97.74
	Duyarlılık (%)	99.06	99.06	79.54	79.54	97.24
	Seçicilik (%)	100	100	0.0	0.0	100
	ROC eğrisi altında kalan alan	0.981	0.981	0.495	0.495	0.944
Üstel benzerlik katsayısı (exponential similarity coefficient)	*Sınıflama Doğruluğu (%)*	83.45	77.45	59.38	54.88	78.94
	Duyarlılık (%)	86.84	88.72	78.89	88.33	86.12
	Seçicilik (%)	63.15	44.45	18.60	27.40	48
	ROC eğrisi altında kalan alan	0.689	0.651	0.36	0.62	0.66

Tablo 6.30 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Kullanılan Benzerlik Fonksiyonu	***Sınıflandırıcı***	***ANFIS***	***C4.5 Karar Ağacı***	***AIRS***	***Fuzzy-AIRS***	***YSA***
Kosinüs genliği (cosine amplitude)	*Sınıflama Doğruluğu (%)*	85.96	98.24	56.14	47.37	96.49
	Duyarlılık (%)	80.56	96.77	85.71	50	96.67
	Seçicilik (%)	95.23	100	52	47.28	96.67
	ROC eğrisi altında kalan alan	0.853	0.981	0.581	0.498	0.964
Min-Mak. metodu	*Sınıflama Doğruluğu (%)*	96.50	98.24	63.16	68.42	89.47
	Duyarlılık (%)	96.67	96.77	84.61	87.50	85.29
	Seçicilik (%)	96.29	100	56.81	60.97	95.65
	ROC eğrisi altında kalan alan	0.964	0.981	0.646	0.696	0.964
Mutlak üstel (absolute exponential)	*Sınıflama Doğruluğu (%)*	75.43	87.71	84.21	87.71	84.21
	Duyarlılık (%)	78.57	87.09	95.65	100	88.89
	Seçicilik (%)	72.41	88.46	76.47	79.41	80
	ROC eğrisi altında kalan alan	0.756	0.875	0.848	0.883	0.842
Üstel benzerlik katsayısı (exponential similarity coefficient)	*Sınıflama Doğruluğu (%)*	73.68	98.24	85.96	84.21	92.98
	Duyarlılık (%)	69.23	96.77	100	100	93.34
	Seçicilik (%)	83.34	100	77.14	75	92.59
	ROC eğrisi altında kalan alan	0.740	0.981	0.866	0.85	0.929

Tablo 6.31 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Kullanılan Benzerlik Fonksiyonu	***Sınıflandırıcı***	***ANFIS***	***C4.5 Karar Ağacı***	***AIRS***	***Fuzzy-AIRS***	***YSA***
Kosinüs genliği (cosine amplitude)	*Sınıflama Doğruluğu (%)*	93.75	98.43	90.63	92.19	95.31
	Duyarlılık (%)	89.47	97.14	85	87.17	91.89
	Seçicilik (%)	100	100	100	100	100
	ROC eğrisi altında kalan alan	0.933	0.983	0.90	0.916	0.95
Min-Mak. metodu	*Sınıflama Doğruluğu (%)*	67.18	84.37	46.88	46.88	84.37
	Duyarlılık (%)	65.85	85.29	0.0	0.0	85.29
	Seçicilik (%)	69.56	83.34	46.88	46.88	83.34
	ROC eğrisi altında kalan alan	0.663	0.843	0.50	0.50	0.843
Mutlak üstel (absolute exponential)	*Sınıflama Doğruluğu (%)*	45.31	9.37	62.50	60.93	50
	Duyarlılık (%)	48.64	14.70	67.85	62.16	52.38
	Seçicilik (%)	40.74	3.34	58.34	59.25	45.45
	ROC eğrisi altında kalan alan	0.448	0.09	0.629	0.604	0.50
Üstel benzerlik katsayısı (exponential similarity coefficient)	*Sınıflama Doğruluğu (%)*	84.37	100	65.62	70.31	85.93
	Duyarlılık (%)	78.57	100	100	100	80.48
	Seçicilik (%)	95.65	100	57.69	61.22	95.65
	ROC eğrisi altında kalan alan	0.835	1.00	0.676	0.720	0.851

Tablo 6.32 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, benzerlik tabanlı veri ağırlıklandırma ön-işleme yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Kullanılan Benzerlik Fonksiyonu	***Sınıflandırıcı***	***ANFIS***	***C4.5 Karar Ağacı***	***AIRS***	***Fuzzy-AIRS***	***YSA***
Kosinüs genliği (cosine amplitude)	*Sınıflama Doğruluğu (%)*	90.56	94.33	98.11	98.11	98.11
	Duyarlılık (%)	96	96.29	96.55	96.55	96.55
	Seçicilik (%)	85.71	92.30	100	100	100
	ROC eğrisi altında kalan alan	0.908	0.944	0.98	0.98	0.98
Min-Mak. metodu	*Sınıflama Doğruluğu (%)*	67.92	30.18	54.71	54.71	69.81
	Duyarlılık (%)	70.37	32	100	100	73.07
	Seçicilik (%)	65.38	28.57	51.02	51.02	66.67
	ROC eğrisi altında kalan alan	0.679	0.302	0.571	0.571	0.699
Mutlak üstel (absolute exponential)	*Sınıflama Doğruluğu (%)*	73.58	60.37	67.92	67.92	75.47
	Duyarlılık (%)	71.87	62.96	64.86	64.86	80
	Seçicilik (%)	76.19	57.69	75	75	71.42
	ROC eğrisi altında kalan alan	0.730	0.603	0.668	0.668	0.757
Üstel benzerlik katsayısı (exponential similarity coefficient)	*Sınıflama Doğruluğu (%)*	92.45	100	83.01	83.01	98.11
	Duyarlılık (%)	92.85	100	88	88	100
	Seçicilik (%)	92	100	78.57	78.57	96.15
	ROC eğrisi altında kalan alan	0.924	1.00	0.832	0.832	0.982

Yukarıda elde edilen sonuçlara göre benzerlik tabanlı veri ağırlıklandırma ile sınıflandırıcı birleşimi hibrid sistem, kosinüs genliği benzerlik fonksiyonu veri ağırlıklandırma yöntemi için 6 medikal veri kümesi içinden kalp hastalığı, E.coli promoter gen dizileri, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile macular hastalığı veri kümelerinde; minimum-maximum benzerlik fonksiyonu veri ağırlıklandırma yöntemi kalp hastalığı ve E.coli promoter gen dizileri veri kümelerinde; mutlak üstel (absolute exponential) benzerlik fonksiyonu veri ağırlıklandırma yöntemi kalp hastalığı ve SPECT görüntüleri veri kümelerinde; üstel benzerlik katsayısı benzerlik fonksiyonu veri ağırlıklandırma yöntemi ise kalp

hastalığı, E.coli promoter gen dizileri, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile macular hastalığı veri kümelerini sınflamada ön-işlemesiz elde edilen sonuçlara göre üstün sonuçlar elde etmiştir.

6.3. Boyut Azaltımı ve Özellik Seçme Yöntemlerinin Etkisi

Bu tez çalışmasında önerilen ve kullanılan boyut azaltımı ve özellik seçme yöntemleri, *Temel Bileşen Analizi*, *Kernel F-skor özellik seçme* yöntemi ve *bilgi kazancına dayanan özellik seçme* yöntemleridir.

6.3.1. Temel bileşen analizi ile boyut azaltımı yönteminin sonuçları

Temel bileşen analizinde, medikal veri kümelerinin boyutları yeni bir uzaya dönüştürülerek eski özelliklerden farklı yeni özellikler elde edilir. Bu yeni uzayın özelliklerini seçerken değişkenlerin varyansları hesaplanır ve büyükten küçüğe doğru sıralanır ve içlerinden en yüksek varyansı sağlayanlar seçilir, diğerleri elimine edilir. Boyut azaltımı olayı veri kümesindeki özellik sayısına ve veri sayısına göre değişmektedir.

Tablo 6.33 kalp hastalığı, Tablo 6.34 E.coli promoter gen dizileri, Tablo 6.35 SPECT görüntüleri, Tablo 6.36 Doppler sinyali ile damar sertliği, Tablo 6.37 VEP sinyali ile optik sinir hastalığı, Tablo 6.38 ise PERG sinyali ile Macular hastalığı veri kümesi için temel bileşen analizi boyut azaltımı yöntemi ile sınıflandırıcıların birleşiminden elde edilen sonuçları göstermektedir.

Tablo 6.33 Kalp hastalığı veri kümesinin sınıflandırılmasında, temel bileşen analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***TBA ile azaltılan özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	5	82.23	84.61	80.73	0.822
C4.5 Karar Ağacı	5	83.70	81.67	85.34	0.837
AIRS	5	85.19	83.34	86.67	0.851
Fuzzy-AIRS	5	82.96	89.36	79.54	0.82
YSA	5	80.74	84.00	78.83	0.807

Tablo 6.34 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, temel bileşen analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***TBA ile azaltılan özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	23	78.84	89.47	72.72	0.788
C4.5 Karar Ağacı	23	80.76	80.76	80.76	0.807
AIRS	23	82.69	81.48	84	0.826
Fuzzy-AIRS	23	88.46	100	81.25	0.884
YSA	23	78.84	80	77.78	0.788

Tablo 6.35 SPECT görüntüleri veri kümesinin sınıflandırılmasında, temel bileşen analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***TBA ile azaltılan özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	7	94.73	95.41	91.67	0.894
C4.5 Karar Ağacı	7	84.21	90.47	60.71	0.756
AIRS	7	91.73	92.03	90	0.817
Fuzzy-AIRS	7	93.23	92.17	100	0.826
YSA	7	95.48	96.29	92	0.913

Tablo 6.36 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, temel bileşen analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***TBA ile azaltılan özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	4	89.47	96.15	83.87	0.890
C4.5 Karar Ağacı	4	100	100	100	1.00
AIRS	4	98.25	96.77	100	0.982
Fuzzy-AIRS	4	100	100	100	1.00
YSA	4	100	100	100	1.00

Tablo 6.37 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, temel bileşen analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***TBA ile azaltılan özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	11	75	73.68	76.92	0.745
C4.5 Karar Ağacı	11	75	73.68	76.92	0.745
AIRS	11	81.25	89.28	75	0.817
Fuzzy-AIRS	11	81.25	78.94	84.61	0.807
YSA	11	76.56	80.64	72.72	0.767

Tablo 6.38 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, temel bileşen analizi ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***TBA ile azaltılan özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	7	84.90	79.41	94.73	0.842
C4.5 Karar Ağacı	7	94.33	96.29	92.30	0.944
AIRS	7	84.61	81.25	90.47	0.844
Fuzzy-AIRS	7	75.43	77.78	73.07	0.755
YSA	7	98.11	96.56	100	0.980

Yukarıda elde edilen sonuçlara göre temel bileşen analizi ile sınıflandırıcı birleşimi hibrid sistem, 6 medikal veri kümesini sınflamada ön-işlemesiz elde edilen sonuçlara göre bütün veri kümelerinde üstün sonuçlar elde etmiştir.

6.3.2. Kernel F-skor özellik seçme yönteminin sonuçları

F-skor ölçüm yöntemi, iki sınıflı veri kümelerini sınıflamada kullanılan her bir özelliğin, sınıfları birbirinden ayırt edebilme yeteneğini ölçmek için kullanılır. Fakat lineer bir dağılıma sahip olmayan veri kümelerinde çok iyi sonuçlar elde etmemiştir. Bu dezavantajı önlemek için veri kümesinin özellik uzayı *RBF* veya *Lineer* kernel fonksiyonları ile kernel uzayına dönüştürülür. Bu sayede lineer olarak ayrılamayan veri kümeleri kernel fonksiyonları ile yüksek boyutlu lineer olarak ayrılabilir bir veri kümesine dönüştürülür. Yapılan uygulamalarda ilk olarak *F-skor özellik seçmenin* performansı incelendi. Daha sonra *RBF* ve *Lineer* kernel fonksiyon ön-işlemeli *F-skor özellik seçme* yöntemlerinin başarıları incelenmiştir.

F-skor özellik seçme kullanılarak medikal veri kümeleri sınıflamada elde edilen sonuçlar Tablo 6.39 ile Tablo 6.44 arasında verilmiştir. Tablo 6.39 kalp hastalığı, Tablo 6.40 E.coli promoter gen dizileri,Tablo 6.41 SPECT görüntüleri, Tablo 6.42 Doppler sinyali ile damar sertliği, Tablo 6.43 VEP sinyali ile optik sinir hastalığı ve Tablo 6.44 ise PERG sinyali ile Macular hastalığı veri kümesinde, *f-skor özellik seçme* ile sınıflandırıcıların birleşiminin elde ettiği sonuçları göstermektedir.

Tablo 6.39 Kalp hastalığı veri kümesinin sınıflandırılmasında, f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	13	6	97.78	96.72	98.65	0.98
C4.5 Karar Ağacı	13	6	77.04	77.04	82.43	0.8
AIRS	13	6	80	80	68	0.711
Fuzzy-AIRS	13	6	80	80	68	0.711
YSA	13	6	77.61	77.61	88.23	0.821

Tablo 6.40 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	57	17	59.61	58.62	60.86	0.596
C4.5 Karar Ağacı	57	17	84.61	84.61	84.61	0.846
AIRS	57	17	51.92	50.98	100	0.519
Fuzzy-AIRS	57	17	51.92	50.98	100	0.519
YSA	57	17	67.30	63.63	73.68	0.673

Tablo 6.41 SPECT görüntüleri veri kümesinin sınıflandırılmasında, f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	22	7	79.70	89.90	50	0.723
C4.5 Karar Ağacı	22	7	59.39	100	34.17	0.742
AIRS	22	7	79.70	92.47	50	0.767
Fuzzy-AIRS	22	7	78.95	92.40	48.75	0.762
YSA	22	7	75.70	89.90	50	0.723

Tablo 6.42 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	61	17	98.25	96.77	100	0.981
C4.5 Karar Ağacı	61	17	96.49	96.67	96.30	0.964
AIRS	61	17	98.25	96.77	100	0.981
Fuzzy-AIRS	61	17	98.25	96.77	100	0.981
YSA	61	17	98.25	96.77	100	0.981

Tablo 6.43 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	63	20	81.25	77.50	87.50	0.805
C4.5 Karar Ağacı	63	20	73.68	83.87	75.76	0.784
AIRS	63	20	79.69	76.92	84	0.791
Fuzzy-AIRS	63	20	78.31	85.71	72.23	0.786
YSA	63	20	78.12	76.31	80.76	0.776

Tablo 6.44 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	63	18	84.90	85.71	84	0.848
C4.5 Karar Ağacı	63	18	84.90	85.71	84	0.848
AIRS	63	18	75.47	70.27	87.50	0.744
Fuzzy-AIRS	63	18	75.47	74.19	77.28	0.750
YSA	63	18	94.33	96.30	92.30	0.944

Yukarıda elde edilen sonuçlara göre *f-skor özellik seçme* ile sınıflandırıcı birleşimi hibrid sistem, kalp hastalığı, E.coli promoter gen dizileri, Doppler sinyali ile damar sertliği, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile macular hastalığı veri kümelerini sınflamada ön-işlemesiz elde edilen sonuçlara göre üstün sonuçlar elde etmiştir.

Lineer kernel f-skor özellik seçme kullanılarak medikal veri kümelerini sınıflamada elde edilen sonuçlar Tablo 6.45 ile Tablo 6.50 arasında verilmiştir. Tablo 6.45 kalp hastalığı, Tablo 6.46 E.coli promoter gen dizileri, Tablo 6.47 SPECT görüntüleri, Tablo 6.48 Doppler sinyali ile damar sertliği, Tablo 6.49 VEP sinyali ile optik sinir hastalığı, Tablo 6.50 ise PERG sinyali ile Macular hastalığı veri kümesi için *Lineer kernel f-skor özellik seçme* ile sınıflandırıcıların birleşiminin elde ettiği sonuçları göstermektedir.

Tablo 6.45 Kalp hastalığı veri kümesinin sınıflandırılmasında, Lineer kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***Lineer Kernel ile oluşturulan uzay boyutu***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	13	270	128	HAFIZA YETERSİZ (SONUÇ YOK)			
C4.5 Karar Ağacı	13	270	128	85.18	83.34	86.67	0.841
AIRS	13	270	128	76.30	71.87	80.28	0.763
Fuzzy-AIRS	13	270	128	78.52	89.47	73.95	0.792
YSA	13	270	128	80.74	78.95	85.50	0.811

Tablo 6.46 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, Lineer kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***Lineer Kernel ile oluşturulan uzay boyutu***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	57	106	40	51.92	51.61	52.38	0.519
C4.5 Karar Ağacı	57	106	40	76.92	76.92	76.92	0.769
AIRS	57	106	40	59.62	69.23	56.41	0.596
Fuzzy-AIRS	57	106	40	55.77	80	53.19	0.557
YSA	57	106	40	63.46	65.21	62.06	0.634

Tablo 6.47 SPECT görüntüleri veri kümesinin sınıflandırılmasında, Lineer kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***Lineer Kernel ile oluşturulan uzay boyutu***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	22	267	148	HAFIZA YETERSİZ (SONUÇ YOK)			
C4.5 Karar Ağacı	22	267	148	66.16	100	37.50	0.787
AIRS	22	267	148	70.68	92.40	38.89	0.733
Fuzzy-AIRS	22	267	148	72.18	90.58	39.58	0.715
YSA	22	267	148	79.69	88.34	50	0.707

Tablo 6.48 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, Lineer kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***Lineer Kernel ile oluşturulan uzay boyutu***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	61	114	57	98.25	96.77	100	0.981
C4.5 Karar Ağacı	61	114	57	100	100	100	1.00
AIRS	61	114	57	98.25	96.77	100	0.981
Fuzzy-AIRS	61	114	57	98.25	96.77	100	0.981
YSA	61	114	57	96.49	96.67	96.30	0.964

Tablo 6.49 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, Lineer kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***Lineer Kernel ile oluşturulan uzay boyutu u***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	63	129	62	68.75	70.58	66.67	0.686
C4.5 Karar Ağacı	63	129	62	40.62	43.75	37.50	0.405
AIRS	63	129	62	57.81	68.42	53.34	0.591
Fuzzy-AIRS	63	129	62	51.56	52.38	0.0	0.485
YSA	63	129	62	73.43	54.28	72.41	0.732

Tablo 6.50 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, Lineer kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***Lineer Kernel ile oluşturulan uzay boyutu utu***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	63	106	49	75.47	74.19	77.27	0.762
C4.5 Karar Ağacı	63	106	49	11.32	19.35	0.0	0.114
AIRS	63	106	49	50.94	57.14	48.71	0.51
Fuzzy-AIRS	63	106	49	50.94	57.14	48.71	0.51
YSA	63	106	49	94.33	96.29	92.30	0.951

Yukarıda elde edilen sonuçlara göre *Lineer kernel f-skor özellik seçme* ile sınıflandırıcı birleşimi hibrid sistem, kalp hastalığı, E.coli promoter gen dizileri ve Doppler sinyali ile damar sertliği veri kümelerini sınflamada ön-işlemesiz elde edilen sonuçlara göre üstün sonuçlar elde etmiştir.

RBF (Radial Basis Function) kernel f-skor özellik seçme kullanılarak medikal veri kümelerini sınıflamada elde edilen sonuçlar Tablo 6.51 ile Tablo 6.56 arasında verilmiştir. Tablo 6.51 kalp hastalığı, Tablo 6.52 E.coli promoter gen dizileri, Tablo 6.53 SPECT görüntüleri, Tablo 6.54 Doppler sinyali ile damar sertliği, Tablo 6.55 VEP sinyali ile optik sinir hastalığı, Tablo 6.56 ise PERG sinyali ile Macular hastalığı veri kümesi için *RBF kernel f-skor özellik seçme* ile sınıflandırıcıların birleşiminin elde ettiği sonuçları göstermektedir.

Tablo 6.51 Kalp hastalığı veri kümesinin sınıflandırılmasında, RBF kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***RBFKernel ile oluşturulan uzay boyutu***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	13	270	117	HAFIZA YETERSİZ (SONUÇ YOK)			
C4.5 Karar Ağacı	13	270	117	74.81	69.69	79.71	0.75
AIRS	13	270	117	76.30	70.59	82.08	0.766
Fuzzy-AIRS	13	270	117	80	77.04	82.43	0.798
YSA	13	270	117	76.30	71.21	81.15	0.765

Tablo 6.52 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, RBF kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***RBFKernel ile oluşturulan uzay boyutu***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	57	106	39	50	50	54.54	0.50
C4.5 Karar Ağacı	57	106	39	67.30	66.67	68	0.673
AIRS	57	106	39	59.62	69.23	56.41	0.596
Fuzzy-AIRS	57	106	39	55.77	100	53.06	0.557
YSA	57	106	39	63.46	60	70.59	0.634

Tablo 6.53 SPECT görüntüleri veri kümesinin sınıflandırılmasında, RBF kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***RBFKernel ile oluşturulan uzay boyutu***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	22	267	118	HAFIZA YETERSİZ (SONUÇ YOK)			
C4.5 Karar Ağacı	22	267	118	99.24	99.06	100	0.981
AIRS	22	267	118	66.17	88.60	33.34	0.663
Fuzzy-AIRS	22	267	118	64.66	91.54	33.87	0.695
YSA	22	267	118	81.20	89.52	53.34	0.730

Tablo 6.54 Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında, RBF kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***RBFKernel ile oluşturulan uzay boyutu***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	61	114	52	98.25	96.77	100	0.981
C4.5 Karar Ağacı	61	114	52	100	100	100	1.00
AIRS	61	114	52	98.25	96.77	100	0.981
Fuzzy-AIRS	61	114	52	98.25	96.77	100	0.981
YSA	61	114	52	98.25	96.77	100	0.981

Tablo 6.55 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, RBF kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***RBFKernel ile oluşturulan uzay boyutu***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	63	129	45	70.31	71.42	68.96	0.70
C4.5 Karar Ağacı	63	129	45	64.06	68.96	60	0.644
AIRS	63	129	45	75	78.12	71.87	0.75
Fuzzy-AIRS	63	129	45	56.25	71.42	52	0.563
YSA	63	129	45	85.93	90.32	81.81	0.861

Tablo 6.56 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, RBF kernel f-skor özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***Orijinal Özellik Sayısı***	***RBFKernel ile oluşturulan uzay boyutu***	***F-Skor ile seçilen özellik sayısı***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	63	106	39	69.81	65.79	80	0.686
C4.5 Karar Ağacı	63	106	39	73.58	75	72	0.735
AIRS	63	106	39	69.81	75.87	76.19	0.730
Fuzzy-AIRS	63	106	39	75.47	70.27	87.50	0.744
YSA	63	106	39	90.56	92.59	88.46	0.906

Yukarıda elde edilen sonuçlara göre *RBF kernel f-skor özellik seçme* ile sınıflandırıcı birleşimi hibrid sistem, kalp hastalığı, E.coli promoter gen dizileri, Doppler sinyali ile damar sertliği ve PERG sinyali ile macular hastalığı veri kümelerini sınflamada ön-işlemesiz elde edilen sonuçlara göre üstün sonuçlar elde etmiştir.

6.3.3. Bilgi kazancına dayanan özellik seçme yönteminin sonuçları

Bu özellik seçme yöntemi, veri kümesinde bulunan özelliklerinin çıkış bilgisine göre bilgi kazançlarının hesaplanarak önemli özelliklerin (yüksek bilgi kazancına sahip) seçilmesi, ilgisiz veya fazla özelliklerin ise veri kümesinden ayıklanmasına dayanır. Bu sayede hem işlem maliyeti azalacak hem de yüksek bir sınıflama performası elde edilebilecektir. Bilgi kazancını hesaplamadan önce veri kümesi C4.5 ayrıcı ile ayrık değerli bir hale dönüştürülür ve her bir özelliğin bilgi kazancı hesaplanır. Hesaplanan bilgi kazançlarının ortalaması eşik değer olarak seçilir. Özelliğin bilgi kazancı bu eşik değerden yüksekse seçilir, değilse özellik uzayından uzaklaştırılır.

Tablo 6.57 kalp hastalığı, Tablo 6.58 E.coli promoter gen dizileri, Tablo 6.59 SPECT görüntüleri, Tablo 6.60 VEP sinyali ile optik sinir hastalığı, Tablo 6.61 ise PERG sinyali ile Macular hastalığı veri kümesi için *bilgi kazancına dayanan özellik seçme* ile sınıflandırıcıların birleşiminin elde ettiği sonuçları göstermektedir.

Tablo 6.57 Kalp hastalığı veri kümesinin sınıflandırılmasında, bilgi kazancına dayanan özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***BK' dayanarak seçilen özellik sayısı ve seçilen özellikler***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	(9)-A1, A2, A3, A8, A9, A10, A11, A12, A13	71.12	64	80	0.72
C4.5 Karar Ağacı	(9)-A1, A2, A3, A8, A9, A10, A11, A12, A13	79.25	76.67	81.34	0.79
AIRS	(9)-A1, A2, A3, A8, A9, A10, A11, A12, A13	66.67	65.95	67.04	0.651
Fuzzy-AIRS	(9)-A1, A2, A3, A8, A9, A10, A11, A12, A13	65.93	71.87	64.07	0.631
YSA	(9)-A1, A2, A3, A8, A9, A10, A11, A12, A13	78.51	76.27	80.26	0.781

Bilgi Kazancı tabanlı özellik seçme algoritması ile 13 özelliğe sahip olan kalp hastalığı veri kümesinden ilgisiz veya fazla özellikler arındırılarak özellik sayısı 9' a indirilir.

Tablo 6.58 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında, bilgi kazancına dayanan özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***BK' dayanarak seçilen özellik sayısı ve seçilen özellikler***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	(4)- A6, A15, A16, A17	59.61	61.90	58.06	0.596
C4.5 Karar Ağacı	(4)- A6, A15, A16, A17	84.61	84.61	84.61	0.846
AIRS	(4)- A6, A15, A16, A17	82.69	79.31	86.95	0.826
Fuzzy-AIRS	(4)- A6, A15, A16, A17	82.69	81.48	84	0.826
YSA	(4)- A6, A15, A16, A17	78.84	77.78	80	0.788

Bilgi Kazancı tabanlı özellik seçme algoritması ile 57 özelliğe sahip olan promoter gen dizisi veri kümesinden ilgisiz veya fazla özellikler arındırılarak özellik sayısı 4' e indirilir.

Tablo 6.59 SPECT görüntüleri veri kümesinin sınıflandırılmasında, bilgi kazancına dayanan özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***BK' dayanarak seçilen özellik sayısı ve seçilen özellikler***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	(8)-A8, A13, A16, A17, A18, A20, A21, A22	80.45	90	51.51	0.739
C4.5 Karar Ağacı	(8)-A8, A13, A16, A17, A18, A20, A21, A22	78.94	86.79	48.14	0.674
AIRS	(8)-A8, A13, A16, A17, A18, A20, A21, A22	78.20	92.30	47.61	0.766
Fuzzy-AIRS	(8)-A8, A13, A16, A17, A18, A20, A21, A22	79.70	90.72	50	0.748
YSA	(8)-A8, A13, A16, A17, A18, A20, A21, A22	80.45	90	51.51	0.739

Bilgi Kazancı tabanlı özellik seçme algoritması ile 22 binary (ikili) özelliğe sahip olan SPECT görüntüleri kullanılarak kalp hastalığı veri kümesinden ilgisiz veya fazla özellikler arındırılarak özellik sayısı 8' e indirilir.

Bilgi Kazancı tabanlı özellik seçme algoritması ile 61 sürekli özelliğe sahip olan Doppler sinyali ile damar sertliği hastalığı teşhisi veri kümesinde ilgisiz veya fazla özellikler olmadığı veya bütün özelliklerin bilgi kazançları, özellik seçmede kullanılan bilgi kazancı eşiğinden büyük olduğu için bütün özelliklerin seçilme yoluna gidilmiştir.

Tablo 6.60 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında, bilgi kazancına dayanan özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***BK' dayanarak seçilen özellik sayısı ve seçilen özellikler***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	(6)-A18, A21, A24, A28, A29, A43	64.06	67.74	60.60	0.642
C4.5 Karar Ağacı	(6)-A18, A21, A24, A28, A29, A43	82.81	84.84	80.64	0.828
AIRS	(6)-A18, A21, A24, A28, A29, A43	87.50	84.21	92.30	0.870
Fuzzy-AIRS	(6)-A18, A21, A24, A28, A29, A43	89.06	84.61	96	0.885
YSA	(6)-A18, A21, A24, A28, A29, A43	85.93	83.78	88.89	0.855

Bilgi Kazancı tabanlı özellik seçme algoritması ile 63 sürekli özelliğe sahip olan VEP sinyalleri kullanılarak optik sinir hastalığı teşhisi veri kümesinden ilgisiz veya fazla özellikler arındırılarak özellik sayısı 6' ya indirilir.

Tablo 6.61 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında, bilgi kazancına dayanan özellik seçme ile sınıflandırıcıların birleşiminden elde edilen sonuçlar

Sınıflandırıcı	***BK' dayanarak seçilen özellik sayısı ve seçilen özellikler***	***Sınıflama Doğruluğu (%)***	***Duyarlılık (%)***	***Seçicilik (%)***	***ROC eğrisi altında kalan alan***
ANFIS	(6)-A16, A17, A18, A29, A31, A52	83.01	85.18	80.76	0.844
C4.5 Karar Ağacı	(6)-A16, A17, A18, A29, A31, A52	92.45	92.85	92	0.924
AIRS	(6)-A16, A17, A18, A29, A31, A52	94.34	96.30	92.30	0.944
Fuzzy-AIRS	(6)-A16, A17, A18, A29, A31, A52	86.79	92	82.14	0.870
YSA	(6)-A16, A17, A18, A29, A31, A52	84.90	81.25	90.47	0.830

Bilgi Kazancı tabanlı özellik seçme algoritması ile 63 sürekli özelliğe sahip olan PERG sinyalleri kullanılarak Macular (sarı nokta) hastalığı teşhisi veri kümesinden ilgisiz veya fazla özellikler arındırılarak özellik sayısı 6'ya indirilir.

Yukarıda elde edilen sonuçlara göre *Bilgi Kazancı tabanlı özellik seçme* ile sınıflandırıcı birleşimi hibrid sistem, kalp hastalığı, E.coli promoter gen dizileri, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile macular hastalığı veri kümelerini sınflamada ön-işlemesiz elde edilen sonuçlara göre üstün sonuçlar elde etmiştir.

6.4. Ön-işleme ve Özellik Seçme Yöntemleri Birleştirilerek Oluşturulan Hibrid Sistemler

Özellik seçme, veri ağırlıklandırma ön-işleme ve sınıflama algoritmaları birleştirilerek yeni hibrid sistemler oluşturulmuş ve tezde kullanılan altı medikal veri kümesine ilk kez uygulanmıştır. Daha sonra her bir medikal veri küme için en iyi hibrid sistem seçilmiştir. Kullanılan hibrid sistemlerin listesi Tablo 6.62' de verilmiştir. Medikal veri kümeleri için en iyi hibrid sistemi gösteren liste de Tablo

6.63’ de görülmektedir. Tablo 6.64’ den Tablo 6.69’a kadar olanlar her bir medikal veri kümesi için önerilen hibrid sistemlerin sınıflama sonuçlarını göstermektedir. Tablo 6.64 kalp hastalığı, Tablo 6.65 E.coli promoter gen dizileri, Tablo 6.66 SPECT görüntüleri, Tablo 6.67 Doppler sinyali ile damar sertliği, Tablo 6.68 VEP sinyali ile optik sinir hastalığı, Tablo 6.69 ise PERG sinyali ile Macular hastalığı veri kümesi için önerilen hibrid sistemlerin elde ettiği sınıflama sonuçlarını göstermektedir.

Tablo 6.62 Önerilen hibrid sistemler

Önerilen Hibrid Sistemler
TBA, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem
TBA, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem
TBA, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem
TBA, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem
BKÖS, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem
BKÖS, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem
BKÖS, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem
BKÖS, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem
Kernel F-skor Özellik Seçme yöntemi (KFSÖS) (RBF Kernel fonksiyonu), Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem
KFSÖS (RBF Kernel fonksiyonu), k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem
KFSÖS (RBF Kernel fonksiyonu), Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem
KFSÖS (RBF Kernel fonksiyonu), Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem

Tablo 6.63 Medikal veri kümelerini sınıflandırılmasında en iyi sonucu elde eden hibrid sistemler

Medikal Veri Kümesi	***Hibrid Sistem***
Kalp hastalığı veri kümesinin sınıflandırılmasında	BKÖS, Genelleştirilmiş Ayrışım Analizi ve YSA' nın birleşimi hibrid bir sistem
E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında	BKÖS, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve C4.5 karar ağacı veya Fuzzy-AIRS sınıflandırıcılarından birinin kullanımı ile oluşan hibrid bir sistem
SPECT görüntüleri veri kümesinin sınıflandırılmasında	TBA, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve YSA' nın birleşimi hibrid bir sistem
Doppler sinyali ile damar sertliği veri kümesinin sınıflandırılmasında	TBA, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve ANFIS veya C4.5 karar ağacı veya YSA sınıflandırıcılarından birinin kullanımı ile oluşan hibrid bir sistem
VEP sinyali optik sinir hastalığı veri kümesinin sınıflandırılmasında	KFSÖS (RBF Kernel fonksiyonu), Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve C4.5 karar ağacı veya YSA sınıflandırıcılarından birinin kullanımı ile oluşan hibrid bir sistem
PERG sinyali macular hastalığı veri kümesinin sınıflandırılmasında	TBA, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve AIRSveya Fuzzy-AIRS sınıflandırıcılarından birinin kullanımı ile oluşan hibrid bir sistem

Tablo 6.64 Kalp hastalığı veri kümesinin sınıflandırılmasında önerilen hibrid sistemlerden elde edilen sınıflama doğrulukları

Önerilen Hibrid Sistem	***Sınıflama Doğruluğu (%)***
TBA, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 72.59 C4.5: 77.03 AIRS: 79.25 Fuzzy-AIRS: 74.07 YSA: 71.85
TBA, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 77.78 C4.5: 77.78 AIRS: 84.44 Fuzzy-AIRS: 82.96 YSA: 82.96
TBA, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 80.74 C4.5: 77.78 AIRS: 81.48 Fuzzy-AIRS: 77.78 YSA: 81.48
TBA, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 98.51 C4.5: 99.25 AIRS: 77.78 Fuzzy-AIRS: 74.81 YSA: 100
BKÖS, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 77.03 C4.5: 100 AIRS: 75.56 Fuzzy-AIRS: 75.56 YSA: 80.74
BKÖS, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 70.37 C4.5: 76.29 AIRS: 65.92 Fuzzy-AIRS: 64.44 YSA: 77.78
***BKÖS, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem**	ANFIS: 97.03 C4.5: 98.51 AIRS: 98.51 Fuzzy-AIRS: 95.56 YSA: 100
BKÖS, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 100 C4.5: 99.25 AIRS: 55.56 Fuzzy-AIRS: 55.56 YSA: 100
Kernel F-skor Özellik Seçme yöntemi (KFSÖS) (RBF Kernel fonksiyonu), Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: Yetersiz hafıza C4.5: 100 AIRS: 77.03 Fuzzy-AIRS: 76.30 YSA: 80
KFSÖS (RBF Kernel fonksiyonu), k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: Yetersiz hafıza C4.5: 100 AIRS: 75.56 Fuzzy-AIRS: 70.37 YSA: 77.03
KFSÖS (RBF Kernel fonksiyonu), Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: Yetersiz hafıza C4.5: 72.55 AIRS: 77.03 Fuzzy-AIRS: 76.30 YSA: 75.56
KFSÖS (RBF Kernel fonksiyonu), Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: Yetersiz hafıza C4.5: 98.51 AIRS: 99.25 Fuzzy-AIRS: 99.25 YSA: 99.25

* en iyi hibrid sistemi göstermektedir.

Tablo 6.65 E.coli promoter gen dizileri veri kümesinin sınıflandırılmasında önerilen hibrid sistemlerden elde edilen sınıflama doğrulukları

Önerilen Hibrid Sistem	***Sınıflama Doğruluğu (%)***
TBA, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 80.76 C4.5: 57.69 AIRS: 80.74 Fuzzy-AIRS: 82.69 YSA: 90.38
TBA, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 78.84 C4.5: 67.30 AIRS: 82.69 Fuzzy-AIRS: 82.69 YSA: 75
TBA, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 78.84 C4.5: 61.53 AIRS: 78.84 Fuzzy-AIRS: 78.85 YSA: 75
TBA, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 98.07 C4.5: 98.07 AIRS: 98.07 Fuzzy-AIRS: 98.08 YSA: 96.15
BKÖS, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 71.15 C4.5: 67.30 AIRS: 90.38 Fuzzy-AIRS: 90.38 YSA: 84.61
BKÖS, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 61.53 C4.5: 73.07 AIRS: 84.61 Fuzzy-AIRS: 84.62 YSA: 86.53
BKÖS, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 59.61 C4.5: 73.07 AIRS: 82.69 Fuzzy-AIRS: 84.62 YSA: 80.76
***BKÖS, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem**	ANFIS: 98.07 C4.5: 100 AIRS: 98.07 Fuzzy-AIRS: 100 YSA: 98.07
Kernel F-skor Özellik Seçme yöntemi (KFSÖS) (RBF Kernel fonksiyonu), Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 65.38 C4.5: 21.15 AIRS: 69.23 Fuzzy-AIRS: 65.38 YSA: 61.53
KFSÖS (RBF Kernel fonksiyonu), k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 61.53 C4.5: 32.69 AIRS: 98.07 Fuzzy-AIRS: 98.08 YSA: 71.15
KFSÖS (RBF Kernel fonksiyonu), Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 50 C4.5: 36.53 AIRS: 51.92 Fuzzy-AIRS: 51.92 YSA: 63.46
KFSÖS (RBF Kernel fonksiyonu), Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 98.08 C4.5: 98.07 AIRS: 98.07 Fuzzy-AIRS: 98.08 YSA: 98.08

* en iyi hibrid sistemi göstermektedir.

Tablo 6.66 SPECT görüntüleri veri kümesinin sınıflandırılmasında önerilen hibrid sistemlerden elde edilen sınıflama doğrulukları

Önerilen Hibrid Sistem	***Sınıflama Doğruluğu (%)***
TBA, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 87.96 C4.5: 75.18 AIRS: 90.22 Fuzzy-AIRS: 91.73 YSA: 91.72
***TBA, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem**	ANFIS: 93.23 C4.5: 87.96 AIRS: 93.23 Fuzzy-AIRS: 91.73 YSA: 93.98
TBA, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 95.48 C4.5: 74.43 AIRS: 90.97 Fuzzy-AIRS: 93.23 YSA: 91.72
TBA, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 98.49 C4.5: 78.94 AIRS: 20.30 Fuzzy-AIRS: 20.30 YSA: 99.25
BKÖS, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 78.94 C4.5: 61.65 AIRS: 76.69 Fuzzy-AIRS: 78.20 YSA: 80.45
BKÖS, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 75.93 C4.5: 61.65 AIRS: 77.44 Fuzzy-AIRS: 78.20 YSA: 79.69
BKÖS, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 75.93 C4.5: 61.65 AIRS: 78.19 Fuzzy-AIRS: 78.20 YSA: 79.69
Kernel F-skor Özellik Seçme yöntemi (KFSÖS) (RBF Kernel fonksiyonu), Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: Yetersiz hafıza C4.5: 67.66 AIRS: 20.30 Fuzzy-AIRS: 20.30 YSA: 73.68
KFSÖS (RBF Kernel fonksiyonu), k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: Yetersiz hafıza C4.5: 71.42 AIRS: 93.23 Fuzzy-AIRS: 93.23 YSA: 78.19
KFSÖS (RBF Kernel fonksiyonu), Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: Yetersiz hafıza C4.5: 71.42 AIRS: 65.41 Fuzzy-AIRS: 73.68 YSA: 79.69
KFSÖS (RBF Kernel fonksiyonu), Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: Yetersiz hafıza C4.5: 78.94 AIRS: 94.73 Fuzzy-AIRS: 91.72 YSA: 99.24

* en iyi hibrid sistemi göstermektedir.

Tablo 6.67 Doppler siyali ile damar sertliği veri kümesinin sınıflandırılmasında önerilen hibrid sistemlerden elde edilen sınıflama doğrulukları

Önerilen Hibrid Sistem	***Sınıflama Doğruluğu (%)***
TBA, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 78.94 C4.5: 92.98 AIRS: 96.49 Fuzzy-AIRS: 96.49 YSA: 98.25
TBA, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 85.96 C4.5: 98.25 AIRS: 92.98 Fuzzy-AIRS: 92.98 YSA: 96.49
TBA, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 85.96 C4.5: 91.22 AIRS: 94.73 Fuzzy-AIRS: 94.74 YSA: 100
***TBA, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem**	ANFIS: 100 C4.5: 100 AIRS: 94.73 Fuzzy-AIRS: 94.74 YSA: 100
Kernel F-skor Özellik Seçme yöntemi (KFSÖS) (RBF Kernel fonksiyonu), Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 96.49 C4.5: 98.25 AIRS: 94.73 Fuzzy-AIRS: 94.74 YSA: 94.73
KFSÖS (RBF Kernel fonksiyonu), k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 98.25 C4.5: 92.98 AIRS: 94.73 Fuzzy-AIRS: 94.74 YSA: 98.25
KFSÖS (RBF Kernel fonksiyonu), Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 98.25 C4.5: 92.98 AIRS: 94.73 Fuzzy-AIRS: 94.74 YSA: 98.25
KFSÖS (RBF Kernel fonksiyonu), Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 96.49 C4.5: 92.98 AIRS: 94.73 Fuzzy-AIRS: 94.74 YSA: 98.25

* en iyi hibrid sistemi göstermektedir.

Tablo 6.68 VEP sinyali ile optik sinir hastalığı veri kümesinin sınıflandırılmasında önerilen hibrid sistemlerden elde edilen sınıflama doğrulukları

Önerilen Hibrid Sistem	***Sınıflama Doğruluğu (%)***
TBA, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 78.12 C4.5: 100 AIRS: 68.75 Fuzzy-AIRS: 64.06 YSA: 73.43
TBA, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 65 C4.5: 82.81 AIRS: 73.43 Fuzzy-AIRS: 87.50 YSA: 76.56
TBA, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 73.43 C4.5: 70.31 AIRS: 75 Fuzzy-AIRS: 64.06 YSA: 75
TBA, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 96.87 C4.5: 98.43 AIRS: 75 Fuzzy-AIRS: 56.25 YSA: 98.43
BKÖS, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 53.51 C4.5: 12.50 AIRS: 76.56 Fuzzy-AIRS: 71.87 YSA: -
BKÖS, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 71.87 C4.5: 81.25 AIRS: 68.75 Fuzzy-AIRS: 70.31 YSA: 79.68
BKÖS, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 64.06 C4.5: 75 AIRS: 75 Fuzzy-AIRS: 71.87 YSA: 87.50
BKÖS, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 96.87 C4.5: 96.87 AIRS: 68.75 Fuzzy-AIRS: 70.31 YSA: 98.43
Kernel F-skor Özellik Seçme yöntemi (KFSÖS) (RBF Kernel fonksiyonu), Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 53.31 C4.5: 12.50 AIRS: 75 Fuzzy-AIRS: 71.87 YSA: 57.81
KFSÖS (RBF Kernel fonksiyonu), k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 73.43 C4.5: 57.81 AIRS: 75 Fuzzy-AIRS: 71.87 YSA: 89.06
KFSÖS (RBF Kernel fonksiyonu), Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 71.87 C4.5: 56.25 AIRS: 75 Fuzzy-AIRS: 64.06 YSA: 79.68
****KFSÖS (RBF Kernel fonksiyonu), Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem***	ANFIS: 96.87 C4.5: 98.43 AIRS: 75 Fuzzy-AIRS: 71.87 YSA: 98.43

* en iyi hibrid sistemi göstermektedir.

Tablo 6.69 PERG sinyali ile Macular hastalığı veri kümesinin sınıflandırılmasında önerilen hibrid sistemlerden elde edilen sınıflama doğrulukları

Önerilen Hibrid Sistem	***Sınıflama Doğruluğu (%)***
TBA, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 49.05 C4.5: 13.20 AIRS: 58.49 Fuzzy-AIRS: 58.49 YSA: 50.94
TBA, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 92.45 C4.5: 92.45 AIRS: 83.01 Fuzzy-AIRS: 79.25 YSA: 98.11
TBA, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 88.67 C4.5: 86.79 AIRS: 92.45 Fuzzy-AIRS: 92.45 YSA: 96.22
***TBA, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem**	ANFIS: 98.11 C4.5: 98.11 AIRS:100 Fuzzy-AIRS: 100 YSA: 98.11
BKÖS, Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 62.26 C4.5: 35.84 AIRS: 60.37 Fuzzy-AIRS: 69.81 YSA: 66.03
BKÖS, k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 86.79 C4.5: 79.25 AIRS: 84.90 Fuzzy-AIRS: 81.13 YSA: 96.22
BKÖS, Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 77.35 C4.5: 79.25 AIRS: 90.56 Fuzzy-AIRS: 90.56 YSA: 92.45
BKÖS, Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 98.11 C4.5: 98.11 AIRS: 98.11 Fuzzy-AIRS: 98.11 YSA: 98.11
Kernel F-skor Özellik Seçme yöntemi (KFSÖS) (RBF Kernel fonksiyonu), Bulanık veri ağırlıklandırma ön-işleme (BVAÖ) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 49.05 C4.5: 13.20 AIRS: 47.17 Fuzzy-AIRS: 47.17 YSA: 50.94
KFSÖS (RBF Kernel fonksiyonu), k-NN tabanlı veri ağırlıklandırma ön-işleme (k=15) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 75.47 C4.5: 58.49 AIRS: 47.17 Fuzzy-AIRS: 47.17 YSA: 90.56
KFSÖS (RBF Kernel fonksiyonu), Genelleştirilmiş Ayrışım Analizi ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 62.26 C4.5: 60.37 AIRS: 49.05 Fuzzy-AIRS: 49.05 YSA: 90.56
KFSÖS (RBF Kernel fonksiyonu), Benzerlik tabanlı veri ağırlıklandırma ön-işleme (Üstel Benzerlik Katsayı Yöntemi) ve sınıflandırıcının birleşimi hibrid bir sistem	ANFIS: 98.11 C4.5: 98.11 AIRS: 98.11 Fuzzy-AIRS: 98.11 YSA: 98.11

* en iyi hibrid sistemi göstermektedir.

7. SONUÇLAR VE ÖNERİLER

Bu tez çalışmasında, biyomedikal veri kümelerini sınıflamada sınıflama algoritmasının performansını arttırmak için ön-işleme yöntemi olarak çeşitli veri ağırlıklandırma ve özellik seçme yöntemleri önerilmiş ve kullanılmıştır. Bu çalışmadaki diğer bir amaç da bu medikal verileri en iyi sınıflayan hibrid bir model oluşturmaktır. Bu bölümde, tezden elde edilen sonuçlar yorumlanmış ve gelecekte konu üzerinde çalışacaklar için bazı öneriler sunulmuştur.

7.1. Sonuçlar

Medikal veri kümeleri, eksik değerler, yanlış olarak girilmiş değerler, fazla boyutlu özellikler ve gürültü içerebilirler. Bu durum medikal verileri sınıflandırken sınıflama performansını negatif yönde etkilemektedir. Bunu önlemek için; veri normalizasyonu, veri temizleme, boyut azaltımı gibi veri ön-işleme yöntemlerinin kullanılması gerekmektedir. Bu tez çalışmasında veri ön-işleme yöntemleri iki şekilde incelenmiştir. İlki, veri ağırlıklandırma ve veri azaltma yöntemi, diğeri ise boyut azaltımı ve özellik seçme yöntemidir.

Önerilen veri ağırlıklandırma ve veri ön-işleme yöntemleri; *bulanık ağırlıklandırma ön-işleme*, k-*NN (k- en yakın komşu) tabanlı veri ağırlıklandırma ön-işleme*, *genelleştirilmiş ayrışım analizi* (*Kernel Fisher Ayrışım Analizi-Kernel LDA*) ve *benzerlik tabanlı veri ağırlıklandırma ön-işleme* yöntemleridir. Kullanılan ve önerilen özellik seçme yöntemleri ise *temel bileşen analizi*, *bilgi kazancına dayanan özellik seçme* algoritması ve *Kernel F-skor özellik seçme* yöntemleridir. Önerilen bu veri ön-işleme yöntemleri, University of Califoria Irvine (UCI) veritabanından alınan; Kalp hastalığı veri kümesi, *SPECT* (Single Photon Emission Computed Tomography-Tek Foton Yayılmalı Hesaplamalı Tomografi) görüntüleri ile kalp hastalığı veri kümesi ve Promoter gen dizileri veri kümeleri ve Fatih

Üniversitesi Öğretim Üyesi Prof. Dr. Sadık Kara ve Erciyes Üniversitesindeki ekibi tarafından alınan ve Atherosclerosis (damar sertliği) hastalığını sınıflamak için hasta ve normal kişeye ait Doppler sinyallerinden spektral analiz yöntemi ile özelliği çıkarılmış Atherosclerosis hastalığı veri kümesi, Macular (Sarı Nokta Hastalığı) hastalığını sınıflamak için yine hasta ve normal kişeye ait PERG (Pattern Electroretinography) sinyallerinden spektral analiz yöntemi ile özelliği çıkarılmış Macular hastalığı veri kümesi ve son olarak da Optik Sinir Hastalığını sınıflamak için hasta ve normal kişeye ait VEP (Visual Evoked Potentials) sinyallerinden spektral analiz yöntemi ile özelliği çıkarılmış olan Optik Sinir hastalığı veri kümelerine ait çalışmalar yapmışlardır.

Veri ön-işleme ile birlikte medikal verileri sınıflamada *ANFIS*, *C4.5 karar ağacı*, *AIRS*, *Fuzzy-AIRS* ve *YSA* sınıflama algoritmaları kullanılmıştır. Sınıflandırma performanslarını karşılaştırmak ve en iyi yöntemi seçmek için 5-kat çaprazlama, sınıflama doğruluğu, duyarlılık-seçicilik değerleri ve ROC eğrisi altında kalan değerler kullanılmıştır.

Bulanık mantık tabanlı veri ağırlıklandırma yöntemi, 6 medikal veri kümesi içinden; kalp hastalığı veri kümesi, E.coli promoter gen dizileri veri kümesi, *SPECT* görüntüleri ve VEP sinyali ile optik sinir hastalığı veri kümelerinde ön-işlemesiz sınıflama doğruluklarına göre yüksek doğruluklar elde etmiştir.

k-NN tabanlı veri ağırlıklandırma yönteminde, 6 medikal veri kümelerini sınıflamada *k*' nın 5, 10, 15 ve 20 değerleri kullanılmış ve her bir *k* değeri ayrı ayrı değerlendirilmiştir. *k*' nın 5 ve 20 değerleri için, kalp hastalığı, E. coli promoter gen dizileri, *SPECT* görüntüleri, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile Macular hastalığı veri kümelerinde ön-işlemesiz sınıflama doğrulukları dikkate alınırsa yüksek doğruluk oranları elde edilmiştir. *k*' nın 10 değeri için kalp hastalığı, E. coli promoter gen dizileri, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile Macular hastalığı veri kümelerinde ön-işlemesiz sınıflama doğrulukları dikkate alınırsa yüksek doğruluk oranları elde edilmiştir. *k*' nın 15 değeri için ise, kalp hastalığı, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile Macular hastalığı veri kümelerinde ön-işlemesiz sınıflama doğrulukları dikkate alınırsa yüksek doğruluk oranları elde edilmiştir.

Genelleştirilmiş ayrışım analizinde, 6 medikal veri kümelerini sınıflamada; kalp hastalığı, E.coli promoter gen dizileri, *SPECT* görüntüleri ve VEP sinyali ile optik sinir hastalığı veri kümelerinde ön-işlemesiz sınıflama doğrulukları dikkate alınırsa yüksek doğruluk oranları elde edilmiştir.

Benzerlik tabanlı veri ağırlıklandırma ön-işleme yönteminde, kosinüs genliği (cosine amplitude), minimum-maksimum metodu, mutlak üstel (absolute exponential) yöntemi ve üstel benzerlik katsayısı (exponential similarity coefficient) içeren dört çeşit benzerlik fonksiyonları kullanılmış ve veri birlikteliği ile birleştirilmiştir. Bu benzerlik fonksiyonlarının etkileri de ayrıca ayrı ayrı incelenmiştir. Kosinüs genliği benzerlik fonksiyonu veri ağırlıklandırma yöntemi dikkate alındığında, kalp hastalığı, E. coli promoter gen dizileri, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile Macular hastalığı veri kümelerinde ön-işlemesiz sınıflama doğrulukları dikkate alınırsa yüksek doğruluk oranları elde edilmiştir. Minimum-maksimum metodu benzerlik fonksiyonu veri ağırlıklandırma yöntemi dikkate alındığında, kalp hastalığı ve E. coli promoter gen dizileri veri kümelerinde ön-işlemesiz sınıflama doğrulukları dikkate alınırsa yüksek doğruluk oranları elde edilmiştir. Mutlak üstel (absolute exponential) benzerlik fonksiyonu veri ağırlıklandırma yönteminde ise, kalp hastalığı ve *SPECT* görüntüleri veri kümelerinde ön-işlemesiz sınıflama doğruluklarına göre yüksek doğrulukları elde etmiştir. Son olarak üstel benzerlik katsayısı benzerlik fonksiyonu veri ağırlıklandırma yönteminde, kalp hastalığı, E. coli promoter gen dizileri, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile Macular hastalığı veri kümelerinde ön-işlemesiz sınıflama doğrulukları dikkate alınırsa yüksek doğruluk oranları elde edilmiştir.

Temel bileşen analizi ile medikal veri kümelerinin boyutlarında azalma gerçekleşmiş ve bu sayede hem işlem maliyetinde düşüşler olmuş hem de sınıflama performansında artışlar kaydedilmiştir. Ayrıca kullanılan tüm medikal veri kümelerinde ön-işlemesiz sınıflama doğruluklarına göre yüksek oranda doğruluklar tespit edilmiştir.

Bilgi kazancı tabanlı özellik seçme yönteminde, 6 medikal veri kümesi içinden; kalp hastalığı, E. coli promoter gen dizileri, VEP sinyali ile optik sinir

hastalığı ve PERG sinyali ile Macular hastalığı veri kümelerinde ön-işlemesiz sınıflama doğrulukları dikkate alınırsa yüksek doğruluk oranları elde edilmiştir.

F-skor özellik seçme yöntemi ile sınıflandırıcının birleşimi hibrid sistem, 6 medikal veri kümesi içinden; kalp hastalığı, E. coli promoter gen dizileri, Doppler sinyali ile damar sertliği, VEP sinyali ile optik sinir hastalığı ve PERG sinyali ile Macular hastalığı veri kümelerinde ön-işlemesiz sınıflama doğrulukları dikkate alınırsa yüksek düzeyde doğruluk oranları elde edilmiştir.

Bu tez çalışmasında ilk kez önerilen *Lineer Kernel F-skor özellik seçme* yöntemi ile sınıflandırıcının birleşimi hibrid sistem, 6 medikal veri kümesini sınıflamada; kalp hastalığı, E. coli promoter gen dizileri ve Doppler sinyali ile damar sertliği veri kümelerinde ön-işlemesiz sınıflama doğrulukları dikkate alınırsa yüksek düzeyde doğruluk oranları elde edilmiştir.

Bu tez çalışmasında ilk kez önerilen *RBF Kernel F-skor özellik seçme* yöntemi ile sınıflandırıcının birleşimi hibrid sistem, 6 medikal veri kümesini sınıflamada; kalp hastalığı, E. coli promoter gen dizileri, Doppler sinyali ile damar sertliği ve PERG sinyali ile Macular hastalığı veri kümelerinde ön-işlemesiz sınıflama doğrulukları dikkate alınırsa yüksek düzeyde doğruluk oranları elde edilmiştir.

Ayrıca, özellik seçme, veri ağırlıklandırma ve sınıflandırma algoritmaları birleştirilerek 12 yeni hibrid model önerilmiş ve 6 medikal veri kümesine ilk kez uygulanmıştır. Her bir medikal veri kümesi için en iyi sonucu sağlayan hibrid model seçilmiştir. Kalp hastalığı veri kümesini sınıflamada en iyi hibrid model *bilgi kazancına dayanan özellik seçme*, *genelleştirilmiş ayrışım analizi* ve *YSA*' nın birleşimi hibrid bir model; E.coli promoter gen dizileri veri kümesini sınıflamada en iyi hibrid model *bilgi kazancına dayanan özellik seçme*, *benzerlik tabanlı veri ağırlıklandırma ön-işleme* ve *C4.5* karar ağacı veya *Fuzzy-AIRS* sınıflandırıcılarından birinin kullanımı ile oluşan hibrid bir model; SPECT görüntüleri veri kümesini sınıflamada en iyi hibrid model *temel bileşen analizi*, *k-NN tabanlı veri ağırlıklandırma ön-işleme* ve *YSA*' nın birleşimi hibrid bir model; Doppler sinyali ile damar sertliği veri kümesini sınıflamada en iyi hibrid model *temel bileşen analizi, benzerlik tabanlı veri ağırlıklandırma ön-işleme* ve *ANFIS* veya *C4.5* karar ağacı veya *YSA* sınıflandırıcılarından birinin kullanımı ile oluşan hibrid bir model; VEP

sinyali optik sinir hastalığı veri kümesini sınıflamada en iyi hibrid model *KFSÖS* (*RBF Kernel fonksiyonu*), *benzerlik tabanlı veri ağırlıklandırma ön-işleme* ve C4.5 karar ağacı veya YSA sınıflandırıcılarından birinin kullanımı ile oluşan hibrid bir model; PERG sinyali macular hastalığı veri kümesini sınıflamada en iyi hibrid model; *temel bileşen analizi, benzerlik tabanlı veri ağırlıklandırma ön-işleme* ve AIRSveya Fuzzy-AIRS sınıflandırıcılarından birinin kullanımı ile oluşan hibrid bir model olmuştur.

7.2. Öneriler

Burada kullanılan veri ön-işleme teknikleri gerçek hayatta elde edilmiş medikal veri kümelerine uygulanmış ve orijinal haline göre yüksek sınıflama performansı elde edilmiştir. Bu medikal karar verme mekanizmaları bu tez çalışmasında çevrimdışı (offline) çalışmaktadır. Ancak, *DSP* (*Digital Signal Processing*) kartları kullanılarak hastadan elde edilen laboratuvar sonuçları çevrimiçi (online) olarak analiz edilmek suretiyle hastalığın karar verme aşamasında, doktora yardımcı olabilecek olacak bir uzman sistemin tasarlanmasına kapı aralayabilir.

Temel bileşen analizi veya bir boyut azaltımı algoritması kullanma yoluna gidilerek, hastadan alınacak kan veya idrar numunelerinin laboratuvar sonuçlarına göre hastanın zamana göre iyileşip iyileşmediğini gösteren görsel bir uzman sistem gerçekleştirilebilir. Bu öneri, hastadan düzenli olarak veri alınması şartıyla yapılabilir.

Bu tez çalışmasında kullanılan özellik seçme yöntemlerinin dışında, çok sınıflı veri kümelerinde özellik seçme işlemini gerçekleştirirken, sınıf etiketlerini ikili gruplar halinde her bir özelliğin karşılıklı bilgi kazancını hesaplamak için ve içlerinden en yüksek kazancı sağlayan özelliklerin seçilmesi için kullanılabilir. Bu yöntemde, her bir ikili durum için en yüksek kazancı sağlayan özellikler seçilir ve içlerinden en çok tekrarlanan özellikler yeni özellik kümesine atılır. Önerilen özellik seçme işlemi genel olarak 3 veya daha fazla sınıfa sahip veri kümelerinde kullanılabilir.

Bulanık mantık veri ağırlıklandırma yöntemini Matlab' da bir *GUI* programı tasarlanabilir. Böylelikle kullanıcının bölütleme sayısını seçmesine, üyelik fonksiyon tipinin değiştirmesine izin veren bir sistem ortaya çıkarılabilir. Bu şekilde işlenecek veri kümesi için en iyi optimum yapı elde edilebilir.

Değişkenlerin (veya özellikler) kendi aralarındaki ve özellikler ile çıkışlar arasındaki korelasyon hesaplanarak, birbirleriyle ilişkili özelliklerin veri kümesinden çıkartılmasına yani özellikler ile çıkış arasında yüksek korelasyon sağlayan özelliklerin seçilmesine yönelik yeni bir özellik seçme yöntemi oluşturulabilir.

Sınıflama algoritmalarının performanslarını arttırmak için diğer bir yöntem de ön-işleme adımı olarak "*Kümele Yöntemleri*" ni kullanmaktır. İlk olarak veri kümesinde kaç sınıf varsa o sayıda veri kümesini kümeye bölmek daha sonra özelliklerin küme merkezlerine olan uzaklıklarına göre veri kümesini ağırlıklandırmaktır. Bu işlem ile özelliğin küme merkezine en yakın veriye yüksek değerli ağırlık, uzak olana da daha düşük ağırlık verilerek yeni bir özellik uzayı oluşturulması yoluna gidilir.

8. KAYNAKLAR

Akdemir, B., Polat, K., Güneş,S., 2007, "Prediction of E.Coli Promoter Gene Sequences Using a Hybrid Combination Based on Feature Selection, Fuzzy Weighted Pre-processing, and Decision Tree Classifier", KES (1) 2007: 125-131.

Asl, B.M., Setarehdan, S.K., Mohebbi, M., 2008, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal", Artificial Intelligence in Medicine, 44(1), 51-64.

Atilla, H. et all., 2006, "Pattern elektroretinography and visual evoked potentials in optic nerve diseases", Journal of Clinical Neuroscience, 13:55-59.

Bakırcı, Ü., Yıldırım, T., 2004, "Diagnosis of Cardiac Problems From SPECT Images by Feedforward Networks", SİU 2004, IEEE 12. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, sf. 103 –105, Kuşadası.

Baudat,G., Anouar, F., 2000, "Generalized Discriminant Analysis Using a Kernel Approach", Neural Computation 12(10): 2385-2404.

Beale, R., & Jackson, T. 1990, Neural computing: An introduction. Bristol, UK: Institute of Physics Publishing.

Bin Cao, Dou Shen, Jian-Tao Sun, Qiang Yang, Zheng Chen, 2007, "Feature selection in a kernel space", International Conference on Machine Learning (ICML) Oregon USA, June 20-24 2007: 121-128.

Ceylan, M., Ceylan, R., Dirgenali, F., Kara, Özbay, Y., 2007, "Classification of carotid artery Doppler signals in the early phase of atherosclerosis using complex-valued artificial neural network", Computers in Biology and Medicine, 37(1): 28-36.

Chamlin, M., Davidoff, L., Feiring, E., 1955, "Ophthalmologic changes produced by pituitary tumors", Am J Ophthalmol, 40: 353-359.

Cios, K.J. Goodenday, L.S. Shah, K.K. Serpen, G., "A novel algorithm for classification of SPECT images of a human heart", Proceedings Ninth IEEE Symposium on Computer-Based Medical Systems, 1996, 1– 5.

Dart, A.M., Lacombe F., Yeoh J.K., Cameron J.D., Jennings G.L., Laufer E., Esmore D.S., 1991, "Aortic Distensibility in Patients with Isolated Hyper-Cholesterolaemia, Coronary Artery Disease, or Cardiac Transplant", Lancet, 270 –273.

De Castro, L.N., Von Zuben F.J., 2001, "Ainet: An Artificial Immune Network for Data Analysis" (Full version, pre-print). Book Chapter in Data Mining: A Heuristic Approach, Hussein A. Abbass, Ruhul A. Sarker, and Charles S. Newton (Eds.), Idea Group Publishing, USA.

Dirgenali, F., 2003, "Karotid Atar Damar Doppler Sinyallerinden Damar Sertliğinin Erken Dönemde Teşhisi", Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.

Dirgenali, F., Kara, S., 2006, "Recognition of early phase of atherosclerosis using principles component analysis and artificial neural networks from carotid artery Doppler signals", Expert Systems with Applications, 31(3): 643-651.

El-Hanjouri, M., Alkhaldi, W., Hamdy, N., Alim, O.A., "Heart diseases diagnosis using HMM", MELECON 2002. 11th Mediterranean on Electrotechnical Conference, 2002, 489 – 492.

Evans, D., 2000, "Doppler Signal Analysis", Ultrasound in Med. & Biol., Vol. 26(1): 13–15.

Evans, D.H., McDicken W.N., Skidmore R., Woodcock J.P., 1989, "Doppler Ultrasound"

Fawcett, F.,. 2005, "An Introduction to ROC analysis", Pattern Recognition Letters, 27(8): 861-874.

Geoffrey, G.T., Jude, W.S., and Michiel, O.N.,1990, "Refinement of Approximate Domain Theories by Knowledge-Based Neural Networks", Proc. of the eight national conf. on Artificial Intelligence, 861-866.

Güler, İ., Übeyli, E.D., 2005, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients", Journal of Neuroscience Methods, 148(2): 13-121.

Güler, I., Übeyli , E. D., 2003, "Detection of ophthalmic artery stenosis by least-mean squares backpropagation neural network", Computers in Biology and Medicine, 33, 333–343.

Güven, A., 2005, "Göze Ait Elektrofizyolojik Sinyaller Kullanılarak Yapay Sinir Ağları Destekli Bazı Göz Hastalıklarının Teşhisi", Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi.

Güven, A., Polat, K., Kara, S., Güneş, S., 2008, "The effect of generalized discriminate analysis (GDA) to the classification of optic nerve disease from VEP signals", Computers in Biology and Medicine, 38(1):62-68.

Güven, A., Kara, S., 2006, "Diagnosis of the Macular diseases from pattern electroretinography signals using artificial neural networks", Expert Systems with Applications, 30(2): 361-366.

Haibo, G., Wenxue H., Jianxin C, Xu Yonghong, 2007, "Optimization of Principal Component Analysis in Feature Extraction" , International Conference on Mechatronics and Automation, ICMA 2007, Harbin, 5-8 Aug. 2007, 3128– 3132.

Harley, C. and Reynolds, R., 1987, "Analysis of E.coli promoters sequences", Nucl. Acids Res., 15:2343-2361.

Haykin, S. (1994). Neural networks: A comprehensive foundation. New York: Macmillan.

Heckenlively, JR., Arden GB., 1991, Princeples and Practice of Clinical Electrophisiology of Vision, Mosby Year Book, USA.

Hirai, T., Sasayama S., Kawasaki T., Yagi S., 1989, "Stiffness of Systemic Arteries in Patients with Myocardial Infarction, A Noninvasive Method to Predict Severity of Coronary Atherosclerosis", Circulation, 78 –86.

Holder, G.E., 1987, "Significance of abnormal pattern elektroretinography in anterior visual pathway dysfunction", British Journal of Ophthalmology, 71: 166-171.

Isabelle, G., André, E., 2003, "An Introduction to Variable and Feature Selection", JMLR Special Issue on Variable and Feature Selection, 3(Mar):1157-1182.

Ifeachor, E.C., Jervis B.W.2002. Digital Signal Processing-A Practical Approach, 2nd ed. Pearson Education Limited USA.

Jancic, S., J., Kostic, V., Carelli, V., 2007, "Visual evoked potentials (VEP) and brainstem evoked potentials (BAEP) as diagnostic procedures in Leber's hereditary optic neuropathy", Clinical Neurophysiology, 118(4):119.

Jang, J-SR, 1992, "Self-learning fuzzy controllers based on temporal backpropagation", IEEE Transactions on Neural Networks, 3(5):714-723.

Jang, J-SR., 1993, "ANFIS: adaptive-network-based fuzzy inference system", IEEE Transactions on Systems Man and Cybernetics, 23(3):665—85.

Jang, J.-S. R. , Sun, C.-T. , 1997, Mizutani, E., Neuro-Fuzzy and Soft Computing, Prentice Hall, US,

Kahramanli, H., Allahverdi N., 2007, "Design of a hybrid system for the diabetes and heart diseases" Expert Systems with Applications, 35(1-2): 82-89.

Kara, S., 1995, "A Study of Mitral and Tricuspid Valve Blood Flows by Autoregressive Spectral Analysis Method and Doppler Unit", PhD Thesis, Institute of Science of Erciyes University.

Kara, S., Kemaloğlu S., Erdoğan N., 2006, "Comparison of fast Fourier transformation and autoregressive modelling as a diagnostic tool in analysis of lower extremity venous signals", Computers in Biology and Medicine, 36(5), 484-494.

Kara, S, Dirgenali, F., 2007a, "A system to diagnose atherosclerosis via wavelet transforms, principal component analysis and artificial neural networks", Expert Systems with Applications, Volume 32, Issue 2, February 2007, Pages 632-640.

Kara, S., Güven, A., İçer, S., 2007b, "Classification of Macular and optic nerve disease by principal component analysis", Computers in Biology and Medicine, 37(6): 836-841.

Kara, S., Güven, A., 2007c, "Training a learning vector quantization network using the pattern electroretinography signals", Computers in Biology and Medicine, 37(1): 77-82.

Kim, HC, Kim DJ, Bang SY. 2002, “Face recognition using LDA mixture model”, In: Proceedings of the 16th international conference on pattern recognition, vol. 2: p. 486-9.

Knight, T., Timmis J., 2001, "Assessing The Performance of The Resource Limited Artificial Immune System AINE", Computing Laboratory, University of Kent at Canterbury, Technical Report 3-01.

Kodaz, H., 2007, “Bilgi Kazancı Tabanlı Yapay Bağışıklık Tanıma Sistemi”, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi.

Kotsiantis, S.B., Kanellopoulos D., Pintelas P. E., 2006, “Data Preprocessing for Supervised Leaning”, International Journal of Computer Science, Volume 1(2): 111-117.

Kordylewski, H., Graupe D., Liu K., 2001, “A novel large-memory neural network as an aid in medical diagnosis applications”, IEEE Transactions on Information Technology in Biomedicine 5(3): 202-209.

Kurgan, L., Cios K.J., 2002, “Ensemble of Classifiers to Improve Accuracy of the CLIP4 Machine Learning Algorithm”, accepted for the SPIE's International Symposium on Sensor Fusion: Architectures, Algorithms, and Applications VI.

Kwak, N., Choi, C.-H., 2002, “Input feature selection for classification problems”, IEEE Trans. Neural Networks 13 (1), 143–159.

Latifoğlu, F., Polat, K., Kara, S., Güneş S., 2008a, “Medical diagnosis of atherosclerosis from Carotid Artery Doppler Signals using principal component analysis (PCA), k-NN based weighting pre-processing and Artificial Immune Recognition System (AIRS)”, Journal of Biomedical Informatics, 41(1): 15-23.

Latifoglu, F., Şahan, S., Kara, Güneş, S., 2007a, “Diagnosis of atherosclerosis from carotid artery Doppler signals as a real-world medical application of artificial immune systems”, Expert Systems with Applications, 33(3): 786-793.

Latifoglu, F., Kodaz, H., Kara, Güneş, S., 2007b, “Medical application of Artificial Immune Recognition System (AIRS): Diagnosis of atherosclerosis from carotid artery Doppler signals”, Computers in Biology and Medicine, 37(8):1092-1099.

Lavrac, N., 1999, “Selected techniques for data mining in medicine”, Artificial Intelligence in Medicine, 16(1): 3-23.

Lindsay, I. Smith 2002, A tutorial on principal components analysis (http:// kybele.psych.cornell.edu/~edelman/Psych-465 Spring-2003/PCA-tutorial)

Liu, B., Hsu,W., Ma, Y., 1998, “Integrating Classification and Association Rule Mining” Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98), New York, USA.

Lukasz, A. Kurgan, Krzysztof J. Cios, Ryszard Tadeusiewicz, Marek Ogiela, Lucy S. Doodenday, 2001, “Knowledge discovery approach to automated cardiac SPECT diagnosis”, Artifical Intelligence in Medicine, pp.149-169.

Maaten, L.J.P. Van Der, 2007, “An Introduction to Dimensionality Reduction Using Matlab”, Technical Report MICC 07-07. Maastricht University, Maastricht, The Netherlands.

Martinez, AM, Kak, AC., 2001, “PCA versus LDA", IEEE Trans Pattern Anal Mach Intell, 23(2):228—33.

Marwah, G., Boggess L., 2002, “Artificial Immune Systems for Classification: Some Issues”, 1st International Conference on Artificial Immune Systems (ICARIS-2002), University of Kent at Canterbury, UK, September 9-11, 2002.

Mitchell, T.M., 1997, “Machine Learning”, The McGraw-Hill Comp. Press, New York, USA.

Monsoon, H. Hayes, 1996, Statistical Digital Signal Processing and Modeling, New York, USA: Wiley, ISBN-0-471-59431-8.

Nasraoui, O., Gonzalez F., Dasgupta D., 2002, “The Fuzzy Artificial Immune System: Motivations, Basic Concepts, and Application to Clustering and Web Profiling”, Published at IEEE International Conference on Fuzzy Systems. Published in The Proceedings of The IEEE World Congress on Computational Intelligence, Hawaii, May 12-17.

Özbay, Y., Kara, S., Latifoğlu,F., Ceylan, R., Ceylan, M., 2007a, “Complex-valued wavelet artificial neural network for Doppler signals classifying”, Artificial Intelligence in Medicine, 40(2): 143-156.

Özbay,Y., Ceylan, M., 2007b,“Effects of window types on classification of carotid artery Doppler signals in the early phase of atherosclerosis using complex-valued artificial neural network”, Computers in Biology and Medicine, 37(3): 287-295.

Özşen, S, Kara, S., Latifoğlu, F., Güneş, S., 2007, “A new supervised classification algorithm in artificial immune systems with its application to carotid artery Doppler signals to diagnose atherosclerosis”, Computer Methods and Programs in Biomedicine, 88(3): 246-255.

Özşen, S., Güneş, S., 2008, “Effect of feature-type in selecting distance measure for an artificial immune system as a pattern recognizer”, Digital Signal Processing, 18(4): 635-645.

Pazzani, M., 1995, “An Iterative Improvement Approach for The Discretization of Numeric Attributes in Bayesian Classifiers”, Proceedings of The First International Conference on Knowledge Discovery and Data Mining (KDD-95), Montreal, Canada, 228-233.

Polat, K., Güneş, S., 2007, “A hybrid approach to medical decision support systems: Combining feature selection, fuzzy weighted pre-processing and AIRS”, Computer Methods and Programs in Biomedicine, 88(2): 164-174.

Polat, K., Şahan, S., Güneş,S., 2007a, “Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour) based weighting preprocessing”, Expert Systems with Applications, 32(2): 625-631.

Polat, K., Güneş, S., Tosun, S., 2006a, “Diagnosis of heart disease using artificial immune recognition system and fuzzy weighted pre-processing”, Pattern Recognition, 39(11):2186-2193.

Polat, K., Sekerci, R., Günes, S., 2007b, "Artificial Immune Recognition System Based Classifier Ensemble on the Different Feature Subsets for Detecting the Cardiac Disorders from SPECT Images" DEXA 2007: 45-53.

Polat, K., Güneş, S., 2007c, "A novel approach to estimation of E. coli promoter gene sequences: Combining feature selection and least square support vector machine (FS_LSSVM)", Applied Mathematics and Computation, 190(2):1574-1582.

Polat, K., Güneş, S., 2007d, "A new method to forecast of Escherichia coli promoter gene sequences: Integrating feature selection and Fuzzy-AIRS classifier system", Expert Systems with Applications, In Pres.

Polat, K., Kara, S., Latifoglu, F., Güneş, S., 2006b, "A Novel Approach to Resource Allocation Mechanism in Artificial Immune Recognition System: Fuzzy Resource Allocation Mechanism and Application to Diagnosis of Atherosclerosis Disease", ICARIS 2006: 244-255.

Polat, K., Kara, S., Güven, A., Güneş, S. 2008a, "Comparison of Different Classifier Algorithms to Diagnosing of Macular and Optic Nerve Diseases", Expert Systems, Special issue on Medical Decision Making, Article in Press.

Polat, K., Kara, S., Güven, A., Güneş, S. 2008b, "Usage of class dependency based feature selection and fuzzy weighted pre-processing methods on classification of Macular disease", Expert Systems with Applications, In Press.

Polat, K., Kara, S., Güven, A., Güneş, S., 2007e, "A Hybrid Automated Detection System Based on Least Square Support Vector Machine Classifier and k -NN Based Weighted Pre-processing for Diagnosing of Macular Disease", ICANNGA (2) 2007: 338-345.

Polat, K., Latifoğlu, F., Kara, S., Güneş, S., 2008d, "Usage of Novel Similarity Based Weighting Method to Diagnose the Atherosclerosis from Carotid Artery Doppler Signals", Medical & Biological Eng. & Computing, Vol. 46, pp. 353-362.

Polat, K., 2004, "Özellik Seçme (FS) İle Yapay Bağışıklık Tanıma Sistemi (AIRS) Kullanılarak Medikal Teşhise Gidiş", Yüksek Lisans Tezi, Selçuk Üniversitesi.

Proakis, J.G., Rader C.M., Ling F., Nikias C.L. 1992. Advanced Digital Signal Processing, Macmillian Publishing Company, USA.

Quinlan, J.R., 1986, "Induction of Decision Trees", Machine Learning, 1(1): 81-106.

Rumelhart, D. E., Hilton, G. E., & Williams, R. J. 1986, "Learning representations by back-propagating errors", Nature, 323, 533–536.

Ruther, K., Ehlich, P., Philipp, A., Eckstein, A., Zrenner, E., 1998, "Prognostic value of the pattern electroretinogram in cases of tumors affecting the optic pathway", Graefe's Arch Clin Exp Ophthalmol, 236: 259-263.

Şahan, S., Polat, K., Kodaz, H., Güneş, S., 2005, "The Medical Applications of Attribute Weighted Artificial Immune System (AWAIS): Diagnosis of Heart and Diabetes Diseases", ICARIS 2005: 456-468.

Semmlov, J.L. 2004. Biosignal and Biomedical İmage Processing, Marcel Deccer Inc. USA.

Setiono, R., 2000, "Generating concise and accurate classification rules for breast cancer diagnosis", Artificial Intelligence, vol.18, pp.205-219.

Simpson, P. K. (1989). Artificial neural systems. New York: Pergamon Press.

Stefanadis, C., Stratos C., Boudoulas H., Kourouklis C., Toutouzas P., 1990, "Distensibility of The Ascending Aorta, Comparison of Invasive and Non-invasive Techniques in Healthy Men and in Men with Coronary Artery Disease", Eur Heart J., 990 –996.

Stoll, W., Buse, H., Kroll, P., 1988, "Decompression of the orbit and the optic nerve in different diseases", Journal of Cranio-Maxillofacial Surgery, 16: 308-311.

Tasman, W., 1992, Duane's Foundations of Clinical Ophthalmology, Lippincott Williams and Wilkins, United States of America.

The UCI KDD Archive Information and Computer Science University of California, Irvine http://kdd.ics.uci.edu/databases/

Timmis, J., Neal M., 2000, "Investigating The Evolution and Stability of A Resource Limited Artificial Immune System", In The Proceedings of International Conference on Genetic and Evolutionary Computation Conference (GECCO), Special Workshop on Artifical Immune Systems, Las Vegas, Nevada, USA, 40-41.

Türkoğlu, I., Arslan, A., İlkay, E., 2002, "An expert system for diagnosis of the heart valve diseases", Expert Systems with Applications, 23, 229–236.

Vaitkus, P.J., Cobbold R.S., Johnston K.W., 1988, "A Comparative Study and Assessment of Doppler Ultrasound Spectral Estimation Techniques Part II: Methods and Results", Ultrasound Med. Biol., Vol. 14(8): 673–688.

Watkins, A., 2005, "Exploiting Immunological Metaphors in The Development of Serial, Parallel, and Distributed Learning Algorithms", PhD Thesis, Kent University.

Watkins, A., 2001, "AIRS: A Resource Limited Artificial Immune Classifier", M.S. Thesis, Mississippi State University.

Watkins, A., Timmis J., 2002a, "Artificial Immune Recognition System (AIRS): Revisions and Refinements", In 1st International Conference on Artificial Immune Systems (ICARIS), University of Kent at Canterbury, UK, September 9th-11th, 2002.

Watkins, A., Boggess L.C., 2002b, "A Resource Limited Artificial Immune Classifier", In The Proceedings of The Special Sessions on Artificial Immune Systems in Congress on Evolutionary Computation, IEEE World Congress on Computational Intelligence, Honolulu, Hawaii, May 2002.

Wang, J., 2003, Data mining: opportunities and challenges, IGI Publishing Hershey, PA, USA

West, D., West, V., 2000, "Improving diagnostic accuracy using a hierarchical neural network to model decision subtasks", Internat. J. Med. Inform. 57 (1), 41–55.

Wierzchon, S.T., Kuzelewska U., 2002, “Stable Clusters Formation in An Artificial Immune System”, First International Conference on Artificial Immune Systems (ICARIS), University of Kent at Canterbury, UK, September 9-11, 68-75.

William, P., Honan, James R. Heron, David H. Foster, Graham K. Edgar and Mark O. Scase, 1993, “Abnormalities of visual function in hereditary motor and sensory neuropathy”, Journal of the Neurological Sciences, 114(2):188-192.

Wright, I. A., Gough, N. A. J., 1999, “Artificial neural network analysis of common femoral artery Doppler shift signals: Classification of proximal disease”, Ultrasound in Medicine and Biology, 24(5), 735–743.

Yan, H., Jiang Y., Zheng J., Peng C., Li Q., 2006, “A multilayer perceptron-based medical decision support system for heart disease diagnosis”, Expert Systems with Applications, 30(2), 272-281.

Yan, H., Zheng J., Jiang Y., Peng C., Li Q., 2003, “ Development of a decision support system for heart disease diagnosis using multilayer perceptron”, ISCAS '03. Proceedings of the 2003 International Symposium on Circuits and Systems, 25-28 May 2003, V-709 - V-712 vol.5.

Yi-Wei, C., Chih-Jen, L.,“Combining SVMs with Various Feature Selection Strategies”, NIPS 2003 Feature Selection Challenge, 1-10.

Yi, S., Z. Dezheng, W. ZhongJie, Z. Lisheng, 2005, “A New Weighting Method and Its Application, Neural Networks and Brain”, ICNN&B '05. International Conference, vol.1, pp.266 – 269.

Zhang, F., Kuo, M.D., Brunkhors, A., “E. coli Promoter Prediction Using Feed-Forward Neural Networks”, 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006, EMBS '06, 2025 – 2027.

Zheng, J., Jiang Y., Yan H., “Committee Machines with Ensembles of Multilayer Perceptron for the Support of Diagnosis of Heart Diseases”, 2006 International Conference on Communications, Circuits and Systems Proceedings, Guilin, 25-28 June 2006, 2046 – 2050.