

**ANKARA ÜNİVERSİTESİ  
FEN BİLİMLER ENSTİTÜSÜ**

**YÜKSEK LİSANS TEZİ**

**BULANIK UZMAN SİSTEMLER KULLANILARAK  
TIPTA HASTALIK TEŞHİSİ**

**Negar ZIASABOUNCHI**

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**ANKARA  
2014**

**Her hakkı saklıdır**

## ETİK

Ankara Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım bu tez içindeki bütün bilgilerin doğru ve tam olduğunu, bilgilerin üretilmesi aşamasında bilimsel etiğe uygun davrandığımı, yararlandığım bütün kaynakları atıf yaparak belirttiğimi beyan ederim.

22.12.2014

Negar Ziasabounchi

## ÖZET

Yüksek Lisans Tezi

BULANIK UZMAN SİSTEMLER KULLANILARAK TIPTA HASTALIK TEŞHİSİ

Negar ZIASABOUNCHI

Ankara Üniversitesi  
Fen Bilimleri Enstitüsü  
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. İman ASKERBEYLİ

Kalp hastalıklarının teşhis süreci, hastanın sağlığında çok önemli ve kritik bir rol oynamaktadır. Ayrıca, hastalığı daha ileri bir seviyeye ulaşmasını engellemektedir. Makine öğrenme teknikleri ve veri madencilik algoritmalarının kalp hastalıklarının tanısında önemli katkıda bulunmaktadır. Bu çalışmanın amacı, adaptif sinirsel bulanık çıkarım sistemi (ANFİS) kullanılarak hastanın kalp hastalık derecesini belirlemek için bir yöntem geliştirmektir. Ayrıca geliştirilen yöntem, kümeleme yöntemlerinden K-means ve bulanık C-means yöntemleri ile kıyaslanmıştır. Kullanılan tüm algoritmalar Cleveland kalp hastalığı veri seti üzerinde uygulandıktan sonra test edilmektedirler. ANFİS ve kümeleme sonuçlarının karşılaştırılmasında ANFİS'in daha yüksek başarıma sahip olduğu görülmüştür. Sonuç olarak geliştirilen ANFİS modeli kalp hastalıklarının teşhisi için karar destek sistemi olarak kullanılabileceğini göstermiştir.

**Aralık 2014, 57 sayfa**

**Anahtar Kelimeler:** Uzman sistemler, Bulanık mantık, Hastalık teşhisi

## **ABSTRACT**

Master Thesis

### **MEDICAL DISEASES DIAGNOSTICS USING FUZZY EXPERT SYSTEMS**

Negar ZIASABOUNCHI

University of Ankara  
Graduate School of Natural and Applied Science  
Department of Computer Engineering

Supervisor: Prof. Dr. Iman ASKERBEYLİ

Heart disease diagnosis procedure is very vital and critical issue for the patient's health. In addition, this prevents the disease from progressing to further stages. The role of using machine learning techniques and data mining algorithms in diagnosis of heart disease is very considerable. The aim of this study is to develop a method of classifying for heart disease degree of patient based characteristic data using adaptive neuro fuzzy inference system (ANFIS). Furthermore, the developed method is compared with the K-means clustering method and fuzzy C-means methods. All the algorithms are implemented and tested on a Cleveland heart disease dataset. Compared results are suggesting that the ANFIS has a much probability of success at heart disease diagnosis. Obtained results show that the proposed analysis system can be used as an expert system for heart decision support system.

**December 2014, 57 pages**

**Key Words:** Fuzzy expert system, Fuzzy logic, Disease diagnosis

## TEŞEKKÜR

Tezimin hazırlanması sırasında büyük bir özveriyle beni yönlendiren ve yardımlarını hiçbir zaman eksik etmeyen değerli hocam Sayın Prof. Dr. İman ASKERBEYLİ'ye (Ankara Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı) sonsuz teşekkürlerimi sunarım.

Ayrıca çalışmalarım sırasında manevi desteğini daima hissettiğim aileme özellikle anne ve babama sonsuz teşekkürlerimi sunmaktan onur duyarım.

Negar ZIASABOUNCHI

Ankara, Aralık 2014

## İÇİNDEKİLER

|                                                   |      |
|---------------------------------------------------|------|
| TEZ ONAY SAYFASI                                  |      |
| ETİK.....                                         | i    |
| ÖZET.....                                         | ii   |
| ABSTRACT .....                                    | iii  |
| TEŞEKKÜR.....                                     | vi   |
| SİMGELER DİZİNİ .....                             | vii  |
| ŞEKİLLELR DİZİNİ .....                            | viii |
| ÇİZELGELER DİZİNİ.....                            | ix   |
| 1. GİRİŞ .....                                    | 1    |
| 2. BULANIK SİSTEMLER.....                         | 3    |
| 2.1 Bulanık Mantık Kavramı .....                  | 3    |
| 2.2 Bulanık Kümeler ve Üyelik Fonksiyonları ..... | 3    |
| 2.3 Bulanık Küme İşlemleri.....                   | 7    |
| 2.3.1 Birleşim .....                              | 7    |
| 2.3.2 Kesişim .....                               | 8    |
| 2.3.3 Tümleme .....                               | 8    |
| 2.4 Sözel Değişkenler.....                        | 9    |
| 2.5 Bulanık Kurallar .....                        | 10   |
| 2.6 Bulanık Çıkarım Sistemi .....                 | 11   |
| 2.6.1 Mamdani çıkarım metodu.....                 | 12   |
| 2.6.2 Takagi - Sugeno bulanık metodu .....        | 14   |
| 2.7 Durulaştırma .....                            | 14   |
| 2.7.1 Ağırlık merkezi yöntemi .....               | 15   |
| 2.7.2 En büyüğün ortası yöntemi .....             | 15   |
| 2.7.3 Açığortay yöntemi .....                     | 16   |
| 3. YAPAY ZEKA TEKNİKLERİ.....                     | 17   |
| 3.1 Makine Öğrenme.....                           | 17   |
| 3.1.1 Gözetimli öğrenme .....                     | 17   |
| 3.1.2 Gözetimsiz öğrenme .....                    | 17   |
| 3.2 Uzman Sistemler.....                          | 17   |
| 3.3 ANFİS .....                                   | 18   |
| 3.3.1 ANFİS modelin mimarisi.....                 | 18   |
| 3.3.2 Melez öğrenme algoritması .....             | 20   |
| 3.4 KÜMELEME YÖNTEMİ.....                         | 22   |
| 3.4.1 K-means Algoritması .....                   | 22   |
| 3.4.2 Bulanık C-means algoritması .....           | 23   |
| 3.5 Temel Bileşenler Analizi (PCA) .....          | 24   |
| 3.6 Performans Ölçütleri .....                    | 25   |
| 4. KALP HASTALIKLARI.....                         | 27   |
| 4.1 Kalp Hastalıklarında Risk Faktörleri .....    | 27   |
| 4.1.1 Göğüs ağrısı .....                          | 27   |
| 4.1.2 Yüksek tansiyon .....                       | 27   |
| 4.1.3 Kolesterol .....                            | 28   |
| 4.1.4 Şeker hastalığı .....                       | 28   |
| 4.1.5 Elektrokardiyografi (EKG) .....             | 28   |

|                                                                        |    |
|------------------------------------------------------------------------|----|
| 4.1.6 Maksimum kalp hızı .....                                         | 28 |
| 4.1.7 Talyum testi .....                                               | 29 |
| 4.1.8 Cinsiyet.....                                                    | 29 |
| 4.2 Cleveland Veri Seti.....                                           | 29 |
| 5. TASARIM ve UYGULAMA .....                                           | 31 |
| 5.1 Kalp Hastalığının Teşhisi İçin Bulanık Uzman Sistem Tasarımı ..... | 31 |
| 5.1.1 Giriş ve çıkış üyelik fonksiyonların tasarımı .....              | 31 |
| 5.1.2 Kural tabanı.....                                                | 34 |
| 5.1.3 Bulanık çıkarım ve durulaştırma .....                            | 35 |
| 5.1.4 Sonuç değerlendirme .....                                        | 38 |
| 5.2 ANFİS ile teşhis .....                                             | 38 |
| 5.2.1 Veri toplama .....                                               | 38 |
| 5.2.2 ANFİS modelin tasarımı.....                                      | 39 |
| 5.2.3 ANFİS modelin eğitimi .....                                      | 40 |
| 5.2.4 Cross validation yöntemi .....                                   | 41 |
| 5.2.5 Performans değerlendirme.....                                    | 41 |
| 5.2.6 Sonuçların analizi.....                                          | 41 |
| 5.2.7 Sonuç yüzeyi .....                                               | 43 |
| 5.2.8 Sonuç değerlendirmesi.....                                       | 45 |
| 5.3 Kümeleme Algoritmaları.....                                        | 45 |
| 5.3.1 Kümele algoritmaları ile teşhis .....                            | 45 |
| 5.3.2 Sonuçların analizi.....                                          | 47 |
| 5.3.3 Sonuç değerlendirmesi.....                                       | 50 |
| 6. SONUÇ.....                                                          | 52 |
| KAYNAKLAR .....                                                        | 53 |
| ÖZGEÇMİŞ.....                                                          | 57 |

## SİMGELER DİZİNİ

|             |                              |
|-------------|------------------------------|
| $\mu(x)$    | x Elemanının üyelik derecesi |
| $\cup$      | Birleşim işlemi              |
| $\cap$      | Kesişim işlemi               |
| $\subseteq$ | Alt Küme                     |
| $\subset$   | Öz alt küme                  |
| $\in$       | Kümenin elemanıdır           |
| $\notin$    | Kümenin elemanı değildir     |
| $\bar{A}$   | A kümesinin tümleyeni        |
| $\sum$      | Kartezyen toplam             |
| $\leq$      | Küçük eşittir                |
| $\geq$      | Büyük eşittir                |
| $\circ$     | Ve                           |
| $\vee$      | Veya                         |
| Min         | En Küçük                     |
| Max         | En Büyük                     |

## Kısaltmalar

|       |                                          |
|-------|------------------------------------------|
| ANFİS | Adaptif Sinirsel Bulanık Çıkarım Sistemi |
| BCM   | Bulanık C-means algoritması              |
| BÇS   | Bulanık Çıkarım Sistemi                  |
| DN    | Doğru Negatif                            |
| DP    | Doğru Pozitif                            |
| DSÖ   | Dünya Sağlık Örgütü                      |
| EKG   | Elektrokardiyografi                      |
| FSM   | Feature Space Mapping                    |
| GMP   | Genelleştirilmiş modes pones             |
| GMT   | Genelleştirilmiş modus tollens           |
| HDL   | High-density lipoprotein                 |
| INT   | Incremental Net                          |
| LDA   | Linear Discriminant Analysis             |
| LDL   | low-density lipoprotein                  |
| NN    | Neural Network                           |
| PCA   | Temel Bileşenler Analizi                 |
| PC    | Temel Bileşenler                         |
| STD   | Standard Deviation                       |
| SNB   | Semi Naive Bayes                         |
| YN    | Yanlış Pozitif                           |
| YP    | Yanlış Negatif                           |
| YSA   | Yapay Sinir Ağları                       |

## ŞEKİLLER DİZİNİ

|                                                                                                                                                               |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Şekil 2.1 Uzun boylu kişinin klasik kümesi .....                                                                                                              | 4  |
| Şekil 2.2 Uzun boylu kişinin bulanık kümesi.....                                                                                                              | 5  |
| Şekil 2.3 Bulanık kümeler için örnek üyelik fonksiyonları.....                                                                                                | 6  |
| Şekil 2.4 X evrenine ait iki bulanık küme. a. Bulanık küme A, b. Bulanık küme B .....                                                                         | 7  |
| Şekil 2.5 Bulanık birleşim işlemlerinin grafiksel gösterimi .....                                                                                             | 7  |
| Şekil 2.6 Bulanık kesişim işlemlerinin grafiksel gösterimi .....                                                                                              | 8  |
| Şekil 2.7 Bulanık tümlleme işlemlerinin grafiksel gösterimi .....                                                                                             | 8  |
| Şekil 2.8 Kısa, orta, uzun kişinin bulanık kümeleri .....                                                                                                     | 10 |
| Şekil 2.9 Bulanık çıkarım sistemi .....                                                                                                                       | 12 |
| Şekil 2.10 Bulanık VE ve VEYA işlemleri için sırasıyla minimizasyon ve<br>maksimizasyon operatörlerini kullanan Mamdani tipi bulanık çıkarım<br>sistemi ..... | 13 |
| Şekil 2.11 Birinci dereceden sugeno bulanık modeli .....                                                                                                      | 14 |
| Şekil 2.12 Ağırlık merkezi durulaştırma metodu .....                                                                                                          | 15 |
| Şekil 2.13 En büyüğün ortası durulaştırma metodu .....                                                                                                        | 15 |
| Şekil 2.14 Açıortay durulaştırma metodu .....                                                                                                                 | 16 |
| Şekil 3.1 ANFİS'in mimarisi.....                                                                                                                              | 19 |
| Şekil 5.1 Kan şekeri ve maksimum kalp hızı için üyelik fonksiyonu.....                                                                                        | 32 |
| Şekil 5.2 Kolesterol ve egzersiz için üyelik fonksiyonları .....                                                                                              | 32 |
| Şekil 5.3 Cinsiyet ve yaş için üyelik fonksiyonları.....                                                                                                      | 33 |
| Şekil 5.4 Göğüs ağrısı ve EKG için üyelik fonksiyonları.....                                                                                                  | 33 |
| Şekil 5.5 Old peak ve kan basıncı için üyelik fonksiyonları .....                                                                                             | 33 |
| Şekil 5.6 Talyum testi ve hastalık teşhisi için üyelik fonksiyonları.....                                                                                     | 34 |
| Şekil 5.7 Matlab fuzzy toolbox ile oluşturulan kural tabanı .....                                                                                             | 35 |
| Şekil 5.8 Mamdani yöntemi ile cinsiyet ve kan basıncı yönlerindeki<br>değişikliklerin hastalık sonucu üzerindeki etkisi.....                                  | 36 |
| Şekil 5.9 Mamdani yöntemi ile kolesterol ve kan basıncı yönlerindeki<br>değişikliklerin hastalık sonucu üzerindeki etkisi.....                                | 36 |
| Şekil 5.10 Matlab fuzzy toolbox Rule Viewer arayüzü.....                                                                                                      | 37 |
| Şekil 5.11 Üyelik fonksiyonları eğitimden önce .....                                                                                                          | 40 |
| Şekil 5.12 Kalp hastalık teşhisi için gelişen kural tablosu.....                                                                                              | 40 |
| Şekil 5.13 Üyelik fonksiyonları eğitimden sonra .....                                                                                                         | 41 |
| Şekil 5.14 Hata oranı.....                                                                                                                                    | 42 |
| Şekil 5.15 ANFİS kontrol sistemi sonucu maksimum kalp hızı ve tansiyon üyelik<br>fonksiyonları için .....                                                     | 43 |
| Şekil 5.16 ANFİS kontrol sistemi sonucu kolesterol ve göğüs ağrısı üyelik<br>fonksiyonları için.....                                                          | 43 |
| Şekil 5.17 PCA ve K-means yöntemi ile kümelene sonuçları .....                                                                                                | 49 |
| Şekil 5.18 PCA ve C-means yöntemi ile kümelene sonuçları.....                                                                                                 | 49 |
| Şekil 5.19 Performans sonuçları.....                                                                                                                          | 50 |

## ÇİZELGELER DİZİNİ

|                                                                                         |    |
|-----------------------------------------------------------------------------------------|----|
| Çizelge 3.1 Melez öğrenme algoritması .....                                             | 22 |
| Çizelge 4.1 Cleveland veri setinde kullanılan 13 değişkenin açıklaması .....            | 30 |
| Çizelge 5.1 Bulanık çıkarım sistemi ile elde edilen sonuçlar .....                      | 37 |
| Çizelge 5.2 ANFİS için kullanılan üyelik derecelerinin sayısı .....                     | 39 |
| Çizelge 5.3 Hata değerleri .....                                                        | 42 |
| Çizelge 5.4 Cleveland veri seti için ANFİS ve diğer yöntemlerin karşılaştırılması ..... | 44 |
| Çizelge 5.5 Metotların karşılaştırılması .....                                          | 47 |
| Çizelge 5.6 Sınıflandırma sonuçlarının analizi .....                                    | 48 |

## 1. GİRİŞ

Kardiyovasküler hastalıklar denilen, kalp ve damar hastalıkları çeşitli sorunlar içeren, kalp damarlarının herhangi bir nedenle daralması, tıkanması ve kalp kasının oksijenlenme ve beslenmesinin bozulmasıdır. Dünya Sağlık Örgütü'nün (DSÖ) Raporu'na göre, kardiyovasküler hastalıkları dünya çapında ölümlerin en büyük nedenidir ve bu hastalığın ölüm oranlarının % 80'i, düşük ve orta gelirli ülkelerde gerçekleşmektedir (Anonymous 2014). Dolayısıyla, kalp hastalığının erken dönemde tahmin edilmesi ve önlenmesi büyük önem taşımaktadır. Son yıllarda tıbbi karar verme yöntemlerinin geliştirilmesiyle hastalık oluşmadan önce ve hastalığın tedavi maliyeti artmadan önlenmesi mümkündür. Öte yandan, bazen laboratuvar sonuçlarına ve doktorların tecrübesine dayanan tıbbi teşhisler yetersiz kalmaktadır ve yanlış teşhisler verilebilmektedir. Bu hataları ortadan kaldırmak ve doğru teşhis yapabilmek için, insan faktörünü eleyerek, basit ve hızlı çözümler üretmek amaçlanmaktadır. Bilgisayar teknolojisinin gelişmesiyle birlikte, akıllı yöntemler ve algoritmalar (örneğin sinir ağı, bulanık mantık ve genetik algoritma) karmaşık ve belirsiz tıbbi görevlerde kritik rol oynamaya başlamıştır (Obanijesu ve Emuoyibofarhe 2012). Son yıllarda, birçok makine öğrenme teknikleri ve algoritmaları tıbbi tanı tahmininde kullanılmaktadır. Sınıflandırma ve kümeleme algoritmaları çok sık kullanılan yöntemler arasındadırlar. Sınıflandırma ve kümeleme algoritmaları sırasıyla denetimsiz öğrenme ve denetimli öğrenme için kullanılmaktadır. Sınıflandırma yöntemleri arasında olan ANFİS, melez uzman sistemlerinden biridir. ANFİS, Neuro-bulanık entegrasyonu, sinir ağlarının ilişkisel yapısı ile bulanık mantığın, insan gibi düşünen ve karar veren özelliğini bir arada kullanarak melez akıllı sistem sunmaktadır (Neshat ve Yaghobi 2009). Bu çalışmanın amacı, sınıflandırma ve kümeleme yöntemleri kullanarak kalp hastalarının, sağlıklı veya hasta gruplara doğru sınıflandırılması için uzman sistem hazırlamaktır.

Bu bölümden sonra, tezin ikinci ve üçüncü bölümde sırasıyla bulanık mantık, sınıflandırma ve kümeleme yöntemleri ile ilgili genel bilgi verilmiştir. Dördüncü bölümde ise kalp hastalıkları ve kalp hastalıklarının teşhisi için kullanılan veri seti ile ilgili bilgi verilmiştir. Beşinci bölümde, kalp hastalığının teşhisi için bulanık uzman sistem, ANFİS destekli uzman sistem tasarlanmıştır ve kıyaslama yapabilmek için,

kümeleme algoritmaları uygulanmıştır. Daha sonra sınıflandırma ve kümeleme sonucu elde edilen verilerin analizi yapılmış ve aralarındaki ilişkiler değerlendirilmiştir. Altıncı yani sonuç bölümünde, araştırma süresince denenen modellerden hangisinin teşhis için daha uygun olduğu belirlenmiştir.

Bu tez çalışmasının sonunda kalp hastalıklarının teşhisinde kullanılan yapay zeka tekniklerin hangisinin hastalıklarının teşhisinde daha başarılı olduğu sonucu verilecektir.

## 2. BULANIK SİSTEMLER

### 2.1 Bulanık Mantık Kavramı

Klasik mantık sistemleri bir değerin sadece doğruluğu veya yanlışlığını belirtmektedir ve çok katı sınırlar çizerek kesin olmayan durumlarla ilgilenmemektedir. Klasik kümelerde bir varlık kümenin elemanı olduğunda “1” ve kümenin elemanı değil ise “0” değerini alır. Bulanık mantık, bulanık küme teorisine dayalı ve klasik küme teorisinin aksine 0 ile 1 arasındaki değerlerde bir gerçek sayı olarak alınır. Bulanık kümelerde her bir varlığın belli bir üyelik derece ile bir unsuru temsil eder ve üyelik fonksiyonu  $\mu(x)$  ile gösterilir. Ayrıca bu değerler sözel terimler ile de karakterize edilebilir. Örneğin 'Ali çok uzun boylu', burada 'çok' sözel terim, 'uzun' bulanık değişken ve  $A(\text{uzun})=\{\text{az, çok, çok az, biraz, biraz fazla, çok fazla}\}$  terim kümesidir. Bulanık mantıkta, az, çok, çok az, biraz, biraz fazla, çok fazla, hızlı, yavaş gibi günlük konuşma dilinde kullandığımız küme üyeleri derecelendirilerek, olayların hangi oranlarda gerçekleştirildiği gösterilebilir. Böylece, bulanık kümeler gerçek hayatta kullandığımız terimleri temsil eder. Bulanıklık kavramı ilk olarak Zadeh (1965) tarafından önerilmiştir. Zadeh, bulanık yaklaşımı kullanarak karmaşık ve komplike sistemleri basitleştirmek amacı ile bulanık kümeleri tanıtmıştır.

### 2.2 Bulanık Kümeler ve Üyelik Fonksiyonları

Bulanık kümeler klasik küme teorisinin bir uzantısı olarak kabul edilebilir. Örnek olarak “uzun boylu kişi kime denir?” sorusuna cevap vermek için, bir kişinin uzun boylu kümesine ait olup olmadığı belirlemek gerekmektedir. Bulanık küme mantığında uzun boylu kümesi derecelere ayrılabilir ve kümenin elemanları ait olma dereceleri ile gösterilir. A, uzun boylu kümesi ve  $\mu_A(x) \in [0,1]$  uzun boylu kümesindeki üyelik derecesidir.  $\mu_A$  fonksiyonu ise bulanık A kümesinin üyelik fonksiyonu adı verilir.

$$A = \{(x, \mu_A(x)) | x \in A, \mu_A(x) \in [0,1]\} \quad (2.1)$$

$\mu_A(x) = 1$ , Eğer x, A'nın tamamen içindeyse;

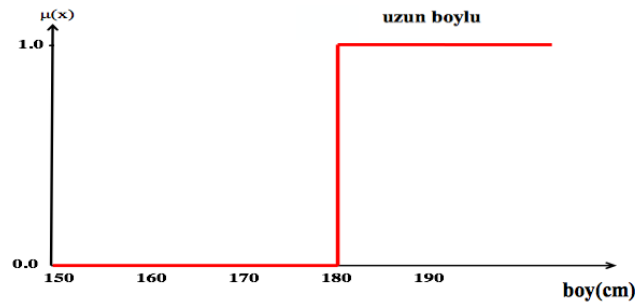
$\mu_A(x) = 0$ , Eğer x, A'nın tamamen dışındaysa;

$0 < \mu_A(x) < 1$ , Eğer x, A'nın bir miktar içindeyse.

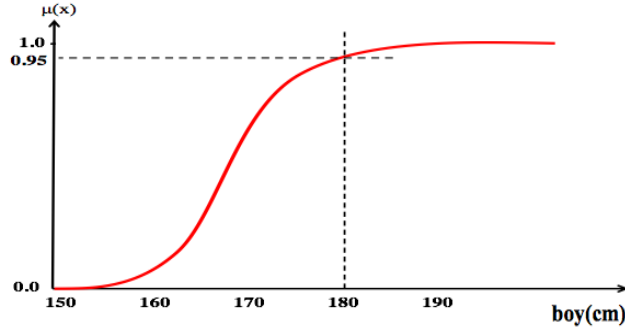
Örnek olarak uzun boylu erkeklerin bir tanımı, “kişinin boyu 180 cm den fazla olursa, uzun boylu bir adam olabilir”. Bu ifade klasik kümelerde şöyle tanımlanabilir;

uzun boylu =  $\{x \mid x \geq 180\}$ .

Şekil 2.1’de görüldüğü gibi klasik küme mantığına göre 180 cm boyundaki bir kişi uzun boyludur ve boyu 180 cm den daha kısa olanlar ise uzun boylu kümesine ait değildir. Hatta 179 cm boyundaki bir kişi kısa boylu olarak kabul edilir. Ancak, bulanık küme metodolojisinde, üyelik derecesi kavramı uygulanmaktadır. Bu mantığa göre 170 cm boyundaki bir kişi kısa değildir çünkü kısmen de olsa uzun boylu insanlar kümesi içindedir. Bulanık mantıkta 160 cm boyundaki kişinin üyelik derecesi 0.6, 170 cm boyundaki kişinin üyelik derecesi 0.7 ve 180 cm boyundaki kişi ise üyelik derecesi 1’dir (Çobanoğlu 2000). Şekil 2.2’de görüldüğü gibi, kişinin boyu 180’den büyük olduğu halde, “uzun boylu” bulanık kümesine tam üyelikle dahil olmaktadır ve “1” üyelik derecesine sahiptir. Fakat kişinin boyu 180’den kısa olduğu halde, üyelik derecesi küçülmeye başlar.



Şekil 2.1 Uzun boylu kişinin klasik kümesi



Şekil 2.2 Uzun boylu kişinin bulanık kümesi

Bir üyelik fonksiyonu, küme üyelerini değerleri ile değişiklik gösteren bir eğridir (Mahmood 2010). Üyelik fonksiyonu grafiğinde  $x$  eksenini küme üyelerini gösterir ve  $y$  eksenini üyelik derecelerini belirler. Bir bulanık küme yapısı iki şeye bağlıdır, birincisi uygun bir evrensel kümenin tanımlanması ve ikincisi uygun bir üyelik fonksiyonu belirlenmesidir (Jang ve Sun, 1995). Genellikle, bulanık kümelerin belirlenmesi uzmanların bilgisine ve tecrübelerine dayanır.

Üyelik fonksiyonlarının (Üçgen, Yamuk, Gauss ve Çan) örnekleri şekil 2.3 gösterilmiştir ve aşağıdaki formüller ile tarif edilmiştir.

(a) Üçgen üyelik fonksiyonu

$$\mu(x; a_1, b, a_2) = \begin{cases} a_1 \leq x \leq b \text{ ise } \frac{x-a_1}{b-a_1} \\ b \leq x \leq a_2 \text{ ise } \frac{a_2-x}{a_2-b} \\ x > a_2 \text{ veya } x < a_1 \text{ ise } 0 \end{cases} \quad (2.2)$$

(b) Yamuk üyelik fonksiyonu

$$\mu(x; a_1, b_1, b_2, a_2) = \begin{cases} a_1 \leq x \leq b_1 \text{ ise } \frac{x-a_1}{b_1-a_1} \\ b_1 \leq x \leq b_2 \text{ ise } 1 \\ b_2 \leq x \leq a_2 \text{ ise } \frac{a_2-x}{a_2-b_2} \\ x > a_2 \text{ veya } x < a_1 \text{ ise } 0 \end{cases} \quad (2.3)$$

(c) Gaussian üyelik fonksiyonu

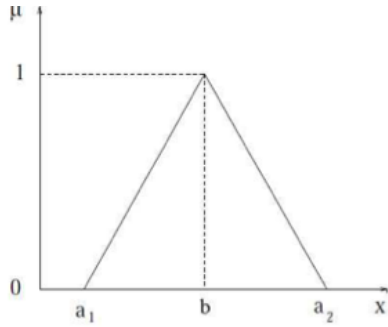
Bu üyelik fonksiyonu iki m ve s parametreleri ile belirlenir.

$$\mu_A(x, m, s) = \exp \left[ \frac{-(x-m)^2}{2s^2} \right] \quad (2.4)$$

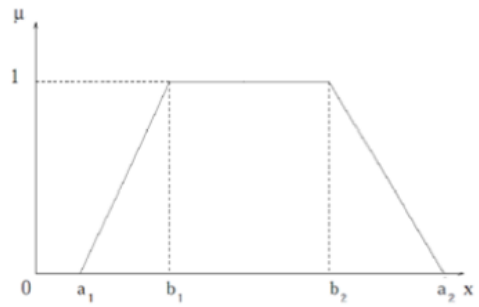
Bu fonksiyonda m fonksiyon merkezini ve s ise genişliğini ifade eder.

(d) Çan şekli üyelik fonksiyonu

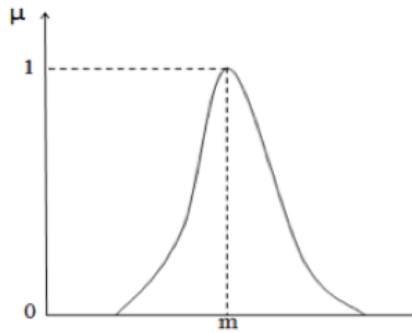
$$\mu(x; a_1, a_2, a_3) = \left\{ \frac{1}{1 + \left| \frac{x-a_3}{a_1} \right|^{a_2}} \right\} \quad (2.5)$$



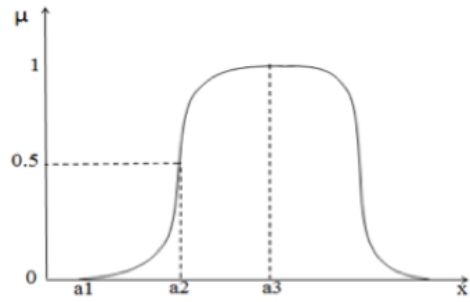
(a) Üçgen üyelik fonksiyonu



(b) Yamuk Üyelik fonksiyonu



(c) Gaussian üyelik fonksiyonu

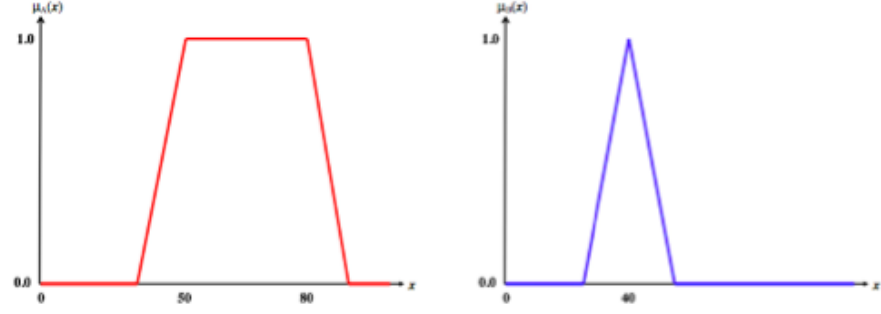


(d) Çan şekilli üyelik fonksiyonu

Şekil 2.3 Bulanık kümeler için örnek üyelik fonksiyonların gösterimi

## 2.3 Bulanık Küme İşlemleri

Bulanıklaştırma işlemlerinden, bulanık birleşim (AND), bulanık kesişim (OR) ve bulanık tamamlayıcısı (NOT) Zadeh tarafından tanımlanmıştır. Şekil 2.4.a, b’de görüldüğü gibi, X evreninde A ve B,  $\mu_A$  ve  $\mu_B$  üyelik fonksiyonları ile iki bulanık küme farz edilerek her bir operatörün işlemi şekil 2.5; 2.7’de gösterilmiştir.



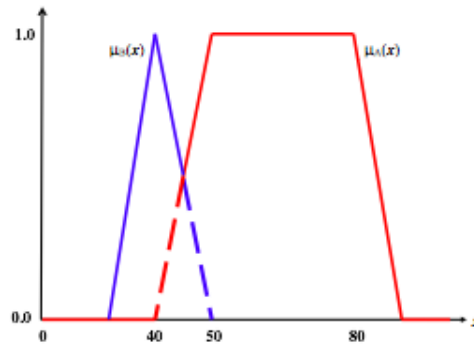
Şekil 2.4 X evrenine ait iki bulanık küme

a. Bulanık küme A, b. Bulanık küme B

### 2.3.1 Birleşim

Her bir kümeye ait ögenin ait oldukları kümelerdeki üyelik derecelerinin en büyüğü alınarak iki tane bulanık alt kümenin birleşim işlemi hesaplanır (Baykal ve Beyan 2004). İki tane A ve B bulanık alt kümelerin birleşimi aşağıdaki gibidir;

$$U[\mu_A(x), \mu_B(x)] = \max[\mu_A(x), \mu_B(x)] \text{ veya } \mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)] \quad (2.6)$$



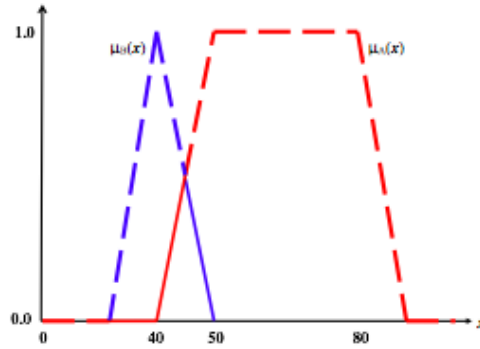
Şekil 2.5 Bulanık birleşim işlemelinin grafiksel gösterimi

### 2.3.2 Kesişim

Her bir kümeye ait ögenin ait oldukları kümelerdeki üyelik derecelerinin en küçüğü alınarak iki tane bulanık alt kümenin kesişim işlemi hesaplanır (Baykal ve Beyan 2004).

İki tane A ve B bulanık alt kümelerin kesişimi aşağıdaki gibidir;

$$I[\mu_A(x), \mu_B(x)] = \min[\mu_A(x), \mu_B(x)] = \text{veya } \mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)] \quad (2.7)$$

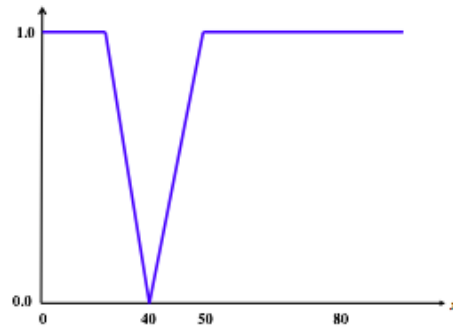


Şekil 2.6 Bulanık kesişim işlemelinin grafiksel gösterimi

### 2.3.3 Tümleme

Bir bulanık kümenin tümleneni aşağıdaki gibidir;

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad (2.8)$$



Şekil 2.7 Bulanık tümleme işleminin grafiksel gösterimi

Aşağıda küme teorisi için gerekli bazı özellikler verilmiştir:

De Morgan kanunu:

$$(\overline{A \cap B}) = \overline{A} \cup \overline{B} \text{ ve } (\overline{A \cup B}) = \overline{A} \cap \overline{B} \quad (2.9)$$

Dağılma özelliği:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \text{ ve } A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad (2.10)$$

Yutma ve tek kuvvet özelliği:

$$A \cup (A \cap B) = A, A \cap (A \cup B) = A \quad (2.11)$$

$$A \cup A = A, A \cap A = A \quad (2.12)$$

Yukarıda anlatılan bu özellikler klasik kümelerle benzemektedir. Ama klasik kümeler ve bulanık kümelerin birbirinden tamamen farklı özelliklerde bulunmaktadır. Bulanık kümeler klasik kümelerin aksine  $A$  bulanık kümesinin tümleyeni ile birleşimi evrensel küme değildir ve  $A$  bulanık kümesinin tümleyeni ile kesişimi boş küme değildir (Çobanoğlu 2000). Bunun nedeni ise  $A$ 'nın tümleyeninin sınırlarının bulanık olmasıdır (Baykal ve Beyan 2004). Matematiksel olarak;

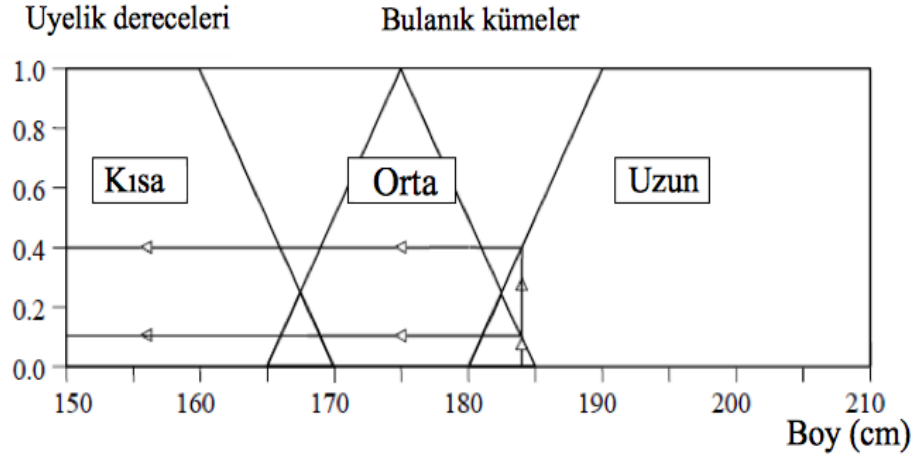
$$A \cap \bar{A} \neq \emptyset \quad (2.13)$$

$$A \cup \bar{A} \neq U \quad (2.14)$$

ifade edilir.

## 2.4 Sözel Değişkenler

Profesör Zadeh (1975) yılında, sözel değişken kavramını ortaya çıkarmış ve sayısal ifadeler yerine sözel ifadeler tanımlayarak, kümeler arası geçişe esneklik kazandırmış. Örnek olarak, bir kişinin boyu uzun, denildiğinde, Sözel değişken; kişinin boyu, değeri ise uzun olduğu anlaşılmaktadır.



Şekil 2.8 Kısa, orta, uzun kişinin bulanık kümeleri

Şekil 2.8’de görüldüğü gibi Boy kelimesi için kullanabileceğimiz değerler; kısa, orta ve uzun kelimelerdir. Bu değerler bulanık kümelerle ifade edilir. Daha sonra ‘eğer...ise...o halde’ kuralları ile bulanık kurallar dizisi oluşturulur.

## 2.5 Bulanık Kurallar

Bulanık mantık kural tabanlı bir sistemdir ve kurallar genelde uzman bir kişinin görüşüne göre oluşur ve bu kurallar istenilen sistem davranışını belirler. Bu kuralların yazılmasında sadece girdi verileri ile çıktılar arasında olabilecek tüm ara (bulanık küme) bağlantıları düşünülür (Yılmaz ve Arslan 2005). Genellikle bulanık kurallar aşağıdaki gibi oluşur:

EĞER öncül (ler) SONRA soncul (ler)

Örneğin X değeri A ise, o halde Y değeri B’dir şeklinde sözel girdi ve çıktı değerlerine sahip olmalıdır (Mahmood 2010).

$R(x, y)$ : Eğer  $A(x)$ , ise  $B(y)$  veya

$R(x, y)$ :  $A(x) \rightarrow B(y)$

Bir Eđer ise o halde kuralı iki ayrı bölümünden oluşur: Öncül ve Soncul. İlk olarak öncül kısmı ifadelerin doğruluđu ve yanlışlığını araştırır. İkinci kısım ise, hüküm kısmı olarak bilinir. birinci kısım doğru olması durumunda ikinci kısımda yer alan ifadeye hükmedilir ve öncül yanlış olduđu aksi halde hüküm kısmı atlanır (Yıldız ve Gedik 2004 ). Bir kuralda öncüller birden fazla parçadan oluşabilir. Tüm çoklu öncüller uygun bulanık operatörlerle bağlandıktan sonra, tüm parçalar aynı anda hesaplanır ve bir çıkış verilir. Daha sonra tüm bu çıkışlar tek bir çıkış olarak toplanır.

Bulanık kümeleri içeren kural olguları: Genelleştirilmiş modes pones ve genelleştirilmiş modus tollens dir.

Genelleştirilmiş modes pones (GMP):

OLGU:  $x \in A$  'dir.  $(R(x))$

KURAL: Eđer  $x \in A$  ise, o halde  $y \in B$  'dir.  $(R(x, y))$

SONUÇ:  $y \in B$  'dir.  $(R(y) = R(x) \circ R(x, y))$

Genelleştirilmiş modus tollens (GMT):

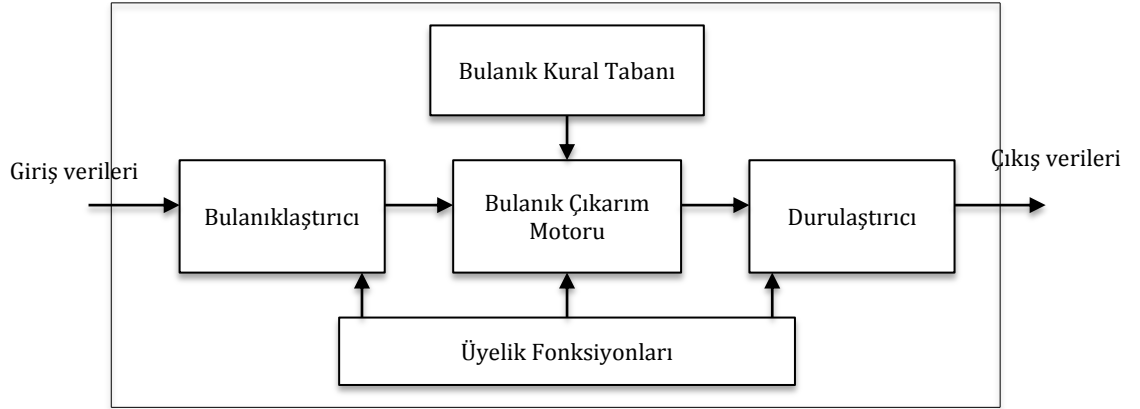
OLGU:  $y \in B$  'dir.  $(R(y))$

KURAL: Eđer  $x \in A$  ise, o halde  $y \in B$  'dir.  $(R(x, y))$

SONUÇ:  $x \in A$  'dir.  $(R(x) = R(y) \circ R(x, y))$

## 2.6 Bulanık Çıkarım Sistemi

Bulanık çıkarım sistemi (BÇS) bulanık kümeler teorisini kullanarak tüm girdi değişkenleri, çıkış değişkenlerine bağlayan bir süreçtir. Şekil 2.9 beş fonksiyonel blokları ile BÇS'in bir örneğini göstermektedir (Hishammuddin 2008).



Şekil 2.9 Bulanık çıkarım sistemi

Genel olarak bulanık model sistemin esas aşamaları aşağıdaki gibidir :

- Bulanıklaştırıcı, kesin girdileri, üyelik fonksiyon mekanizmasına dayalı aralıklara böler ve her üyelik fonksiyonu için uygun sözel ifade atanır.
- Bulanık çıkarım motoru, kural tabanındaki her bir kuralın sonuçlarını birleştirir ve bir çıktı bulur. Başka bir deyişle bu motor her bir kuralın çıkarımlarını birleştirerek tüm sistemin girdileri altında nasıl bir çıktı vereceğini belirler (Yılmaz ve Arslan 2005).
- Durulaştırıcı belirli bir durulama yöntemi ile bulanık çıkışı, kesin çıkışlı bir değere dönüştürür.

En yaygın olarak kullanılan bulanık çıkarım metotları; Mamdani metodu ve Takagi-Sugeno metodudur. Bu iki bulanık çıkarım yöntemlerin arasındaki fark soncul kısmının özelliğidir.

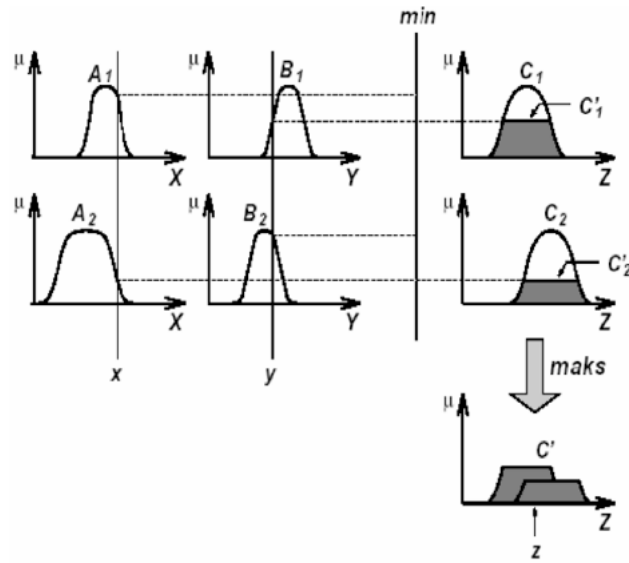
### 2.6.1 Mamdani çıkarım metodu

Mamdani yöntemi ilk defa bir buhar motorunu kontrol etmek için kullanılmıştır (Mamdani ve Assilian 1975). Bu yöntem kolay oluşturulan ve insan davranışlarına

yakın olduğu için yaygın bir kullanıma sahiptir (Yılmaz ve Arslan 2005). Mamdani yönteminde sonculler, bulanık kümelerdir ve bu bulanık kümeler birleştirme işleminden sonra tek bir bulanık çıkışı olarak elde edilir ve bu bulanık çıkış çeşitli durulama yöntemleri ile bir kesin çıkışa dönüştürülür. Şekil 2.10 iki değişkenli ve iki kurallı bir Mamdani tipi bulanık modeldir. Burada  $x$  ve  $y$  giriş değişkenleridir ve  $z$  ise çıkış değişkendir. Bu yöntemde önce giriş değişkenlerin kural tablosuna göre üyelik dereceleri belirlenir ve Min operatörle çıkış bulanık kümesi elde edilir. Daha sonra her bir kural için elde edilen çıkış bulanık kümeleri birleştirerek tek bir bulanık küme elde edilir. Bu birleşim işlemini Max operatörü ile yapılır ve son olarak kesin bir değere ulaşmak için durulaştırma işlemi yapılır.

Kural1: Eğer  $x=A_1$  VE  $y=B_1$  İse  $z=C_1$

Kural2: Eğer  $x=A_2$  VE  $y=B_2$  İse  $z=C_2$

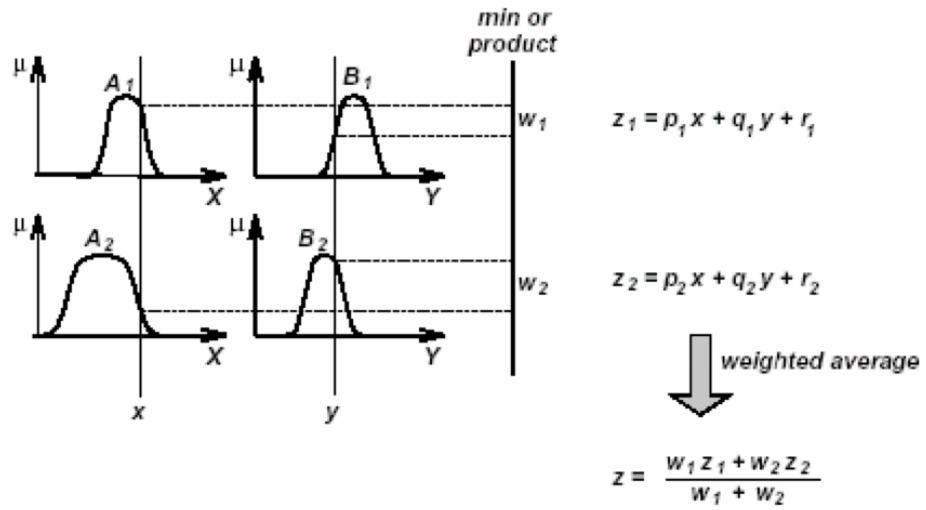


Şekil 2.10 Bulanık VE ve VEYA işlemleri için sırasıyla Min ve Max operatörlerini kullanan Mamdani tipi BÇS

### 2.6.2 Takagi - Sugeno bulanık metodu

Sugeno metodunda (Sugeno 1985) sonuculler, girişlerden oluşan polinomdur. Son çıkış, her kuralın çıkış ağırlıklı ortalamasıdır. Birinci dereceden Sugeno tarzı bulanık kural formu aşağıdaki gibidir:

Eğer  $x$ ,  $A$  ve  $y$  de  $B$  ise o halde  $z=ax+by+c$



Şekil 2.11 Birinci dereceden Sugeno bulanık modeli (Jang ve Sun 1995)

$A$  ve  $B$  öncül kısmındaki bulanık kümelerdir. Soncul kısmında ise  $ax+by+c$  doğrusal bir matematiksel fonksiyondur. Şekil 2.11 birinci dereceden Sugeno bulanık modeli için bulanık çıkarım mekanizmasını göstermektedir.

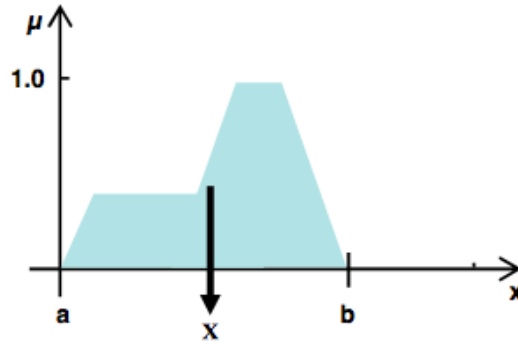
### 2.7 Durulaştırma

Önceden belirtildiği gibi çıkarım mekanizmasından elde edilen dilsel terimler içeren bulanık kümeyi, kesin değere dönüştürmek için durulaştırma yapılır. Çeşitli durulama yöntemleri olmak üzere burada üç yöntemden bahsedilecek.

### 2.7.1 Ağırlık merkezi yöntemi

Durulaştırma yöntemleri arasında en yaygın olan yöntemdir ve ağırlık merkezini bulma kavramına dayanmaktadır. Matematiksel olarak aşağıdaki gibi ifade edilir. Şekil 2.12 Bulanık küme çıkışı için  $[a, b]$  aralığındaki hesaplandığı, ağırlık merkez noktasını göstermektedir.

$$x = \frac{\sum_a^b \mu_c(x) \cdot x}{\sum_a^b \mu_c(x)} \quad (2.15)$$

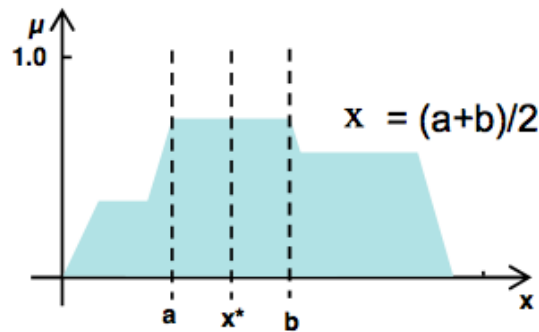


Şekil 2.12 Ağırlık merkezi durulaştırma metodu

### 2.7.2 En büyüğün ortası yöntemi

En büyük üyelik derecelere sahip olan elemanların ortasını bulur. Matematiksel olarak aşağıdaki gibi ifade edilir. Yöntemin grafiksel açıklaması şekil 2.13'de gösterilmiştir.

$$x = \sum_{j=1}^m \frac{x_j}{m} \quad (2.16)$$



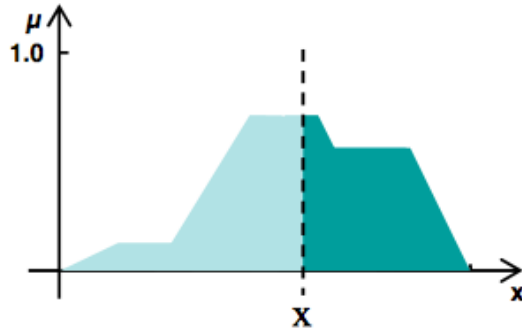
Şekil 2.13 En büyüğün ortası durulaştırma metodu

### 2.7.3 Açıortay yöntemi

Bu yöntem elde edilen alanı, iki eşit alana böler. Şekil 2.14’de bu yöntemin grafiksel gösterimi verilmiştir.

$$\int_{\alpha}^x \mu_A(x)dx = \int_x^{\beta} \mu_A(x)dx \quad (2.17)$$

Burada  $\alpha = \min\{x|x \in X\}$  ve  $\beta = \max\{x|x \in X\}$  dir.



Şekil 2.14 Açıortay durulaştırma metodu

### **3. YAPAY ZEKA TEKNİKLERİ**

#### **3.1 Makine Öğrenme**

Makine öğrenmesinde öncelikle öğrenme yapılacak veri setinin uygulanacak öğrenme metoduna uygun bir şekilde hazırlanması gerekir ve yeni bir metod bulunduğundan sonra bu metodun performansı ölçülür ve diğer metodlarla karşılaştırılması yapılır (Uzun 2007). Son yıllarda, birçok makine öğrenme teknikleri ve algoritmalar tıbbi tanı tahmininde kullanılmıştır. Veri madenciliğinde kullanılan iki temel algoritma kümeleme ve sınıflandırma algoritmalarıdır.

##### **3.1.1 Gözetimli öğrenme**

Gözetimli öğrenme, eğitim verileri üzerinden öğrenmeyi sağlayan bir makine öğrenmesi tekniğidir. Bu öğrenme tekniğinde algoritma giriş değerleri ve çıkış değerleri arasındaki ilişkiyi gösteren bir fonksiyon üretir. Eğitim verisi, giriş değerleri ve hedef değerlerden oluşur.

##### **3.1.2 Gözetimsiz öğrenme**

Gözetimsiz öğrenme modeli gözlemlere bağlı bir makine öğrenmesi tekniğidir (uzun 2007). Bu yöntem, gözetimli öğrenmeden farklı olarak çıkış değerleri kullanmadan sadece girdiler üzerinden öğrenme işlemini gerçekleştirir. Burada ağ kendisi, veri setinin ortak özelliklerini, benzerliklerini bulması ve giriş verileri uygun kategorilere sınıflandırmak için öğrenmesi gerekir.

#### **3.2 Uzman Sistemler**

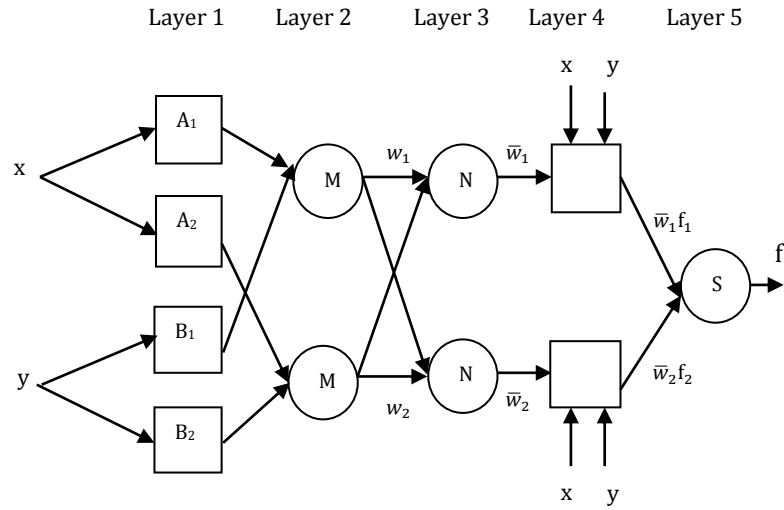
Uzman sistemler bir uzmanın kendi uzmanlık alanındaki bilgisini ve tecrübesinden yararlanarak, bilgisayar ortamına geçiren bilgisayar programlarıdır. Mimari olarak uzman sistemleri diğer karar destek sistemlerinden ayıran en önemli özellik uzman sistemler ‘deki “Bilgi Tabanı” ve çıkarım mekanizmasıdır (Bilge 2007).

### 3.3 ANFİS

Bulanık kurallar bir uzmanın fikri veya deneyimine dayanarak oluşturulur. Ancak her uzman kişi bilgilerini paylaşmak istemeyebilir ve uzman bilgi kullanmak için var olan standart bir yöntem yoktur. Sonuç olarak, yapay sinir ağları (YSA) öğrenme algoritmaları tarafından otomatik olarak bilgi elde edebilmek ve bulanık kurallar oluşturabilmek için, YSA ve bulanık sistemleri birleştirilmiştir. Jang (1993) iki tekniği birleştirerek ANFİS'i önerdi. Bu yöntem, YSA öğrenme yeteneğini ve bulanık mantığın karar verme yeteneğinden birlikte yararlanır.

#### 3.3.1 ANFİS modelin mimarisi

ANFİS yapısı, Takagi-Sugeno tipi bulanık sistemlerine sahip bir ağ yapısıdır. ANFİS bu çıkarım sistemini uygulamak için verilen giriş-çıkış veri seti ve uygun üyelik fonksiyonları ile bulanık kurallar üretir. ANFİS, YSA'nın öğrenme yeteneğini uygulamak için melez öğrenme algoritmasını (en küçük kareler yöntemi ve geri yayılmalı öğrenme algoritması) kullanarak üyelik fonksiyonların aralıklarını değiştirir ve düzenler. Bu düzenleme bulanık sistemlerin modellediği veriler yardımıyla çevresel bilgiyi kullanarak ilgili sistemi öğrenmesini ve kendi kendini güncellemesini sağlar (Ok vd. 2011). Şekil 3.1'de ANFİS'in ağ yapısı şematik olarak gösterilmiştir.



Şekil 3.1 ANFİS'in mimarisi

ANFİS mimarisi 5 katmandan oluşur. ANFİS yapısındaki her katmana ait düğüm işlevleri aşağıda açıklanmıştır.

Katman1: bulanıklaştırma katmanı olarak bilinir ve bu katmandaki düğümler, kare şeklindedir ve bulanık kümelerle ifade edilir. Bu düğümlerin çıkışı ise, seçilen üyelik fonksiyonun, üyelik dereceleridir.

$$O_i^1 = \mu_{A_i}(x), \quad i = 1, 2 \quad (3.1)$$

$$O_i^1 = \mu_{B_{i-2}}(y), \quad i = 1, 2 \quad (3.2)$$

Burada düğümlerin giriş değerleri x ve y'dir. A<sub>i</sub> ve B<sub>i</sub> bulanık kümelerin çıkış değerleri ise bu kümelerle ilişkili üyelik fonksiyonlarının üyelik dereceleridir ve  $\mu_{A_i}$  ve  $\mu_{B_i}$  ile gösterilir. girişlerin üyelik derecesini hesaplamak için Gauss üyelik fonksiyonu kullanılmıştır ve eşitlik (3.3)' de verilmiştir.

$$O_i^1 = \mu_{A_i}(x) = e^{-\frac{1}{2}\left(\frac{x-c_i}{\sigma_i}\right)^2} \quad (3.3)$$

Burada c üyelik fonksiyonunun merkezini temsil eder ve  $\sigma$  üyelik fonksiyonunun genişliğini belirler. Bu parametrelere öncül parametreler denir.

Katman 2: İkinci katman, kural katmanıdır. Bu katmandaki giriş değerleri bir önceki katmandan elde edilen üyelik dereceleridir. Böylece her düğümdeki üyelik derecelerinin çarpımı sonucu çıkış değeri hesaplanır ve bu çıkış ilgili kuralın ateşleme ağırlığını temsil eder.

$$O_i^2 = w_i = \mu_{A_i}(x_1) \times \mu_{B_i}(x_2) \quad i = 1,2 \quad (3.4)$$

Katman 3: Normalizasyon katmanıdır. Bu katmandaki her bir düğüm, kural katmanından gelen tüm düğümleri giriş değeri olarak kabul etmekte ve bu katmanda her bir kuralın normalleştirilmiş ateşleme seviyesi hesaplanmaktadır (Şentürk 2010). Burada  $i$ . düğümün çıkışı,  $i$ . kuralın ateşleme ağırlığının tüm kuralların toplam ateşleme ağırlığının oranı hesaplanır.

$$O_i^3 = \bar{w} = \frac{w_i}{\sum w_i} \quad i = 1,2 \quad (3.5)$$

Katman 4: Bu katmandaki düğümler adaptif olarak adlandırılır ve kare şeklindedir. Bu katmanın  $i$ . düğümün çıkışı eşitlik (3.6) verilmiştir.

$$O_i^4 = \bar{w}_i f_i + \bar{w}_i (p_i x_1 + q_i x_2 + r_i) \quad i = 1,2 \quad (3.6)$$

Burada  $(\bar{w})$ , 3'üncü katmanın çıkış değeridir ve  $\{p_i, q_i, r_i\}$  ise bulanık çıkarım sisteminin soncul parametreleridir.

Katman 5: Bu katmanda tek düğüm vardır. Bu düğüm gelen tüm sinyallerin toplamını hesaplayarak sistemin nihai çıkışını elde eder.

$$O_i^4 = y = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{w_i f_i} \quad i = 1,2 \quad (3.7)$$

### 3.3.2 Melez öğrenme algoritması

Melez öğrenme algoritmasında, ANFİS ağ yapısına ait iki parametre kümesi vardır. Bu kümeler giriş ve çıkış parametrelerden oluşur ve sırasıyla  $S1$  ve  $S2$  ile ifade edilir.

Toplam parametre kümesi ise ;  $S=S1+S2$  şeklinde ifade edilir (Özçalık ve Uygur 2003). Ayrıca bu öğrenme algoritması iki aşamadan oluşur; İleri besleme ve geri besleme. İleri beslemede giriş parametreleri sabit tutularak en küçük kareler yöntemi ile,  $S2$  kümesindeki çıkış parametreler lineer bir şekilde oluşur. Çıkış aşağıdaki gibidir:

$$f = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 = \bar{w}_1(p_1x + q_1y + r_1) + \bar{w}_2(p_2x + q_2y + r_2) \\ = (\bar{w}_1x)p_1 + (\bar{w}_1y)q_1 + (\bar{w}_1)r_1 + (\bar{w}_2x)p_2 + (\bar{w}_2y)q_2 + (\bar{w}_2)r_2 \quad (3.8)$$

$$y = f(i, s) \quad (3.9)$$

Burada  $i$  girdi değişkenleri vektörü,  $s$  parametreler kümesidir. ANFİS ağın çıkışının matrisel şekli aşağıda gösterilmiştir.

$$A\theta = Y, \quad (3.10)$$

Burada  $\theta$ , çıkış parametrelerinin elemanlarından oluşan, bilinmeyen vektördür. Bu bir standart lineer en küçük kareler problemidir ve  $\theta$  için en iyi çözüm  $\|A\theta - Y\|^2$ 'nin minimum değeri olan, en küçük kareler tahmini (LSE)  $\theta^*$ 'dir (Özçalık ve Uygur 2003).

$$\theta^* = (A^T A)^{-1} A^T y \quad (3.11)$$

İkinci aşamada ise sonuç parametreleri sabit alınır ve ağın çıkışındaki hata sinyali geriye doğru giderek, geri yayımlı öğrenme yöntemi ile giriş parametreleri güncellenir. Aşağıdaki denklem giriş parametreleri güncellemek için kullanılır.

$$\Delta x = -\eta \frac{\partial E}{\partial x} \quad (3.12)$$

$x$  giriş parametreleri,  $\eta$  öğrenme oranını,  $E$  ise sistemin çıkışındaki hata değerini ifade eder. İleri besleme ve geri beslemedeki işleyiş döngü süreci, tüm sistem hatası belirlenen hata değerinden küçük olana kadar veya fazla bir değişim göstermeyinceye kadar tekrar edilir (Şentürk 2010). İleri besleme ve geri besleme işlemlerin özeti çizelge 3.1'de gösterilmiştir.

Çizelge 3.1 Melez öğrenme algoritması (Şentürk 2010)

|                           |                          |                        |
|---------------------------|--------------------------|------------------------|
| Melez Öğrenme Algoritması | İleri Besleme            | Geri Besleme           |
| Giriş Parametreleri       | Sabit                    | Geri Yayılmalı Öğrenme |
| Sonuç Parametreleri       | En Küçük Kareler Yöntemi | Sabit                  |

### 3.4 Kümeleme Yöntemi

Kümeleme analizi, örnekleme popülasyonu karakterize etmek için, benzer özelliklere sahip olan veri örneklerini gruplara sınıflandırır (Maimon ve Rokach 2010). Bu grupların her birine küme denir. Küme içindeki elemanların benzerliği fazla ve kümeler arasındaki benzerlik az olmalıdır (Demiralay ve Çamurcu 2005). Kümeleme algoritmalarının hedefi, veri modelleri arasındaki benzerlikleri tanıyarak , veri setindeki yapıyı keşfetmek ve düzenlemektir (Fan vd. 2009).

#### 3.4.1 K-means Algoritması

Yaygın olarak kullanılan bölümlenmeli kümeleme tekniklerinden biri, K-means algoritmasıdır. K-means,  $N$  tane nesneden oluşan veri setini, kullanıcı tarafından belirlenen  $K$  tane kümeye böler ve bu kümeler merkez noktaları ile temsil edilir. İlk olarak küme merkezlerinin başlangıç noktaları rastgele hesaplanır. Her nesne Öklid mesafeye (Euclidean Distance) göre kendine en yakın küme merkezine dahil edilir. Daha sonra, her bir kümenin ortalama değeri hesaplanır ve yeni küme merkezleri belirlenir ve tekrar nesne-merkez uzaklıkları incelenir (Işık ve çamurcu 2010). Küme merkezlerinin güncellemesine ait işlemler, küme merkez değerleri fazla bir değişim göstermeyinceye kadar tekrarlanır (Maimon ve Rokach 2010). Burada amaç toplam ortalama hatayı minimize etmektir (Ünal vd. 2011). Küme merkezleri, Denklem (1) ile hesaplanır.

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (3.13)$$

$N_k$  ,  $k$  kümesine ait elemanların sayısı ve  $\mu_k$ ,  $k$  kümesine ait tüm örneklerin ortalamasını göstermektedir. K-means algoritmasının anlatımı aşağıdaki gibi 4 adımda özetlenmiştir.

Girdi :  $N$  (örnek veri seti),  $K$  (küme sayısı), Çıktı : kümeler ;

Adım 1 :  $K$  tane küme merkezi belirlenir.

Adım 2 : Her nesne kendine en yakın küme merkezine atanır.

Adım 3: oluşan kümelerde yeni merkezler belirlenir.

Adım 4: Ortalama değerleri sabit olana kadar adım 2 ve 3' ü tekrarlanır.

### 3.4.2 Bulanık C-means algoritması

Bulanık C-means algoritması (BCM), en iyi bilinen ve yaygın olarak kullanılan bulanık kümeleme yöntemlerinden biridir. Bulanık mantığın üyelik fonksiyon kavramı, bir nesnenin birden fazla kümeye üye olmasını sağlar. Başka bir deyişle bir nesne üyelik dereceleri ile 0 ve 1 arasında değişerek kısmen de olsa tüm kümelere ait olabilir ve her bir nesnenin tüm kümelere ait toplam üyelik dereceleri 1 olmalıdır (Ross 1995). BCM algoritması temel fikri, amaç fonksiyonu ( $J_m$ ), minimize etmektir. Bu algoritma üst üste binen bir veri kümeleme tekniğidir ve burada her nesne  $X = \{x_1, x_2, \dots, x_N\}$  , üyelik fonksiyonun belirttiği derece ile ( $u_{ij}$ ),  $C$  kümesine aittir.

$$J_m(U, x_1, \dots, x_N) = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m \leq \infty \quad (3.14)$$

Burada  $m$ , 1'den büyük herhangi bir gerçektek sayıdır.  $u_{ij}$ , küme  $C$  ait, 0 ve 1 arasında değişen  $x_i$  veri noktasının üyelik derecesi ve ayrıca  $U$  üyelik matrisin bir elemanıdır.  $c = \{c_1, c_2, \dots, c_N\}$  küme merkez vektörüdür.  $\| * \|$ ,  $j$ 'inci küme merkezi ve  $i$ 'inci veri noktası arasındaki Öklid mesafedir. Algoritmanın başlangıcında,  $U$  üyelik matrisi

rasgele hesaplanır ve küme merkezleri güncellenir. Bu işlemler sırasıyla denklem (3.15) ve (3.16)' de gösterilmiştir.

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left\{ \frac{\|x_i - c_j\|^{2/m-1}}{\|x_i - c_k\|} \right\}} \quad (3.15)$$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m} \quad (3.16)$$

Eski  $U$  matrisi ile yeni  $U$  matrisi karşılaştırılır ve fark  $\varepsilon$ 'dan küçük olana kadar işlemler devam eder (Işık ve Çamurcu 2010).

$$\max_{ij} \{u_{ij}^{(k+1)} - u_{ij}^{(k)}\} < \varepsilon \quad (3.17)$$

burada  $\varepsilon$ , 0 ve 1 arasında sonlandırma kriteridir ve  $k$  ise yineleme sayısıdır. Algoritma aşağıdaki adımlardan oluşmaktadır:

Adım 1: kümelerin sayısı belirlenir.

Adım 2: rastgele 0 ve 1 arasındaki üyelik değerleri ile  $U$  üyelik matrisi hesaplanır.

Adım 3:  $k$ 'inci aşamada,  $c_j$  ve  $U^k$  kullanarak  $c$  bulanık küme merkezleri hesaplanır.

Adım 4:  $U$  matrisi güncelleir:  $U^{k+1} = U^k$

Adım 5: denkem (3.17) kullanılarak  $\varepsilon$  bitene kadar, Adım 3 ve 4 tekrarlanır.

### 3.5 Temel Bileşenler Analizi (PCA)

PCA'in ana fikri Orijinal veri setine ait maksimum bilgiyi koruyarak, veri setinin boyutunu azaltmaktır. PCA, temel bileşenler (pc) olarak adlandırılan bir dizi ilişkili değişkenlerin daha az sayıdaki ilişkisi olmayan, doğrusal ve ortogonal özelliklerini

sahip olan yeni deęişkenler setine dönüştüren bir aritmetik işlemi içermektedir. Böylelikle ilk bir kaç deęişken orijinal veri içerisindeki deęişimin büyük miktarını muhafaza etmektedir (Çilli ve Arıtan 2010). Başka bir deęişle, problem aşağıdaki gibi ifade edilebilir:  $X=(X_1, ..., X_d)^T$  gibi  $d$  boyutlu bir uzaydaki deęişkenleri önemli bilgi kaybı olmadan  $Y=(Y_1, ..., Y_D)^T$   $D < d$ , gibi  $D$  boyutlu bir uzaya daha az sayıda deęişkene indirgenmektedir. Birinci temel bileşen, mümkün olan en yüksek bir varyansa sahiptir ve diğer bileşenlerin her biri, bir önceki bileşenlere dikey olacak şekilde ve mümkün olan en yüksek bir deęişkenliğe sahip bir şekilde seçilirler. (Subbuthai, Periasamy ve Muruganand 2012). PCA çoğunlukla veri analizi aracı olarak kullanılır ve bazı makine öğrenme yöntemlerinde öngörü performansını artırır. Temel bileşenlerin özellikleri, kovaryans matrisinin öz vektörleri ve öz deęerlerine dayanır (Jolliffe 2002).

### 3.6 Performans Ölçütleri

Tıbbi karar verme yöntemleri ile erken ve doğru teşhis yapabilmek ve bu yöntemlerin güvenilirliğini kontrol etmek oldukça hassas ve önemli konulardan biridir. Bu yöntemlerin verdiği sonuçları güvenilir bir şekilde incelemek için öncelikle testin gerçek etkinlik düzeyinin denetlenmesi gerekmektedir (Ertorsun vd.). Testin verimliliğini ve güvenilirliğini belirlemek için duyarlılık, özgüllük ve doğruluk gibi istatistiksel ölçütler kullanılır. Bu ölçütleri kullanmak için gerçek tanı sonuçları bilinmelidir. Duyarlılık (Sensitivity) testin, gerçek hastalar içinden hastaları ayırma yeteneğidir ve Özgüllük (Specificity) testin, gerçek sağlamlar içinden sağlamları ayırma yeteneğidir (Ertorsun vd. 2012). Bir tanı testinin doğruluğu (Accuracy) ise doğru sınıflanan örneklerin toplam doğru ve yanlış sınıflanan örneklerin oranı ile tespit edilir. Duyarlılık, özgüllük ve doğruluk DP, DN, YN ve YP ile açıklanır. Hastalığın, bir hastada mevcut olduğu zaman teşhis doğru pozitifdir (DP). Gerçekte hasta olmayan bir kişi, test sonuçlara göre hasta olmadığı belirtilirse, teşhis doğru negatiftir (DN). DP ve DN her ikisi de doğru sınıflanmayı gösterir. Ancak, hiçbir tıbbi test yüzde yüz doğru değildir. Yanlış pozitif (YP) aslında hasta olmayan bir kişide, test hatalı olarak hastalığın varlığını gösterir. Yanlış negatif (YN) tanısı ise hasta olan bir kişi için test yanlışlıkla hastalığın yokluğunu gösterir. YP ve YN her ikisi de yanlış sınıflanmayı gösterir.

$$\text{Duyarlılık} = DP/(DP + YN) \quad (3.18)$$

$$\text{Özgüllük} = DN/(DN + YP) \quad (3.19)$$

$$\text{Doğruluk} = (DN + DP)/(DN + DP + YN + YP) \quad (3.20)$$

#### **4. KALP HASTALIKLARI**

Kardiyovasküler hastalıklar denilen, kalp ve damar hastalıkları çeşitli sorunlar içeren, kalp damarlarının herhangi bir nedenle daralması, tıkanması ve kalp kasının oksijenlenme ve beslenmesinin bozulmasıdır. Kalp damar hastalıklardan, Koroner arter hastalığı (kalp krizi), Kalp ritim bozuklukları, Damar hastalıkları, Kapak hastalıkları gibi bir çok hastalıklardan oluşur. Kalp krizlerinin çoğu koroner arter hastalığından meydana gelir. Koroner arter hastalığında, kalp kasına gerekli oksijen ve besin gitmesine sağlayan atardamarların biri (koroner arterler) daraldığında meydana gelir. Eğer koroner arter aniden tıkanır ise, kalp kasına giden kan akımı engellenir ve kan akımı hızlı bir şekilde geri dönmez ise, kalp kasının bir bölümü ölmeye başlar buda kalp krizine neden olur (Anonim 2011). Kardiyovasküler hastalığı dünya çapında ölümlerin en büyük nedenidir. Son yirmi yıl içinde, kardiyovasküler hastalıklardan kaynaklanan ölüm oranları hızlı bir şekilde artmaktadır. Kalp hastalıklarının ana risk faktörleri, sigara, kolesterol, şeker hastalığı, hipertansiyon ve yaşıdır. Bu risk faktörler değiştirilebilen ve değiştirilemeyen iki gruptan oluşur. Değiştirilebilir risk faktörlerini tanımak ve bunlarla mücadele etmek hastalığı önlemede çok önemlidir (Anonim 2014a).

##### **4.1 Kalp Hastalıklarında Risk Faktörleri**

###### **4.1.1 Göğüs ağrısı**

Kalp hastalıklarının en önemli ve en sık rastlanan belirtilerinden biridir. Koroner arter hastalığı nedeniyle yeterince oksijen alamayan kalp kası, bunu ağrı ile gösterir (Anonim 2011).

###### **4.1.2 Yüksek tansiyon**

Tansiyon kısaca “kan basıncı” olarak tanımlanır. Damarlardaki kan, dolaşım sırasında bir basınç oluşturur ve bu basınç, alınan gıda, yapılan iş, harcanan güçle bağlantılı

olarak gün içinde küçük farklılıklar gösterebilir (Anonim 2014b). Yüksek tansiyon, kan basıncının yüksek olmasından (hipertansiyon) kaynaklanır ve ciddi bir sağlık sorunudur.

#### **4.1.3 Kolesterol**

Kişinin kolesterol düzeyi ne kadar yüksekse, o kadar kalp damar hastalığının riski artar. Kolesterol iyi kolesterol (HDL), kötü kolesterol (LDL) olarak bilinen iki çeşitten oluşur. HDL, 40 mg/dl'den fazla olması ve LDL 130 mg/dl'den düşük olması kalp hastalıklarının riskini azaltır.

#### **4.1.4 Şeker hastalığı**

Eğer şekerin kanda normal bulunması gereken miktardan daha fazla bulunur ise damarlara zarar verir ve bu damarlarda hasar sonucu yeterli miktarda kan dokulara gidemezse kalp krizine yol açabilir. Bu nedenle şeker normal sınırlarda tutulmalıdır. Şeker miktarı 120 mg dl'nin altında olmalıdır.

#### **4.1.5 Elektrokardiyografi (EKG)**

EKG, düzenli veya düzenli olmayan kalp ritim bozukluklarını belirlemek, göğüs ağrısı şikayeti ile başvuran hastalarda kalp krizi tanısını koyabilmek, kalbi etkileyebilecek değişik ilaçların yan etkilerini ortaya çıkarmak ve mekanik aletlerin (kalıcı kalp pili vb.) fonksiyonlarını değerlendirmek amacıyla yapılmaktadır (Anonim 2014c).

#### **4.1.6 Maksimum kalp hızı**

Yaş ilerledikçe nabız sayısı düşerek kalp hızı azalır. Örneğin genç yaşındaki bir kişide maksimum kalp hızı 200 atım/dakikaya kadar artabilirken orta yaşındaki bir kişinin maksimum kalp hızı 176 atım/dakikaya iner. Bazen kalp hızının çok yavaş veya çok hızlı olması, koroner kalp hastalığının belirtisi olabilir.

#### **4.1.7 Talyum testi**

Talyum sintigrafisi, damardan verilen radyoaktif bir ilaç vasıtası ile kalp kasına gelen kan miktarını ölçerek hangi damarda sorun olduğunu saptama esasına dayanan ve kalp hastalıklarının tanısında kullanılan bir tetkiktir (Anonim 2010a).

#### **4.1.8 Cinsiyet**

Değiştirilebilir risk faktörlerin yanı sıra yaş ve cinsiyet gibi değiştirilmeyen risk faktörlerde vardır. Cinsiyet risk faktörleri incelendiğinde koroner arter hastalıklarına bağlı ölümler, erkeklerde kadınlara göre 3 kat daha fazladır ve ayrıca koroner hastalığı erkeklerde daha sık ve erken yaşlarda görülür.

#### **4.2 Cleveland Veri Seti**

Bu çalışmada Cleveland kalp hastalıkları veri seti kullanılmıştır. Bu veri seti California Üniversitesi UCI makine öğrenmesi veri tabanından (<http://archive.ics.uci.edu/ml/>) elde edilmiştir. Veriler California Üniversitesinden, David Aha tarafından toplanmıştır. Bu veri seti sağlıklı ve sağlıksız kişilerden alınan 303 örnekten oluşmuştur. Veri kümesi içinde sağlıklı ve sağlıksız hastaların sayısı sırasıyla, 164 ve 139 dur. Bu veri setinde her örnek için, 13 tan değişken kullanılmıştır. Değişkenler; yaş, cinsiyet, göğüs ağrısı, kan basıncı, kolesterol, kan şekeri, EKG, maksimum kalp hızı, Egzersize bağlı anjina, old peak, Stslope, ca ve talyum testidir. Bu değişkenlerin bilgileri çizelge 4.1'de bulunmaktadır.

Çizelge 4.1 Cleveland veri setinde kullanılan 13 değişkenin açıklaması

|    | Değişken adı | Min | Max | Açıklama                                                                               |
|----|--------------|-----|-----|----------------------------------------------------------------------------------------|
| 1  | Yaş          | 29  | 77  | Yaş (Yıl)                                                                              |
| 2  | Cinsiyet     | 0   | 1   | Cinsiyet ( 0 değeri, Kadın ve 1 değeri, Erkek olarak nitelendirilmiştir.               |
| 3  | Cp           | 1   | 4   | Göğüs ağrısı tipleri (Tipik=1, Atipik=2, Non kardiyak göğüs ağrısı=3, Asemptomatik=4 ) |
| 4  | Trestbps     | 94  | 200 | Kan basıncı                                                                            |
| 5  | Chol         | 126 | 564 | Kolesterol                                                                             |
| 6  | Fbs          | 0   | 1   | Kan şekeri (0=yanlış, 1=doğru) Eğer fbs>120 ozaman doğrudur.                           |
| 7  | Restecg      | 0   | 2   | EkG                                                                                    |
| 8  | Thalach      | 71  | 202 | Maksimum kalp hızı                                                                     |
| 9  | Exang        | 0   | 1   | Egzersize bağlı anjina (0=yanlış, 1=doğru)                                             |
| 10 | Old peak     | 0   | 6,2 | Egzersiz testinde oluşan ST segment depresyonu                                         |
| 11 | Stslope      | 1   | 3   | Egzersiz ile üç ST segment tipi (Yukarı eğimli=1, Düz=2, Aşağı eğimli=3)               |
| 12 | ca           | 0   | 3   | Floroskopi ile görülen ata damarların sayısı (0-3)                                     |
| 13 | Thal         | 3   | 7   | Talyum testi (normal=1, sabit defect =2, reversible defect=3)                          |

## **5. TASARIM ve UYGULAMA**

### **5.1 Kalp Hastalığının Teşhisi İçin Bulanık Uzman Sistem Tasarımı**

Günümüzde bilgisayar teknolojisini kullanarak tıp alanında tanı, hastalıkların tedavisi ve hastaları takibe almak oldukça artmıştır. Bazen pek çok belirsiz risk faktörlerden dolayı, klasik olarak bakılan testler ile tam tanı koyabilmek uzman kişiler için bile zordur. Başka bir deyişle tıbbi kavramların ve bu kavramlar arasındaki ilişkilerin kesin olmayan doğası nedeniyle sağlık ve hastalık arasında hiçbir sınır mevcut değildir. Bu yüzden uzmanların doğru araçlarla, risk faktörleri göz önüne getirerek belli bir sonuç elde etmeleri gerekmektedir.

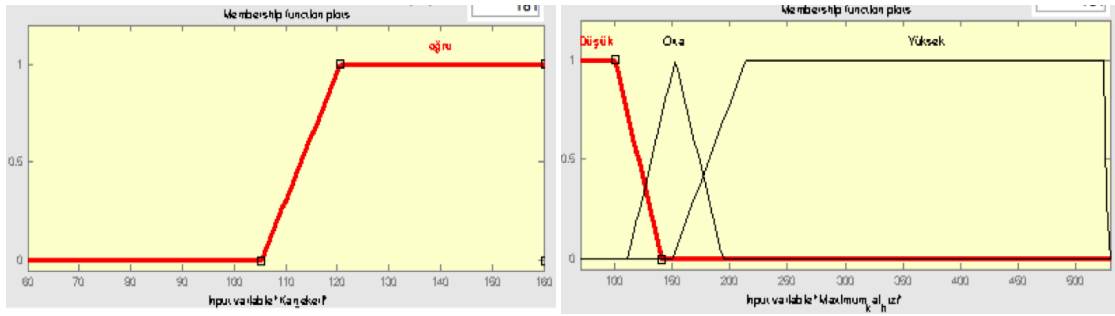
Bu bölümde kalp hastalıklarının teşhisi için bulanık uzman sistemi tasarlanacaktır ve bu sistem bulanık mantık kullanarak kural tabanı oluşturup ve böylece kalp hastalığının varlığı belirlenecektir.

Bu sistem, V.A.Medical Center, Long Beach ve Cleveland Clinic Foundation verilerini kullanarak tasarlanmıştır. Bu verilerin hedefi kalp hastalığının varlığını yada yokluğunu, hastanın yaptığı çeşitli tıbbi testlerin verilen sonuçlarına göre teşhis etmesidir. Bu çalışmada mamdani çıkarım sistemi uygulanmıştır ve böylece hastalık derecesi belirlenmiştir.

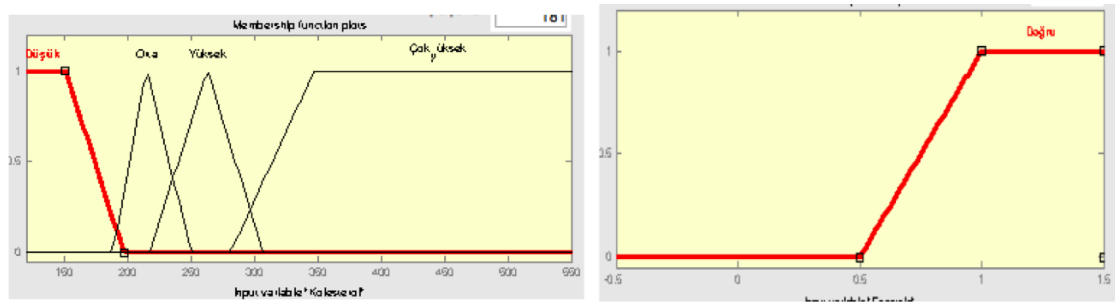
#### **5.1.1 Giriş ve çıkış üyelik fonksiyonlarının tasarımı**

İlk olarak bulanık kontrolde giriş ve çıkış değişkenleri belirlenir. Sonra tüm değişkenlerin üyelik fonksiyonları tanımlanır. Üyelik fonksiyonları, Matlab fuzzy toolbox kullanarak tasarlanmıştır. Bu sistem 11 tane girdi ve bir tane çıktıdan oluşturulmuştur. Giriş değişkenleri; göğüs ağrısı, yüksek tansiyon, kolesterol, kan şekeri, maksimum kalp hızı, egzersiz, old peak, Elektrokardiyografi (EKG), talyum testi, cinsiyet ve yaşıdır. Çıkış değişkeni ise hastada, kalp hastalığının varlığını (hastanın kalp hastalık derecesi) veya yokluğunu göstermektedir. Örnek olarak giriş değişkenlerinden, Göğüs ağrısı 4 çeşitten oluşmaktadır. Burada her hastanın sadece bir

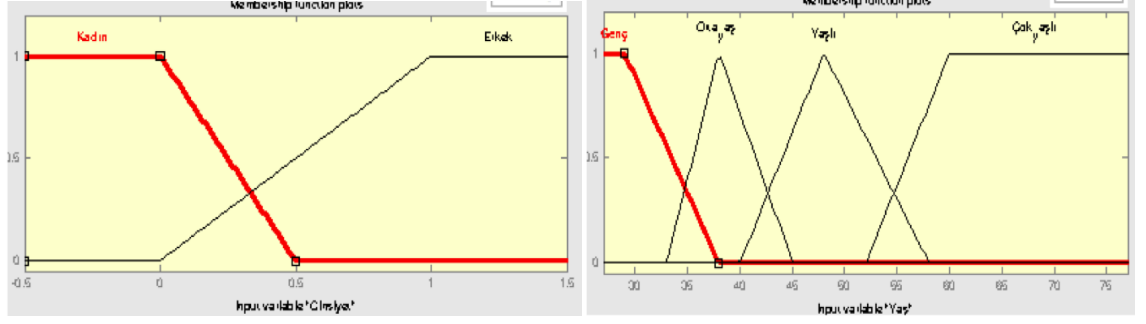
çeşit ağırlı ihtimali olmasından dolayı klasik küme şeklinde kullanılmaktadır. Ağırlı çeşitleri{1=tipik angina, 2=tipik olmayan angina, 3=olmayan angin ağrısı, 4=asymptomatic}. Bir diğer örnek kan basıncıdır. Bu girdi değişken, dört tane sözel terimden oluşmaktadır; Düşük, orta, yüksek ve çok yüksek. Şekil 5.1-5.6'de bu değişkenlerin grafiksel açıklamaları gösterilmektedir. Burada y eksen, bulanık değişkenin üyelik derecesini gösteriyor, x eksen ise sistemin girdilerini ifade ediyor. Üyelik fonksiyonları, Matlab fuzzy logic toolbox kullanarak tasarlanmıştır. Burada iki tip üyelik fonksiyonu kullanılmaktadır; üçgen ve yamuk. Her giriş ve çıkış değişkenleri kendine özgü dilsel değişkenlerden oluşur. Son olarak çıkış üyelik fonksiyonu; sağlıklı, hasta1, hasta 2 , hasta 3 ve hasta 4 dilsel değişkenlerden oluşmuştur.



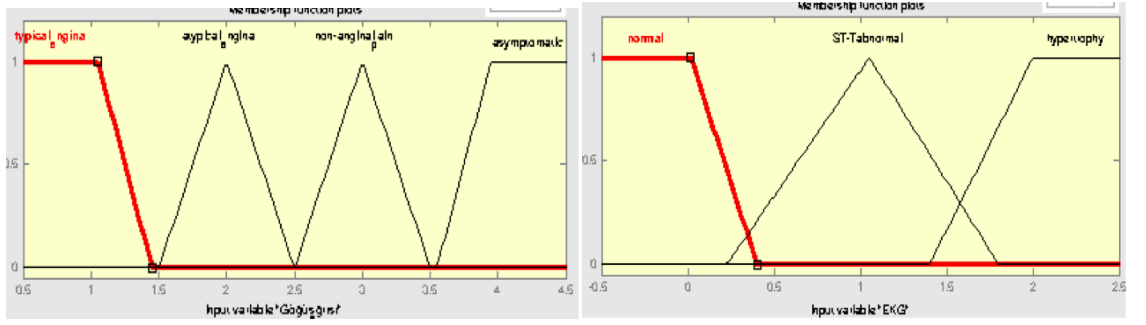
Şekil 5.1 Kan şekeri ve maksimum kalp hızı için üyelik fonksiyonu



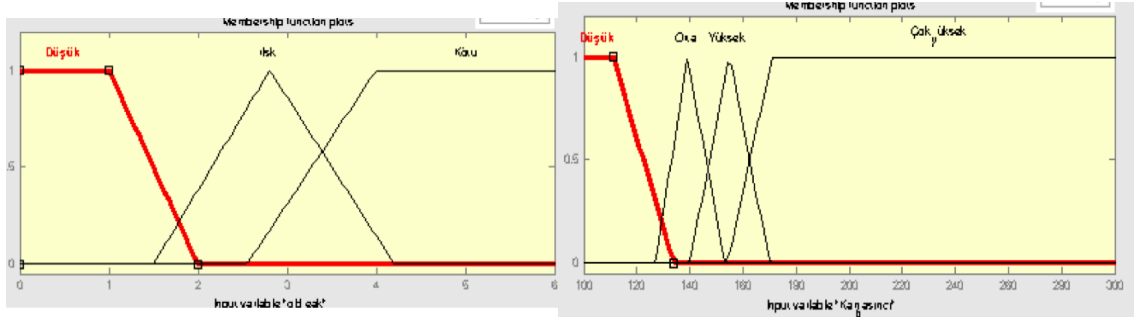
Şekil 5.2 Kolesterol ve egzersiz için üyelik fonksiyonları



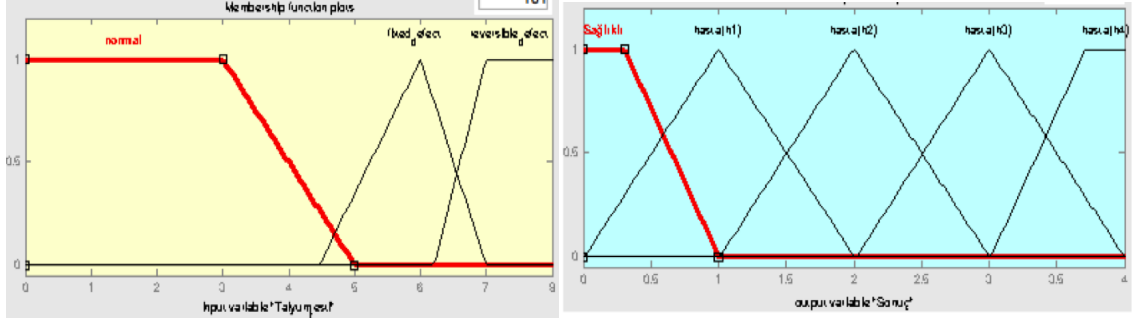
Şekil 5.3 Cinsiyet ve yaş için üyelik fonksiyonları



Şekil 5.4 Göğüs ağrısı ve EKG için üyelik fonksiyonları



Şekil 5.5 Old peak ve kan basıncı için üyelik fonksiyonları

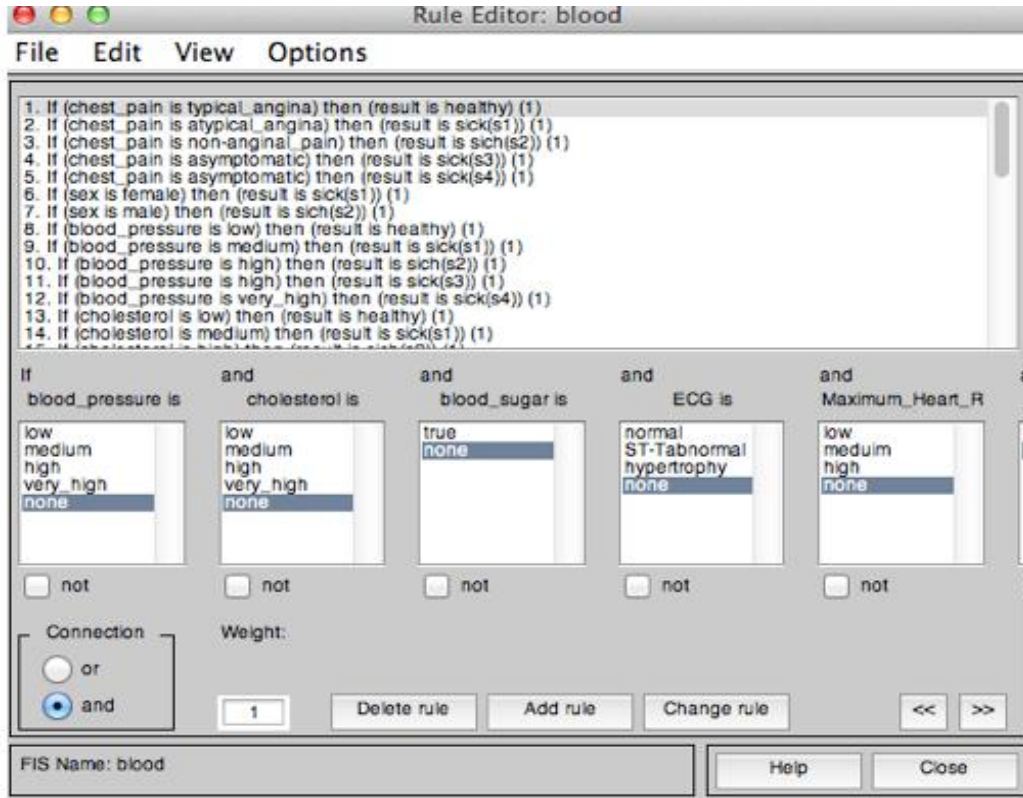


Şekil 5.6 Talyum testi ve hastalık teşhisi için üyelik fonksiyonları

Şekil 5.3’de görüldüğü gibi, Yaş 60’dan büyük olduğu halde, “çok yaşlı” bulanık kümesine tam üyelikle dahil olmaktadır ve “1” üyelik derecesine sahiptir. Aynı şekilde hastanın yaşı 48 olduğu halde, “yaşlı” bulanık kümesine tam üyelikle dahil olur, fakat 48’den büyük veya küçük olduğu halde, üyelik derecesi küçülmeye başlar. Aynı değerin iki farklı bulanık kümeyle kesişmesi mümkündür, örneğin 43 sayısı, “orta yaşlı” ve “yaşlı” bulanık kümeleriyle kesişmektedir ama her biri için farklı üyelik derecelerine sahiptir. Üyelik derecelerini bulmak için eşitlik 2.2 ve 2.3’deki formüller kullanılır.

### 5.1.2 Kural tabanı

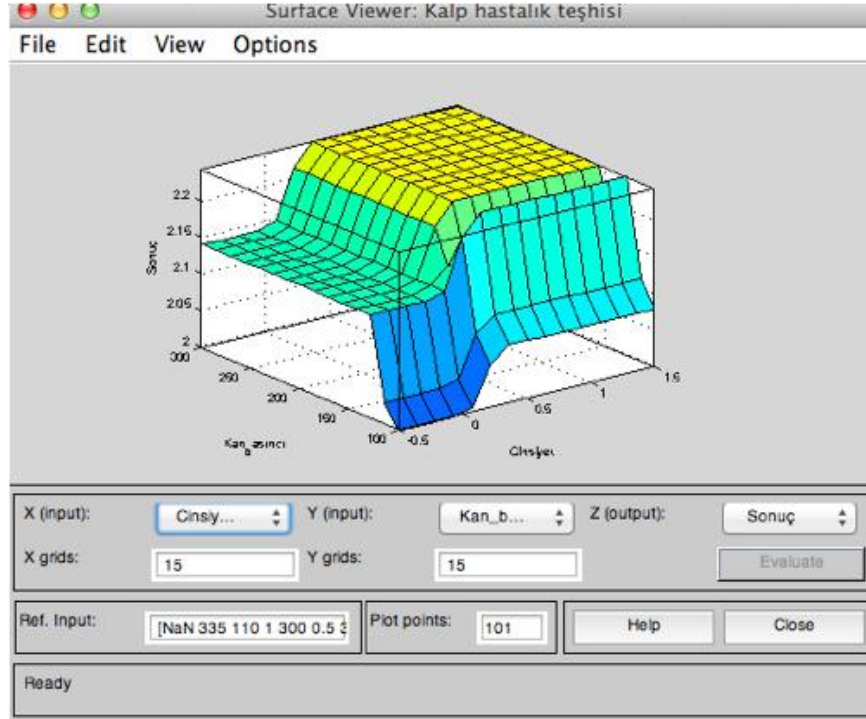
Kurallar, bulanık denetleyicilerin performansını etkileyen en önemli faktördür. Bulanık kuralların eğer- o halde şeklinde olduklarını daha önce de açıklanmıştır. Kalp hastalık teşhisi için 44 tane kural kullanılmıştır. Bulanık kurallar, insan zekasındaki düşünme gücünü kullanarak doğru çözüme ulaşmaya çalışır. Şekil 5.7’de Matlab fuzzy toolbox ile tasarlanan kural tabanı gösterilmektedir.



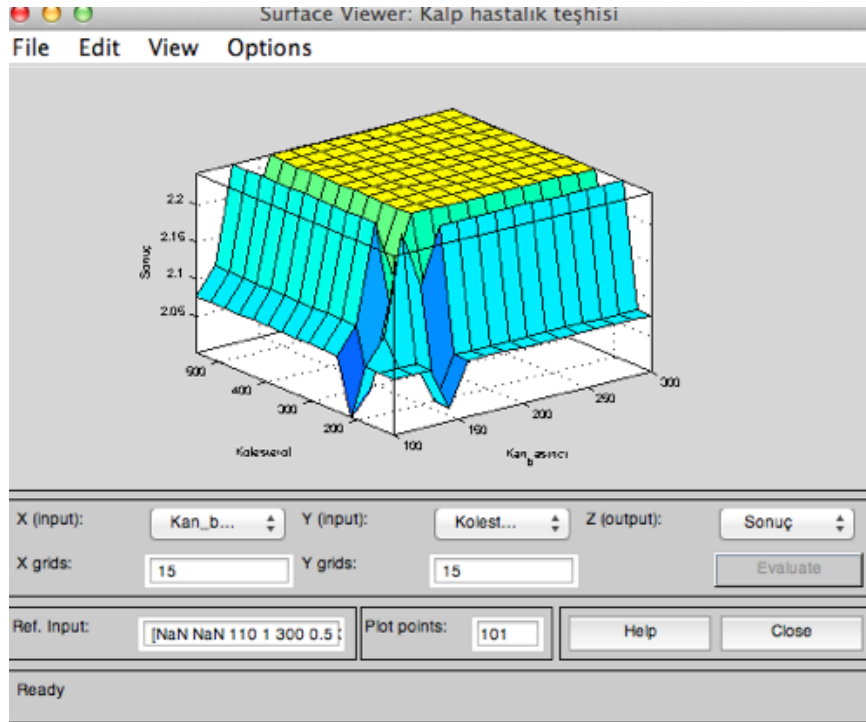
Şekil 5.7 Matlab fuzzy toolbox ile oluşturulan kural tabanı

### 5.1.3 Bulanık çıkarım ve durulaştırma

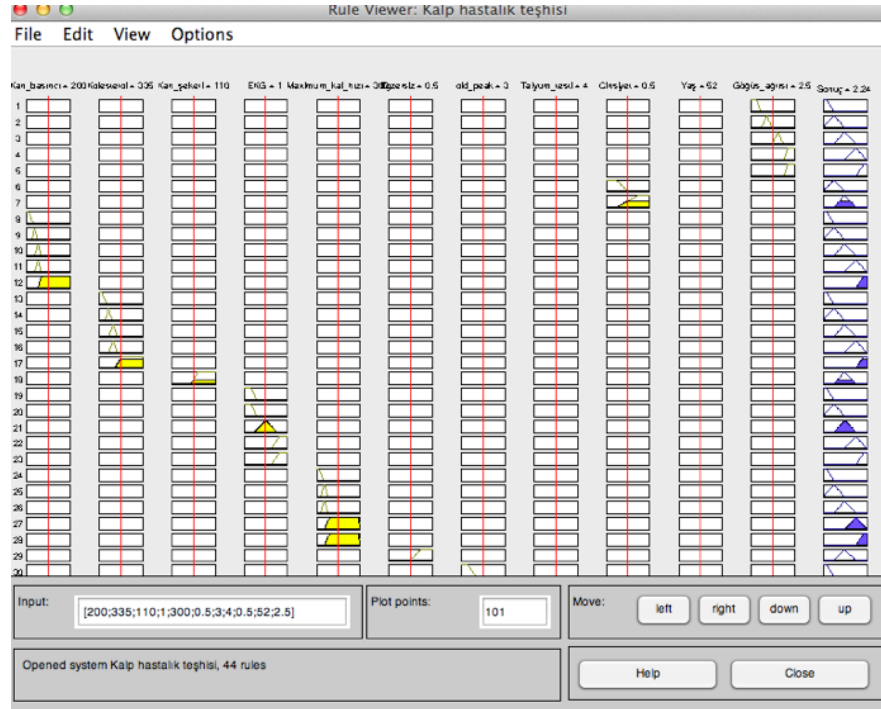
Bu sistemde Mamdani çıkarımı ve durulaştırma yöntemlerinden Ağırlık merkezi yöntemi kullanılmıştır.



Şekil 5.8 Mamdani yöntemi ile cinsiyet ve kan basıncı yönlerindeki değişikliklerin hastalık sonucu üzerindeki etkisi



Şekil 5.9 Mamdani yöntemi ile kolesterol ve kan basıncı yönlerindeki değişikliklerin hastalık sonucu üzerindeki etkisi



Şekil 5.10 Matlab fuzzy toolbox Rule Viewer ara yüzü

Bu sistem, Cleveland Clinic Foundation verilerini kullanarak tasarlanmıştır. 6 hasta bilgisi, yukarıdaki 11 risk faktörleri ile çizelge 5.1’de verilmiştir.

Çizelge 5.1 Bulanık çıkarım sistemi ile elde edilen sonuçlar

| Hastalar | Göğüs ağrısı | Kan basıncı | Kolesterol | Kan şekeri | EKG | MKH | Egzersiz | Old peak | Talyum testi | Cinsiyet | Yaş | Tehis  |
|----------|--------------|-------------|------------|------------|-----|-----|----------|----------|--------------|----------|-----|--------|
| 1        | 4            | 117         | 230        | 1          | 0   | 160 | 1        | 1.4      | 7            | 1        | 60  | Hasta2 |
| 2        | 4            | 120         | 229        | 0          | 2   | 129 | 1        | 2.6      | 7            | 1        | 67  | Hasta1 |
| 3        | 3            | 130         | 250        | 0          | 0   | 187 | 0        | 3.5      | 3            | 1        | 37  | Hasta0 |
| 4        | 2            | 130         | 204        | 0          | 2   | 172 | 0        | 1.4      | 3            | 0        | 41  | Hasta0 |
| 5        | 2            | 120         | 236        | 0          | 0   | 178 | 0        | 0.8      | 3            | 1        | 56  | Hasta0 |
| 6        | 4            | 140         | 268        | 0          | 2   | 160 | 0        | 3.6      | 3            | 0        | 62  | Hasta3 |

#### **5.1.4 Sonuç deęerlendirme**

Günümüzde bir çok kanser ve hastalıkların tanısını koyabilmek için detaylı tetkik gerekmektedir. Klasik olarak bakılan testler ile tam tanı koymak oldukça zordur. Tıpta kullanılan çoęu kavram bulanıktır. Tıbbi kavramların ve bu kavramlar arasındaki ilişkilerin kesin olmayan doğası nedeniyle bulanık mantık yöntemi tıbbi uygulamalar için uygundur. Kesin olmayan tıbbi durumlar bulanık kümelerle tanımlanabilir. Bulanık mantık yaklaşık sonuç çıkarma yeteneğine sahip çözüm üretme metotları önermektedir. Bulanık mantık tıp alanında önemli bir rol oynamaktadır ve pek çok tıbbi uygulamada araştırılmıştır. Bu çalışmada kalp hastalıklarının teşhisi için, üyelik fonksiyonları, giriş deęişkenleri, çıkış deęişkeni ve kural tabanı ile bulanık uzman sistemi tasarlanmıştır. Çalışmadaki simülasyon sonuçları bakıldığında kalp hastalığının teşhisi için kullanılan mamdani çıkarım yöntemi ile iyi sonuçlar elde edilmiştir.

#### **5.2 ANFİS ile Teşhis**

Bu bölümde ANFİS modeli kullanılarak kalp hastalıklarının teşhisi yapılmaktadır.

##### **5.2.1 Veri toplama**

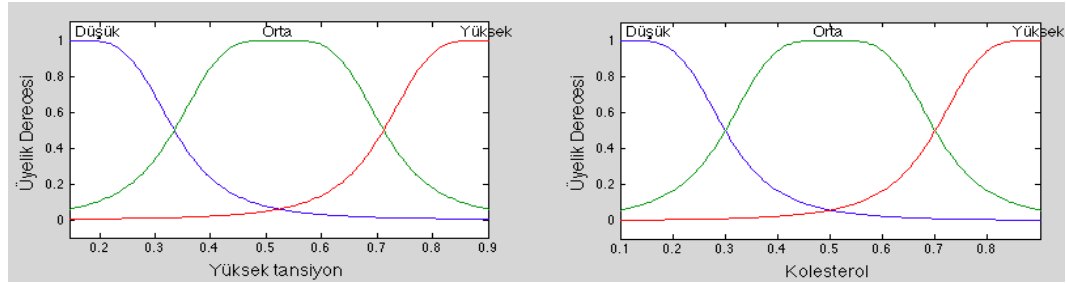
İlk aşamada veriler Cleveland kalp hastalığı veri setinden toplandı ve normalizasyon yöntemi ile MATLAB ortamında kullanmak üzere uygun bir şekle dönüştürüldü. Daha sonra giriş deęişkenleri, üyelik fonksiyonları ile birlikte, ANFİS girişine tanıtıldı. Gaussian üyelik fonksiyon tipi bulanık küme tanımı için kullanıldı. Kalp hastalığı için seçilen parametreler; hastanın yaşı, göęüs ağrısı, tansiyon, kolesterol, şekeri, elektrokardiyografi ve kalp hızı verilerine baęlı olarak kalp hastalığının varlığı (hastanın kalp hastalık derecesi) veya hastalığın yokluğu tahmin edilmiştir.

Çizelge 5.2 ANFİS için kullanılan üyelik derecelerinin sayısı

| Değişken Adı | Min | Max | Üyelik fonksiyonlarının sayısı |
|--------------|-----|-----|--------------------------------|
| age          | 29  | 77  | 3                              |
| cp           | 1   | 4   | 4                              |
| trestbps     | 94  | 200 | 3                              |
| chol         | 126 | 564 | 3                              |
| fbs          | 0   | 1   | 2                              |
| restecg      | 0   | 1   | 3                              |
| talach       | 71  | 202 | 2                              |

### 5.2.2 ANFİS modelin tasarımı

ANFİS mimarisi beş katmanlı ileri beslemeli sinir ağı içeren ve ayrıca sistematik bir yaklaşım sağlamak için belirli bir giriş-çıkış veri seti ile bulanık kurallar üreten Sugeno tipi bulanık çıkarım sistemi uygular. İlk aşamada, kalp hastalıkları için bulanık çıkarım sistemi oluşturmuştur. Eğitim süreci başlamadan önce, bulanık çıkarım sisteminin giriş parametreleri belirlenir ve her girişe atanan üyelik fonksiyonlarının sayısı seçilir. Bu çalışmadaki giriş değişkenler ve üyelik fonksiyonların sayısı çizelge 5.2'de gösterilmiştir. Soncul kısımdaki çıktı değişkeni, girdi değişkenlerinin lineer bir fonksiyonu yada sabit bir fonksiyon şeklindeki üyelik fonksiyonuna sahiptir (Köse vd. 2011). ANFİS model için kural tablosu, Grid Partition yöntemi kullanılarak, tüm girdilerin üyelik fonksiyonlarının tüm olası kombinasyonunu sıralanarak, oluşturulmuştur (Hamdan 2013). Şekil 5.11'de yüksek tansiyon ve kolesterol girişleri için başlangıç üyelik fonksiyonları gösterilmiştir.



Şekil 5.11 Üyelik fonksiyonları eğitimden önce

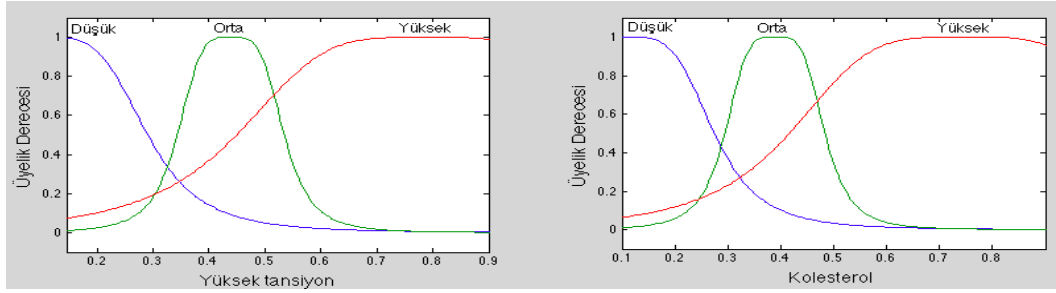
Şekil 5.12’de görüldüğü gibi bulanık çıkarım sistemindeki tüm kuralların öncül kısmı, değişkenlerin tümünü içerir. ANFİS, çıkış katmanında tek bir sonuç üretir.

|     |                                                                                                                                                                                    |     |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 1.  | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is low) and (Blood_sugar is false) and (ECG is normal) and (Max_heart_rate is 100)       | 0.9 |
| 2.  | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is low) and (Blood_sugar is false) and (ECG is normal) and (Max_heart_rate is 110)       | 0.8 |
| 3.  | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is low) and (Blood_sugar is false) and (ECG is ST_abnormal) and (Max_heart_rate is 100)  | 0.7 |
| 4.  | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is low) and (Blood_sugar is false) and (ECG is ST_abnormal) and (Max_heart_rate is 110)  | 0.6 |
| 5.  | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is low) and (Blood_sugar is true) and (ECG is normal) and (Max_heart_rate is 100)        | 0.5 |
| 6.  | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is low) and (Blood_sugar is true) and (ECG is normal) and (Max_heart_rate is 110)        | 0.4 |
| 7.  | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is low) and (Blood_sugar is true) and (ECG is ST_abnormal) and (Max_heart_rate is 100)   | 0.3 |
| 8.  | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is low) and (Blood_sugar is true) and (ECG is ST_abnormal) and (Max_heart_rate is 110)   | 0.2 |
| 9.  | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is high) and (Blood_sugar is false) and (ECG is normal) and (Max_heart_rate is 100)      | 0.1 |
| 10. | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is high) and (Blood_sugar is false) and (ECG is normal) and (Max_heart_rate is 110)      | 0.0 |
| 11. | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is high) and (Blood_sugar is false) and (ECG is ST_abnormal) and (Max_heart_rate is 100) | 0.9 |
| 12. | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is high) and (Blood_sugar is false) and (ECG is ST_abnormal) and (Max_heart_rate is 110) | 0.8 |
| 13. | If (Age is young) and (chest_pain is angina) and (Blood_pressure is high) and (Cholesterol is low) and (Blood_sugar is true) and (ECG is normal) and (Max_heart_rate is 100)       | 0.7 |
| 14. | If (Age is young) and (chest_pain is angina) and (Blood_pressure is high) and (Cholesterol is low) and (Blood_sugar is true) and (ECG is normal) and (Max_heart_rate is 110)       | 0.6 |
| 15. | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is high) and (Blood_sugar is true) and (ECG is normal) and (Max_heart_rate is 100)       | 0.5 |
| 16. | If (Age is young) and (chest_pain is angina) and (Blood_pressure is low) and (Cholesterol is high) and (Blood_sugar is true) and (ECG is ST_abnormal) and (Max_heart_rate is 100)  | 0.4 |
| 17. | If (Age is young) and (chest_pain is angina) and (Blood_pressure is high) and (Cholesterol is low) and (Blood_sugar is false) and (ECG is normal) and (Max_heart_rate is 100)      | 0.3 |
| 18. | If (Age is young) and (chest_pain is angina) and (Blood_pressure is high) and (Cholesterol is low) and (Blood_sugar is false) and (ECG is normal) and (Max_heart_rate is 110)      | 0.2 |
| 19. | If (Age is young) and (chest_pain is angina) and (Blood_pressure is high) and (Cholesterol is low) and (Blood_sugar is false) and (ECG is ST_abnormal) and (Max_heart_rate is 100) | 0.1 |
| 20. | If (Age is young) and (chest_pain is angina) and (Blood_pressure is high) and (Cholesterol is low) and (Blood_sugar is false) and (ECG is ST_abnormal) and (Max_heart_rate is 110) | 0.0 |

Şekil 5.12 Kalp hastalık teşhisi için gelişen kural tablosu.

### 5.2.3 ANFİS modelin eğitimi

Bu çalışmada ANFİS modelin parametreleri tespit etmek için melez öğrenme algoritması kullanılmıştır. Melez algoritma, ileri ve geri besleme olarak iki geçiştten oluşmuştur. Verilerin eğitim sürecinde, ileri beslemede düğümlerin çıkışları katman 4 e kadar ilerler ve soncul parametreleri en küçük kareler tahmini ile belirlenerek hata sinyali her düğüm için hesaplanır. Geriye beslemede ise hata sinyalleri geriye yayılarak öncül parametreleri eğitilir. İleri ve geri geçişler tekrarlanarak öncül ve soncul parametreleri belirlenir. Bu çalışmada ANFİS modeli için, eğitim hata toleransı 0.01 ayarlanmıştır. Melez eğitim metodu ile öncül ve soncul parametreler güncellendikten sonra üyelik fonksiyonların aralık değerleri değişmektedir. Tansiyon ve kolesterol girişlerinin yeni üyelik fonksiyonları şekil 5.13’de gösterilmektedir.



Şekil 5.13 Üyelik fonksiyonları eğitimden sonra

#### 5.2.4 Cross validation yöntemi

Bu çalışmada elde edilen sonuçları test etmek için, Cross validation yöntemi kullanılmıştır. Bu yöntemde, veri seti rastgele eşit olarak k tane kümeye ayrılır ve yöntem k defa tekrarlanır. Her defasında kümelerin biri test verileri için kullanılır ve kalan diğer kümeler eğitim verileri için kullanılır (Kahramanli ve Allahverdi 2008). Bu çalışmada 10 Cross validation kullanılmıştır.

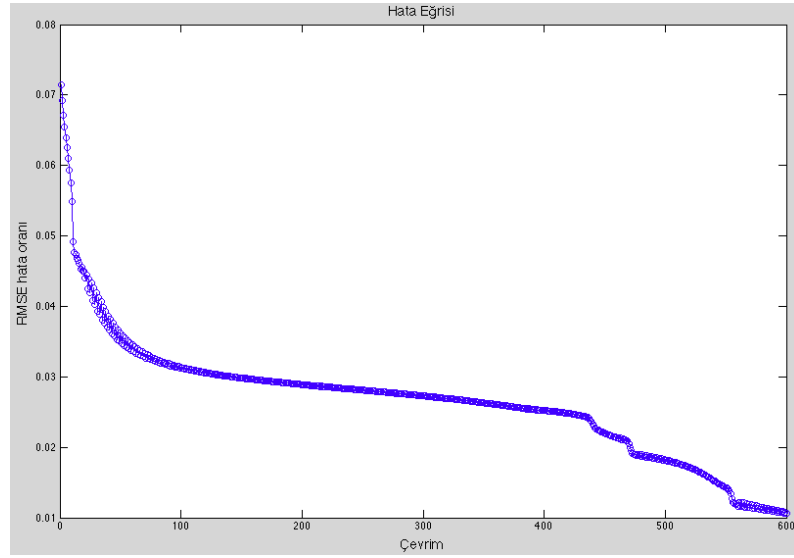
#### 5.2.5 Performansları değerlendirme

ANFİS modelin ne kadar iyi çalıştığını değerlendirmek için, modelin performansı hesaplanır. Bu çalışmadaki modelin performanslarını değerlendirmek için doğruluk derecesi hesaplanmıştır. Doğruluk derecesi, doğru sınıflanan örneklerin, toplam örneklerin oranı ile hesaplanır. Doğruluk derecesi, tıbbi tanı görevleri kullanılan ortak performans ölçütüdür. ANFİS çıkışındaki değerlerin incelenebilmesi için gerçek tanı sonuçları bilinmelidir.

#### 5.2.6 Sonuçların analizi

ANFİS modeli oluşturmak ve eğitime ve test etme işlemleri Matlab 7.0 kullanılarak gerçekleştirilmiştir. Başlangıçta ANFİS modeli oluşturmak ve modelin doğruluğunu test etmek için, Cleveland veri setinden elde edilen 303 adet veri iki guruba ayrılarak eğitim ve test verileri oluşturulmuştur. Bu çalışmada eğitim verileri için 243 adet veri, test verileri için ise 60 adet veri kullanılmıştır. Eğitim için melez öğrenme metodu

kullanılmış ve eğitim hata toleransı 0.01 ayarlanmıştır. Yapılan eğitimde kalp hastalığının çıkışı için 600 çevrim sonucunda 0.01 RMSE hata oranına ulaşılmıştır. Önerilen sistem, 60 örnek ile test edilmiştir ve ANFİS çıkışı için oluşan RMSE hata değeri 0.15 olarak bulunmuştur. Görüldüğü gibi hata oranı çok düşük bir seviyededir. Şekil 5.14’de 600 çevrim sonucunda oluşan hata oranı gösterilmektedir. Çizelge 5.3’de ise eğitim ve test verisi sonuçları verilmiştir.



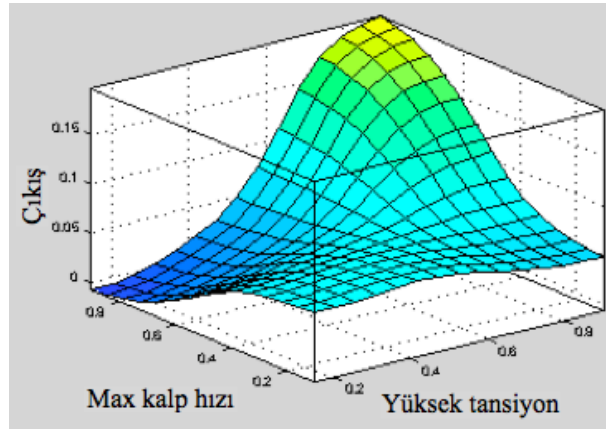
Şekil 5.14 Hata oranı

Çizelge 5.3 Hata değerleri

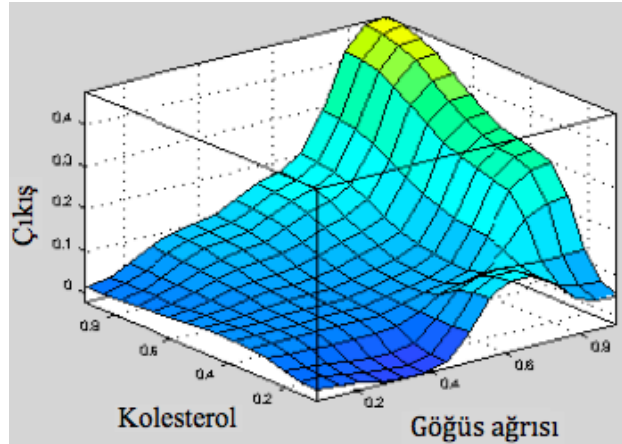
| ANFİS         |      |
|---------------|------|
| Eğitim hatası | 0.01 |
| Test hatası   | 0.15 |

### 5.2.7 Sonuç yüzeyi

ANFİS kontrol mekanizmasında uygulanan melez eğitim metot sonucunda mevcut üyelik fonksiyonları oluşturan öncül parametreler, kalp hastalığının derecesine göre güncellenerek yeniden şekillenmiştir. Daha sonra yeni üyelik fonksiyonları ile kural tabanı oluşturulmuştur ve ağırlıklı ortalama yöntemiyle çıkış değeri hesaplanmıştır. Sonuç yüzeyi, bu çıkış değerini giriş değerlerine bağlayan üç boyutlu bir eğridir. Şekil 5.15-5.16 eğitimden sonra sonuç yüzeyleri gösterilmektedir.



Şekil 5.15 ANFIS kontrol sistemi sonucu için maksimum kalp hızı ve tansiyon üyelik fonksiyonları



Şekil 5.16 ANFIS kontrol sistemi sonucu için kolesterol ve göğüs ağrısı üyelik fonksiyonları

Doğru bir teşhis koyabilmek amacıyla, kalp hastalıkları için tasarlanan ANFİS modelin doğruluk derecesi belirlenmiştir ve elde edilen sonuç, Cleveland kalp hastalığı veri seti üzerinde kullanılan önceki yöntemleri ile karşılaştırılmıştır. Kullanılan bu veri setinde kalp hastalıklarının teşhisi için daha güvenilir doğruluk derecesi bulmak amacıyla, çeşitli veri madenciliği teknikleri ile uygulamalar yapılmıştır. Anonim (2010b) tarafından belirtilen rapora göre, bu veri seti üzerinde uygulanan yöntemler ile elde edilen sınıflandırma doğruluk derecesi % 46.2 ve % 90 arasında değişilmektedir. Çizelge 5.4’de Cleveland veri setini kullanan çeşitli veri madenciliği teknikleri göstermektedir. Melezleştirilmiş veri madenciliği teknikleri kalp hastalığı teşhisinde doğruluk derecesinin arttırıldığı gözlenmiştir. Uygulama sonuçlarına bakıldığında ANFİS yöntemi bu veri seti üzerinde diğer yöntemlerden daha yüksek başarıma sahip olduğu görülmüştür. Çizelge 5.4’ de gösterildiği gibi, Cleveland veri setini kullanan yöntemler arasında ANFİS yöntemi en yüksek doğruluğa sahiptir (Ziasabounchi ve Askerzade 2014a).

Çizelge 5.4 Cleveland veri seti için ANFİS ve diğer yöntemlerin karşılaştırılması (Anonim 2010)

| Metot                            | Doğruluk % | Referans                                |
|----------------------------------|------------|-----------------------------------------|
| Bu çalışmadaki önerilen yaklaşım | 92.3       |                                         |
| Inc Net                          | 90         | Nobert Jankowski                        |
| Hybrid system                    | 86.8       | Humar Kahramanli<br>& Novruz Allahverdi |
| 28-NN, stand, Euclid, 7 features | 85.1± 0.5  | WD KG                                   |
| LDA                              | 84.5       | Ster&Dobnikar                           |
| Fisher discriminate analysis     | 84.2       | Ster&Dobnikar                           |
| K=7, Euclid, std                 | 84.2± 6.6  | WD Gostminer                            |
| 16-NN, stand, Euclid             | 84± 0.6    | WD KG                                   |
| FSM, 82.4-84% on test only       | 84.0       | Rafal Adamczak                          |
| K=1:10, Manhattan, std           | 83.8±5.3   | WD, GostMiner                           |
| Native Bayes                     | 82.5–83.4  | Rafal; ster, Dobnikar                   |
| SNB                              | 83.1       | Ster&Dobnikar                           |

### **5.2.8 Sonuç deęerlendirmesi**

Tezin bu kısmında kalp hastalıklarında erken ve güvenilir tanı koyabilmek ve hastalara yardım amaçlı, ANFİS destekli uzman sistem tasarlanmıştır. ANFİS, sinir ağı ve bulanık mantığın avantajları bir arada kullanarak hastaları doğru guruplara sınıflandırmıştır. Önerilen Model'in, test hata oranı çok düşük ve tatmin edici 0.15 olarak belirlenmiştir. Ayrıca ANFİS'in doğruluk derecesi test etmek için, aynı veri setini kullanan önceki çalışmalardan elde edilen sonuçlarla kıyaslandırılmıştır. Sonuç olarak önerilen modelin yüksek doğruluk derecesine sahip olduğu gözlenmiştir. Dolayısıyla bu model, kalp hastalıklarının sınıflandırması için uygun bir modeldir.

## **5.3 Kümeleme Algoritmaları**

### **5.3.1 Kümele algoritmaları ile teşhis**

Çok boyutlu veri setlerde, makine öğrenmesi için ortaya çıkabilecek sorunlardan biri, overfitting riskidir. Ayrıca tahmin doğruluęu (Predictive Accuracy) azalmasına yol açabilecek dięer sorun, deęişkenlerin fazlalığıdır (Howley ve Madden 2006). Böylece, bu problemlerin üstesinden gelebilmek için, veri madencilięindeki çeşitli uygulamalarda boyut azaltma teknikleri kullanılır. PCA, boyut azaltma yöntemlerden birisidir. PCA denetimsiz boyut indirgeme için yaygın olarak kullanılan istatistiksel tekniktir ve ana fikri verilerin en büyük varyans yönlerini bulmaktır (Ding ve He 2004). Bu bölümde, Kümeleme yöntemlerinin performansını arttırmak amacıyla, PCA yöntemi kullanarak, K-means ve BCM kümeleme yöntemleri uygulanmıştır. Çalışmada 303 satır 13 sütundan ve kalp hastalığının varlığı veya yokluęunu belirten iki sınıftan oluşan Cleveland veri seti kullanılmıştır. Böylece K-means ve BCM algoritmaları 2 küme sayısı için uygulanmıştır. Bu çalışmada ilk olarak Cleveland kalp hastalıkları veri setinin boyutları, PCA yöntemi kullanarak azalmış ve daha az sayıdaki yeni deęişkenler kullanarak kümeleme algoritmaları uygulanmıştır. Yazılım ortamı olarak Matlab programı kullanılmıştır ve kovaryans matrisinin öz deęerleri ve öz vektörleri Matlab'da bulunan hazır fonksiyonlar ile hesaplanmıştır.

PCA ve kümeleme yöntemlerin adımları aşağıda belirtilmiştir:

1. Veri setin giriş matrisi oluşturulur.

2. Veri setine standardize etmek için her girdi değerini her bir değişkene karşı gelen sütunların aritmetik ortalaması çıkarılarak girdi setinin standart sapmasının oranı belirlenir.

$$N = (A - repmat(AMean, [n 1])) ./ repmat(AStd, [n 1]) \quad (5.1)$$

3.  $N$  kovaryans matrisi hesaplandıktan sonra, örnek kovaryans matrisin öz değer ve öz vektörleri hesaplanır.  $V$ , öz vektörler ve  $D$ , öz değerler olmak üzere,

$$[V D] = eig(cov(N)) \quad (5.2)$$

Burada matris  $V$  ve  $D$  sırasıyla temel bileşenlerin katsayıları ve temel bileşenlerin yönünde elde edilen varyansları içerir. Hesaplanan öz değerler kullanılarak her bir temel bileşenin toplam varyansın yüzde kaçını temsil ettiği hesaplanmıştır (Çilli ve Arıtan 2010).

4. Temel bileşenler, standartlaştırılmış veri ve temel bileşenlerin katsayıların çarpımı ile elde edilir.

5. Temel bileşenler, varyansların yüzdesi ve cumulative varyansları belirlendikten sonra, zayıf bileşenlerini eleyerek, büyük değişimi olan ilk bir kaç bileşen seçilir ve yeni  $Y$  matrisi oluşturulur.

6. K-means ve BCM kümeleme algoritmaları ile  $Y$ ,  $k=2$  tane kümeye bölünür.

### 5.3.2 Sonuçların analizi

Çalışmanın bu kısmında ilk olarak Cleveland veri setine PCA yöntemi ile boyut azaltma işlemi yapılmıştır ve daha sonra K-means ve BCM kümeleme algoritmaları uygulayarak veriler sağlıklı veya sağlıklı kümelere sınıflandırılmıştır. K-means ve BCM algoritmaları başlangıç küme merkez değerleri rastgele seçildiği için, bu başlangıç değerler her iki yöntemin performansını etkileyebilmektedir. Dolayısıyla daha iyi sonuçlar elde etmek için, algoritmaların birkaç defa uygulanması önerilmektedir. Bu çalışmada, K-means ve BCM algoritmaları, boyut azaltma işlemi yapılmadan önce ve sonraki uygulamaların sonuçları karşılaştırılmıştır. Çizelge 5.5’de elde edilen kümeleme sonuçları verilmektedir.

Çizelge 5.5 Metotların karşılaştırması

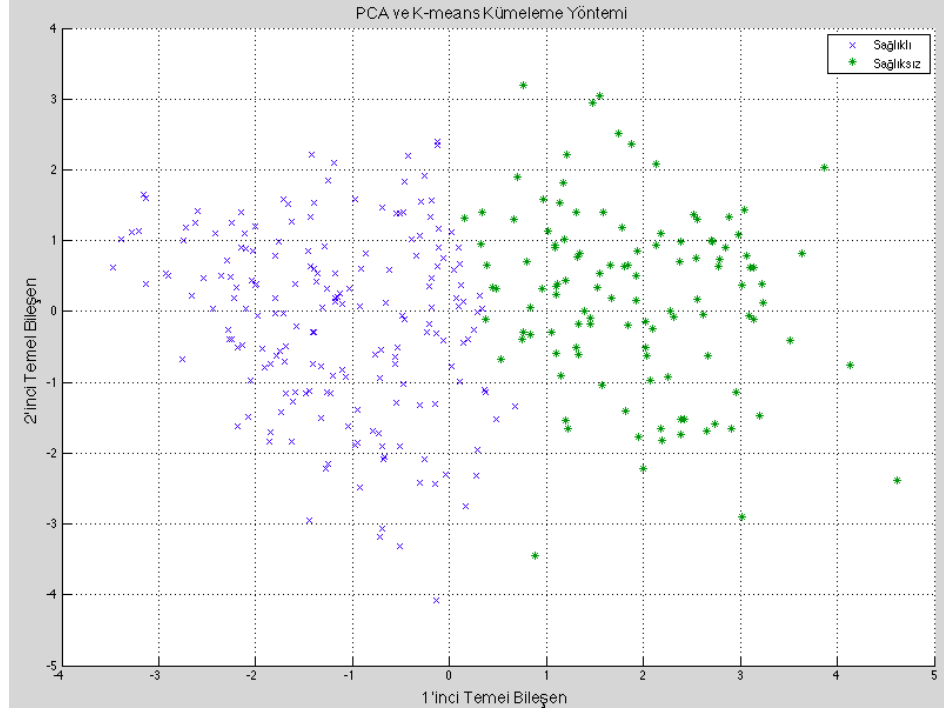
| Metot           | Karşılaştırma         |      |                       |
|-----------------|-----------------------|------|-----------------------|
|                 | Doğruluk derecesi (%) | RMSE | Hesaplama süreci(sec) |
| K-MEANS         | 81.0%                 | 0.45 | 0.28                  |
| BCM             | 80.0%                 | 0.46 | 0,45                  |
| PCA ile K-MEANS | 87.0%                 | 0.40 | 0.1                   |
| PCA ile BCM     | 82.0%                 | 0.44 | 0.32                  |

İlk sütun, her modelin sınıflandırma doğruluğunu temsil eder. Küme merkezleri belirlendikten sonra, her bir özellik vektör ve küme merkezleri arasındaki mesafeye göre, vektörler kendi kümelerine atanır (Hammouda ve Karray 2000). Daha sonra hata ölçüsü hesaplanır. Çizelge 5.5'nin ikinci sütununda görüldüğü gibi, Bu amaçla kök kare hata ortalaması (RMSE) kullanılmıştır. Yapılan çalışmadaki sonuçlara göre, PCA ve K-means kümeleme yöntemi diğer tekniklere göre daha yüksek doğruluk ve düşük RMSE üretir ve daha az hesaplama süresi gerektirir (Ziasabounchi ve Askerzade 2014b). Çizelge 5.6'de bu bölümdeki önerilen kümeleme yöntemlerin performansı ve PCA kullanmadan kümeleme yöntemlerin performansı karşılaştırılmıştır. Burada performans ölçütleri eşitlik 3.18, 2.19 ve 3.20 ile hesaplanmıştır. Her örnek, bu iki sınıftan birine atanır; sağlıklı (normal) veya sağlıklı (anormal).

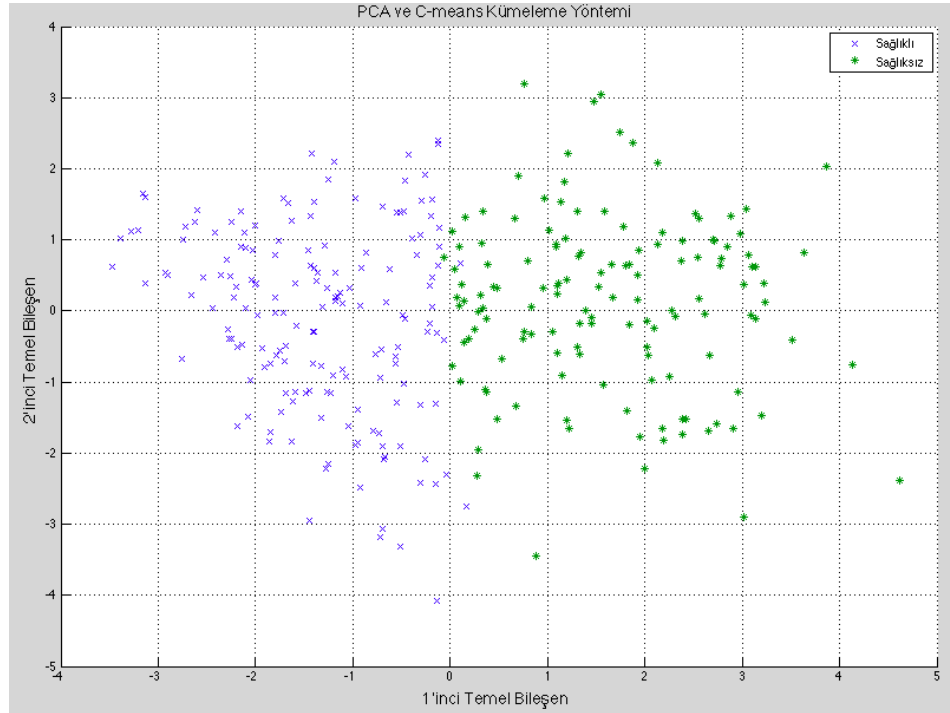
Çizelge 5.6 Sınıflandırma sonuçların analizi

| Metot                  | Örnek    | Kişi sayısı | Test sonucu<br>sağlıklı kişiler | Test sonucu<br>hasta kişiler | Performans ölçütleri |          |          |
|------------------------|----------|-------------|---------------------------------|------------------------------|----------------------|----------|----------|
|                        |          |             |                                 |                              | Duyarlılık           | Özgüllük | Doğruluk |
| <b>K-MEANS</b>         | Sağlıklı | 164         | 132                             | 32                           | 82.0%                | 80.0%    | 81.0%    |
|                        | Hasta    | 139         | 25                              | 114                          |                      |          |          |
| <b>BCM</b>             | Sağlıklı | 164         | 130                             | 34                           | 82.0%                | 79.0%    | 80.0%    |
|                        | Hasta    | 139         | 25                              | 114                          |                      |          |          |
| <b>PCA ile K-MEANS</b> | Sağlıklı | 164         | 151                             | 13                           | 81.0%                | 92.0%    | 87.0%    |
|                        | Hasta    | 139         | 26                              | 113                          |                      |          |          |
| <b>PCA ile BCM</b>     | Sağlıklı | 164         | 135                             | 29                           | 80.0%                | 82.0%    | 82.0%    |
|                        | Hasta    | 139         | 27                              | 112                          |                      |          |          |

Çizelge 5.6'deki sonuçlar değerlendirildiğinde, düşük boyutlu veri seti kullanan kümeleme yöntemleri, oldukça yüksek sınıflandırma doğruluğu ürettiği fark edilmiştir. Bu nedenle kümeleme algoritmaları, PCA yöntemi ile elde edilen daha düşük boyutlu Cleveland veri seti üzerinde daha etkili oldukları belirlenmiştir. Ayrıca PCA ile elde edilen yeni veri seti ile K-means algoritması diğer yöntemlere göre daha iyi sınıflandırma doğruluğu sağlamıştır. Boyut azaltma işleminden sonra uygulanan BCM algoritması ise K-means algoritmasına yakın sonuçlar verdiği gözlenmiştir (Ziasabounchi ve Askerzade 2014b). Burada ise örneklerin sınıflandırılması ilk iki temel bileşen kullanarak görselleştirilmiştir. Şekil 5.17'de PCA yoluyla K-means algoritması kullanarak, örneklerin hasta veya sağlıklı iki kümeden birine sınıflandırışını göstermektedir. Şekil 5.18'de ise PCA yoluyla BCM algoritması ile örnekler sınıflandırılmıştır.

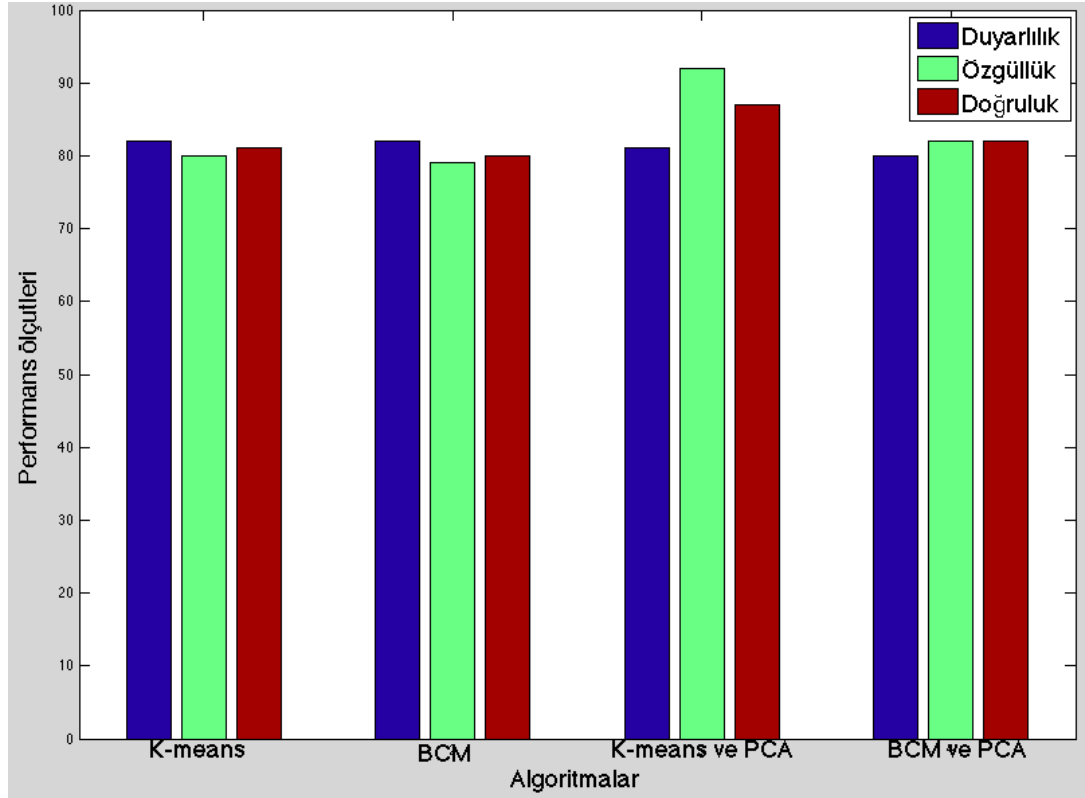


Şekil 5.17 PCA ve K-means yöntemi ile kümelenme sonuçları



Şekil 5.18 PCA ve BCM yöntemi ile kümelenme sonuçları

Şekil 5.19’de yukarıda bahsedildiği yöntemlerin duyarlılığı, özgüllüğü ve doğruluk ölçülerin karşılaştırması gösterilmiştir.



Şekil 5.19 Performans sonuçları

Şekil 5.19’deki duyarlılık, özgüllük ve doğruluk ölçütlerine bakıldığında, PCA ve K-means yöntemi diğer algoritmalara göre daha yüksek bir performans göstermektedir.

### 5.3.3 Sonuç değerlendirmesi

Tezin bu bölümünde PCA yöntemi ile kümeleme algoritmalarının performansı, Cleveland kalp hastalığı veri seti üzerinde analiz edilmiştir. Ayrıca kümeleme algoritmaları, PCA yöntemi kullanmadan, orijinal veri seti üzerinde de uygulanmıştır. Önerilen yöntemleri değerlendirmek için, tüm yöntemlerin performans ölçütleri hesaplanmış ve sonuçları karşılaştırılmıştır. Sonuçlara göre, PCA yöntemi ile veri seti üzerinde yapılan boyut azaltma işlemi, kümeleme algoritmalarının performansını artırmıştır. Ayrıca, yüksek boyutlu veri setlerinde, PCA boyut indirgeme yöntemi kullanarak verilerin görsel

gösterimine kolaylık sağlanmıştır. Cleveland veri seti üzerinde uygulanan yöntemlere bakıldığında, PCA ve K-means yöntemi 87% sınıflandırma doğruluğu ile diğer yöntemlere göre en yüksek ve tatmin edici sonuç elde etmiştir.

## 6. SONUÇ

Son yıllarda uzman sistemlerin kullanımı tıp alanında oldukça önemli bir konu olmuştur. Uzman sistemlerin amacı, insan kaynaklı hataları yok etmek, daha doğru teşhis koyabilmek ve doktorlara yardımcı olabilmektir. Bu tezde ANFİS destekli uzman sistem tasarlanarak, kalp hastalıklarının teşhisi rahatlıkla ve başarılı bir şekilde yapılmıştır. Bu çalışmada ayrıca K-means ve BCM kümeleme algoritmaları uygulanmış ve performansları değerlendirilmiştir. ANFİS yöntemi ile elde edilen doğruluk derecesi (%92.3) ile kümeleme yöntemlerin sonuçları (%81 ile %87 arasında) karşılaştırılmıştır. Bu doğrultuda ANFİS ile elde edilen sonuçlar, Cleveland veri setinin gerçek çıkış değerlerine daha yakın olduğu belirlenmiştir.

## KAYNAKLAR

- Anonim. 2010a. Talyum testi. <http://www.kalbinizesahipcikin.com/?islem=mn&ID=351&tag=%20Talyum%20testi.html>, Erişim Tarihi : 01.12.2013.
- Anonim. 2010b. Data set used for classification (comparison of results), <http://www.is.umk.pl/projects/datasets.html>, Erişim Tarihi : 12.02.2014.
- Anonim. 2011. Kalp hastalıkları. <http://www.florence.com.tr/kalp-damar-cerrahi.html>, Erişim Tarihi: 25.11.2014.
- Anonim. 2014a. Serebrovasküler hastalık (beyin-damar hastalıkları) [http://www.anh.gov.tr/index.php?option=com\\_content&view=article&id=540:serebrovaskueler-hastalk-beyin-damar-hastalklar&catid=27:sa-ki](http://www.anh.gov.tr/index.php?option=com_content&view=article&id=540:serebrovaskueler-hastalk-beyin-damar-hastalklar&catid=27:sa-ki), Erişim Tarihi: 25.11.2014.
- Anonim. 2014b. Yüksek tansiyon. <http://www.academichospital.com.tr/tr/yuksek-tansiyon>, Erişim Tarihi: 1.12.2013.
- Anonim. 2014c. Elektrokardiyograf nedir. <http://www.sanal-hastane.com/elektrokardiyografi-ekg-testi-nedir-nasil-yapilir>, Erişim Tarihi: 1.12.2013.
- Anonymous. 2014. Web sitesi: <http://www.who.org>. Erişim Tarihi: 9.2.2014.
- Baykal, N. ve Beyan, T. 2004. Bulanık mantık ilke ve temelleri. Bıçaklar Kitapevi, 413, Ankara.
- Bilge, U. 2007. Tıpta yapay zeka ve uzman sistemler. 4 Ulusal Tıp Bilişimi Kongresi, Antalya, s113-118.
- Çilli, M. ve Arıtan, S. 2010. Temel bileşenler analizi yardımı ile elde edilen daha az sayıda değişken kullanılarak farklı hızlarda insan koşusunun fourier tabanlı modelinin oluşturulması. Spor Bilimleri Dergisi, 21 (1), 1–12.
- Çobanoğlu, B. 2000. Bulanık mantık ve bulanık küme teorisi. <http://www.cobanoglu.fws1.com/bulanik.htm>, Erişim Tarihi: 15.10.2014.
- Ding, C. and He, X. 2004. K-means Clustering via Principal Component Analysis, International Conference on Machine Learning, Banff, Canada.
- Demiralay, M. ve Çamurcu, Y. 2005. Cure, Agnes ve K-means algoritmalarındaki kümeleme yeteneklerinin karşılaştırılması. İstanbul Ticaret Üniversitesi Fen Bilimler Dergisi, 4(8), 1-18.

- Ertorsun, A.D., Bağ, B., Uzar, G., Turanoğlu, M.A. ve Yazıcı, C. 2012. Web Sitesi : <http://tip.baskent.edu.tr/egitim/mezuniyetoncesi/calismagrp/ogrsmpzsnm12/10.2.pdf>. Erişim Tarihi: 11.08.2014.
- Fan, J., Han, M. and Wang, J. 2009. Single point iterative weighted fuzzy C-means clustering algorithm for remote sensing image segmentation. *Pattern Recognition*, 42(11), 2527–2540.
- Hamdan, H. 2013. An exploration of the adaptive neuro-fuzzy inference system (ANFIS) in modeling survival. PhD thesis, University of Nottingham, School of Computer Science, 194, Nottingham.
- Hammouda, K. and Karray, F. 2000. A comparative study of data clustering techniques, SYDE 625: Tools of intelligent system design, course project.
- Hıshammuddin, A. 2008. Fuzzy methodologies for automated university timetabling solution construction and evaluation. PhD thesis, Computer Science, University of Nottingham.
- Howley, T., Madden, G.M., Connell, M.L. and Ryder, G.A. 2006. The effect of principal component analysis on machine learning accuracy with high dimensional spectral data, *Knowledge Based Systems*, 19(5), 363–370.
- Işık, M. ve Çamurcu, Y. 2010. K-means ve aşırı küresel C-means algoritmaları ile belge madenciliği. *Fen Bilimler Dergisi*, 22, 1-18.
- Jang, J.S.R. and Sun, C.T. 1995. Neuro-fuzzy modeling and control. In *The Proceeding of the IEEE*, 83(3), 378–406.
- Jang, J.-S. R. 1993. ANFIS: Adaptive-network-based fuzzy inference system, *IEEE Transaction on System, Man and Cybernetics*, 23(3), 665–685.
- Jolliffe, I. 2002. *Principal Component Analysis*, Springer, 2nd edition.
- Kahramanli, H. and Allahverdi, N. 2008. Design of a hybrid system for the diabetes and heart disease, *Expert Systems with Applications*, 82-89.
- Köse, M., Terzi, Ö., İlker, A. ve Ergin, G. 2011. Uyarlamalı ağ tabanlı bulanık mantık çıkarım sistemi ile Kızılırmak Nehri'nin akım tahmini. 6th International Advanced Technologies Symposium.
- Maimon, O. and Rokach, L. 2010. *Data Mining and Knowledge Discovery Handbook*, Springer. 1985.
- Mahmood, S.M. 2010. Bulanık mantık kullanılarak trafik kontrolünün tasarımı ve uygulaması. Yüksek lisan tezi, Ankara Üniversitesi, Fen Bilimler Enstitüsü, Bilgisayar Anabilim Dalı, 68, Ankara.

- Mamdani, E.H. and Assilian, S. 1975. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1), 1-13.
- Neshat, M. and Yaghobi, M. 2009. Designing a Fuzzy Expert System of Diagnosing the Hepatitis B Intensity Rate and Comparing in with Adaptive Neural Network Fuzzy System, *Proceeding of the World Congress on Engineering and Computer Science*.
- Obanijesu, O. and Emuoyibofarhe, O. J. 2012. Development of Neuro-fuzzy System for Early Prediction of Heart Attack, *International Journal of Information Technology and Computer Science*, 9, 22-28.
- Ok, Y., Atak, M. ve Akçayol, M.A. 2011. Yalın sinirsel bulanık bir model ile İMKB 100 endeksi tahmini. *Gazi Üniversitesi, Mühendislik ve Mimarlık Fakültesi Dergisi*, 26(4), 897-904.
- Özçalık, H.R. ve Uygur, A.F. 2003. Dinamik sistemlerin uyumlu sinirsel-bulanık ağ yapısına dayalı etkin modellenmesi. *KSÜ Fen ve Mühendislik Dergisi*, 6(1), 36-46.
- Ross, T. 1995. *Fuzzy logic with engineering applications*, McGraw-Hill , New York.
- Sugeno, M. 1985. *Industrial application of fuzzy control*. Elsevier Science Pub.
- Subbuthai, P., Periasamy, A. and Muruganand, S. 2012. Identifying the character by applying PCA method using Matlab, *International Journal of Computer Applications*, 60(1), 8-11.
- Şentürk, S. 2010. Faktöriyel tasarıma adaptif ağ tabanlı bulanık mantık çıkarım sistemi ile farklı bir yaklaşım. *Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22, 57-74.
- Uzun, E. 2007. İnternet tabanlı bilgi erişimi destekli bir otomatik öğrenme sistemi. *Doktora Tezi, Trakya Üniversitesi, Fen Bilimler Enstitüsü, Bilgisayar Anabilim Dalı*, 134, Ankara.
- Ünal, Y., Ekim, U. ve Köklü, M. 2011. Üniversite öğrencilerin ortak zorunlu derslerdeki başarılarının K-means algoritması ile incelenmesi. *Journal of New World Sciences Academy*, 6(1), 342-347.
- Yıldız, B. ve Gedik, H. 2004. Bulanık bütçeleme ve bulanık bütçe kontrolü. *Sosyal Bilimler Dergisi*, 5(2), 141-166.
- Yılmaz, M. ve Arslan, E. 2005. Bulanık mantığın jeodezik problemlerin çözümünde kullanılması. *Harita ve Kadastro Mühendislik Ölçmeleri Sempozyumu*, 23-25 Kasım, İTÜ, İstanbul.
- Zadeh, L.A. 1965. Fuzzy sets. *Information and Control*, 8(3), 338–353.

- Zadeh, L.A. 1975. The concept of a linguistic variable and its application to approximate reasoning - I, II and III. *Information Sciences*, 8;8;9, 199–249; 301–357; 43– 80.
- Ziasabounchi, N. and Askerzade, İ. 2014a. Anfis based classification model for heart disease prediction. *International Journal of Engineering and Computer Science*, 14(2), 7-12.
- Ziasabounchi, N. and Askerzade, İ. 2014b. A comparative study of heart disease prediction based on principal component analysis and clustering methods. *Turkish Journal of Mathematics and Computer Science*.

## ÖZGEÇMİŞ

Adı Soyadı : Negar ZIASABOUNCHI  
Doğum Yeri : Urmiyeh/IRAN  
Doğum Tarihi : 01.12.1985  
Medeni Hali : Bekar  
Yabancı Dili : İngilizce/Farsça/Azerice

### Eğitim Durumu (Kurum ve Yıl)

Lise : Erem Lisesi (2004)  
Lisans : Tabriz Azad Üniversitesi, Bilgisayar Bölümü (2009)  
Yüksek lisans : Ankara Üniversitesi Fen Bilimler Enstitüsü, Bilgisayar Bölümü  
(Eylül 2011- Aralık 2014)

### Yayınları (SCI ve diğer)

Ziasabounchi, N. and Askerzade, İ. 2014. ANFIS based classification model for heart disease prediction. International Journal of Engineering & Computer Science, 14(2), 7-12.

Ziasabounchi, N. and Askerzade, İ. 2014. A comparative study of heart disease prediction based on principal component analysis and clustering methods. Turkish Journal of Mathematics and Computer Science.