



**T.C.  
NECMETTİN ERBAKAN  
ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**



**KALP HASTALIKLARINDA KULLANILAN  
YAPAY ZEKÂ TEKNİKLERİ  
VE  
UYGULAMALARI**

**Mehmet Emin ÇİFCİ**

**YÜKSEK LİSANS TEZİ  
Endüstri Mühendisliği Anabilim Dalı**

**Haziran-2019  
KONYA  
Her Hakkı Saklıdır**

## TEZ KABUL VE ONAYI

Mehmet Emin ÇİFCİ tarafından hazırlanan “Kalp Hastalıklarında Kullanılan Yapay Zekâ Teknikleri ve Uygulamaları” adlı tez çalışması 08/05/2019 tarihinde aşağıdaki jüri tarafından oy birliği ile Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

### Jüri Üyeleri

### İmza

#### Başkan

Doç. Dr. Halife KODAZ

#### Danışman

Dr. Öğr. Üyesi Onur İNAN

#### Üye

Dr. Öğr. Üyesi Şaban GÜLCÜ

Yukarıdaki sonucu onaylarım.

Prof. Dr. S. Savaş DURDURAN  
FBE Müdürü

## **TEZ BİLDİRİMİ**

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

## **DECLARATION PAGE**

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Mehmet Emin ÇİFCİ

08.05.2019

**ÖZET**  
**YÜKSEK LİSANS TEZİ**  
**KALP HASTALIKLARINDA KULLANILAN**  
**YAPAY ZEKÂ TEKNİKLERİ**  
**VE**  
**UYGULAMALARI**  
**Mehmet Emin ÇİFCİ**

**Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü**  
**Endüstri Mühendisliği Anabilim Dalı**  
**Danışman: Dr. Öğr. Üyesi Onur İNAN**

**2019, 66 Sayfa**

**Jüri**

**Dr. Öğr. Üyesi Onur İNAN**

Bu çalışmada, kalp hastalıklarının teşhis edilebilmesi için, Sınıf Nitelik Bağımlılık Maksimizasyonu (SNBM), Temel Bileşenler Analizi (TBA) ve Normalizasyon'dan oluşturulmuş bir Hibrit Sistem Yapay Sinir Ağları (YSA) ve Destek Vektör Makineleri (DVM) ile birleştirilerek Kalp Hastalığı Teşhisine yardımcı bir sistem geliştirilmiştir. Önerilen sistemde iki veri seti kullandık ilki UCI'den alınan kalp hastalığı veri setidir. Çalışmada 270 hastadan alınmış 13 öznitelik içeren veri seti tercih edilmiştir. Veri seti sınıflandırıcılara girmeden önce Hibrit Sistem (SNBM + TBA + Normalizasyon ) içerisinde ön işleme tabi tutulmuştur. Ön işleme için SNBM, TBA ve Normalizasyon algoritmalarının çeşitli versiyonları kullanılmıştır. Ön işleme adımından öznitelik değerleri YSA ve DVM sınıflandırıcılarına ayrı ayrı girmiştir. Hibrit sistem ile en iyi uyum sağlayacak sınıflandırıcıyı bulabilmek için 2 farklı sınıflandırıcı seçilmiştir. Sınıflandırıcıların doğruluk oranları 10-katmanlı çapraz doğrulama metodu ile test edilmiştir. Hibrit sistem ve YSA ile bulunan doğruluk oranı %88.89 olarak bulunmuştur. Hibrit sistem ve DVM sınıflandırıcısı ile bulunan doğruluk oranı %87.77 olarak bulunmuştur. Hibrit sistem kullanılmadan hesaplanan YSA sınıflandırıcısının doğruluk oranı %85.18 olarak bulunmuştur. Hibrit sistem kullanılmadan hesaplanan DVM sınıflandırıcısının doğruluk oranı %84.81 olarak bulunmuştur. Hibrit sistemin sınıflandırıcı sonuçlarına katkı sağlamış olduğu görülmüştür. Önerilen sistemde ikinci olarak yine UCI'den alınan TFEFT veri setiyle çalışılmıştır. Hibrit sistem kullanılmadan hesaplanan YSA sınıflandırıcısının doğruluk oranı %79.40 olarak bulunmuştur. TBA + SNBM + Normalizasyon + YSA sınıflandırıcısı ile bulunan doğruluk oranı %89.91 olarak bulunmuştur. Hibrit sistem kullanılmadan hesaplanan DVM sınıflandırıcısının doğruluk oranı %75.27 olarak bulunmuştur. TBA+SNBM+DVM sınıflandırıcısı ile bulunan doğruluk oranı %89.17 olarak bulunmuştur. Önerilen sistemde kullanılan yöntem ve sonuçlar, araştırmacılar tarafından aynı veri setiyle yapılan literatürdeki çalışmalarla karşılaştırılmıştır. Sonuçlar gelecek için umut verici görülmüştür.

**Anahtar Kelimeler:** Ayırıklaştırma Maksimizasyonu, Destek Vektör Makineleri, Hibrit Sistemler, Kalp Hastalıkları Teşhisi, Normalizasyon, Temel Bileşen Analizi, YSA

**ABSTRACT****MS THESIS****ARTIFICIAL INTELLIGENCE TECHNIQUES  
USED IN HEART DISEASES AND APPLICATIONS****Mehmet Emin ÇİFCİ****THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE  
OF  
NECMETTİN ERBAKAN UNIVERSITY  
THE DEGREE OF MASTER OF SCIENCE  
IN INDUSTRIAL ENGINEERING****Advisor: Assist. Prof. Dr. Onur İNAN****2019, 66 Pages****Jury****Advisor Assist. Prof. Dr. Onur İNAN**

In this study, a heart disease prediction system was developed by combining hybrid system as Class Attribute Interdependence Maximization (CAIM), Principle Component Analysis (PCA), Normalization and classifiers as Artificial Neural Network (ANN) and Support Vector Machines (SVM). In the proposed system, total of 270 heart disease records with 14 attributes from UCI were studied firstly. This dataset is pre-processed before loading to the classifiers. CAIM, PCA and Normalization algorithms were used sequentially for pre-processing. CAIM tags the dataset into new classes. PCA explores the relationship between attributes. Normalization normalizes dataset before inserting dataset into the classifier. After preprocessing, the dataset was entered separately into ANN and SVM classifiers separately. In order to find the best match for the hybrid system, two different classifiers were chosen. The ANN and SVM algorithms classify data loaded from the hybrid system into classes. Accuracy ratios of the classifiers were tested with 10-fold cross validation method. The accuracy ratio with the Hybrid System and ANN was 88.89%. In addition, the accuracy ratio of the Hybrid System and SVM classifier was 87.77%. The Hybrid System has been shown to contribute to the classifier results. Second data set studied with was SPECT. Hybrid System and ANN classifier's accuracy ratio was 89.91%. PCA + CAIM and SVM classifier's accuracy ratio was 89.17%. The methods and results used in the proposed system were compared with the studies conducted by the researchers in the literature with the same data set. The results were promising for the future.

**Keywords:** Artificial Intelligence, Classification and Regression, Dataset, Heart Diseases, Heart Failure Expert System, Hybrid System, Support Vector Machine, Neural Network, Principal Component Analyses

## ÖNSÖZ

Bu tez çalışmasında kalp hastalıklarının teşhisine yardımcı bir sistem önerilmiştir. Ön işleme adımının ardından sınıflandırıcılara giren verilerin doğruluk oranlarına bakılarak önerilen sistem hakkında yorum yapılmıştır. Sistemdeki ilk çalışma UCI'den alınmış 13 öznitelik içeren 270 kayıttan oluşmuş kalp hastalığı veri seti ile yapılmıştır. Sistemdeki ikinci çalışma yine UCI'den alınmış 44 öznitelik içeren 267 kayıttan oluşmuş TFEFT veri seti ile yapılmıştır. Önerilen sistem 2 kısımdan oluşmuştur. İlk kısım Hibrit Sistem olarak adlandırılmış ön işleme adımlarını göstermektedir. SNBM, TBA ve Normalizasyondan oluşmuştur. İkinci kısım ise sınıflandırıcılardan oluşmuştur. Sınıflandırıcılar ön işleme adımından gelen verileri sınıflandırmıştır. Kalp hastası olup olmama durumu program tarafından işletilen sınıflandırıcıdaki sınıf değerine bağlı olarak hesaplanmıştır. Sınıflandırıcıların doğruluk oranları k-katmanlı çapraz doğrulama metodu ile kontrol edilmiştir.

Sonuçlar aynı veri serini kullanan diğer çalışmalarla karşılaştırılmış ve sonuçların gelecek için umut verici olduğu görülmüştür.

Tez çalışmamdaki katkılarından dolayı Dr. Öğr. Üyesi Onur İNAN'a, Dr. Öğretim Üyesi Şaban GÜLCÜ'ye, Doç. Dr. Halife KODAZ'a ve Prof. Dr. Sabri KOÇER'e teşekkür ediyorum.

Verdikleri enerjiden dolayı kızlarım Hilal ve Gülce'ye de ayrıca teşekkür ediyorum.

Mehmet Emin ÇİFCİ

KONYA-2019

## İÇİNDEKİLER

<b>ÖZET</b> .....	<b>iii</b>
<b>ABSTRACT</b> .....	<b>iv</b>
<b>ÖNSÖZ</b> .....	<b>v</b>
<b>İÇİNDEKİLER</b> .....	<b>vi</b>
<b>SİMGELER VE KISALTMALAR</b> .....	<b>vii</b>
<b>1. GİRİŞ</b> .....	<b>1</b>
<b>2. KAYNAK ARAŞTIRMASI</b> .....	<b>7</b>
<b>3. MATERYAL VE YÖNTEM</b> .....	<b>13</b>
3.1. Veri Madenciliği.....	16
3.2. Sınıf Nitelik Bağımlılığı Maksimizasyonu.....	20
3.3. Temel Bileşenler Analizi.....	22
3.4. Veri Normalizasyonu .....	25
3.5. Sınıflandırıcılar .....	26
3.6. Sınıflandırıcı Türleri.....	26
3.6.1. Yapay Sinir Ağları .....	26
3.6.1.1. Eğitim Algoritması .....	31
3.6.1.1.1. Sigmoid fonksiyonu .....	31
3.6.1.1.2. Tanjant Hiperbolik Fonksiyonu.....	32
3.6.1.1.3. Tanjant Sigmoid Fonksiyonu .....	33
3.6.1.1.4. LogSig Fonksiyonu .....	34
3.6.1.2. Geri Yayılım Algoritması .....	34
3.6.2. Destek Vektör Makineleri .....	36
3.6.3. K-katmanlı Çapraz-Doğrulama.....	38
3.7. Performans Değerleme .....	39
3.8. Sınıflandırıcıların Değerlendirilmesi.....	40
3.9. Hassaslık ve Duyarlılık Analizi .....	40
3.10. Kullanılan Veri Setleri ve Özellikleri.....	41
<b>4. ARAŞTIRMA BULGULARI VE TARTIŞMA</b> .....	<b>44</b>
<b>5. SONUÇLAR VE ÖNERİLER</b> .....	<b>59</b>
<b>KAYNAKLAR</b> .....	<b>64</b>
<b>ÖZGEÇMİŞ</b> .....	<b>66</b>

## SİMGELER VE KISALTMALAR

### Kısaltmalar

AHÖM	Aşırı Hırslı Öğrenebilen Makine
ATF	Açısal Tabanlı Fonksiyon
ATYB	Ağ Tabanlı Yapay Bağışıklık
BİÇ	Bayesian İstatistiksel Çıkarım
BM	Bulanık Mantık
BSA	Bulanık Sistem Algoritması
BT	Bilgi Teknolojileri
ÇKA	Çok Katmanlı Ağ
ÇKİBSA	Çok Katmanlı İleri Beslemeli Sinir Ağları
DDA	Doğrusal Diskriminant Analizi
DSÖ	Dünya Sağlık Örgütü
DVM	Destek Vektör Makineleri
EKK	En Küçük Kareler
GA	Genetik Algoritma
GB	Geri Besleme
KAH	Koroner Arter Hastalığı
KDM	Karar Destek Makinesi
KDS	Karar Destek Sistemi
MAFIA	Maksimum Tekrarlı Dizi Algoritması
NB	Naive Bayes
ÖAYBS	Öznitelik Ağırlıklı Yapay Bağışıklı Sistemi
SNBM	Sınıf Nitelik Bağımlılık Maksimizasyonu
TBA	Temel Bileşenler Analizi
TFEBT	Tek Foton Emisyonlu Bilgisayarlı Tomografi
TKDS	Tıbbi Karar Destek Sistemi
TKİBA	Tek Katmanlı İleri Beslemeli Ağ
UATBMÇ	Uyarlamalı Ağ Tabanlı Bulanık Mantık Çıkarımı
UBM	Uyarlamalı Bulanık Mantık
VM	Veri Madenciliği
YSA	Yapay Sinir Ağları

## 1. GİRİŞ

Tez çalışmasının konusunu ihtiva eden kalp, kalp hastalıkları, tıbbi teşhis sistemleri, Dünya Sağlık Örgütü verileri, hibrit sistem algoritmaları, hibrit sistemin amacı ve kapsamı, önerilen sistemde kullanılan metotlar, materyaller, sınıflandırıcılar, yapay zekâ kavramı, yazılım geliştiriciliğinin gelmiş olduğu düzey, dünya üzerinde yapay zekânın bazı kullanım alanları bu kısımda özet olarak anlatılmıştır.

Çalışmanın amacı; sağlık konusunda hastalara, uzman doktorlara, acil hekimlerine, sağlık öğrencilerine, araştırmacılara yardımcı olabilecek “Kalp Hastalık Teşhisine Yardımcı” bir sistemi bilim dünyasına sunabilmek olmuştur.

Uzman doktorların olmadığı uzak bölgelerde kalp hastalıkları teşhisine yardımcı olabilecek bir sistem oluşturulmuştur. Bu sistem ön işleme adımı olan Hibrit Sistem (SNBM + TBA + Normalizasyon) ile Sınıflandırıcıları (YSA ve DVM) bir araya getirmiştir. Hibrit sistem ile daha önce yapılmış çalışmalardaki doğruluk oranlarının çoğundan daha iyi bir doğruluk oranı elde edilmiştir. Hibrit Sistemimiz 3 kısımdan oluşturulmuştur. 3 farklı algoritma ile oluşturulan Hibrit Sistem SNBM, TBA ve Normalizasyon adımlarını içermektedir. Hibrit Sistem sayesinde elde edilen sonuçlar Hibrit Sistem kullanmadan bulunan sonuçlardan daha iyi olduğu görülmüştür.

Hastalık teşhisine yardımcı bir sistem ve yapay zekâ arasında bir bağ kurulabilmek için; hastalıklar, hastalığa bağlı ölüm oranları ve makine öğrenmesi teknikleri analiz edilmiştir. Ölüm oranı yüksek hastalıklar içerisinde en üst sırada olan kalp hastalıkları konusuna ağırlık verilmiştir. Bu çalışmada kalp krizi riskinin tespitine, Koroner Arter Hastalığı'nın (KAH) tespitine (var veya yok) şeklinde, kalp hastalığına etki eden faktörlerin (şeker, tansiyon, kapalı damar, yaş ve cinsiyet gibi öznitelikler tba ile incelendiğinde) tespitine çözüm bulabilmek için bir sistem önerilmiştir. Mevcut hastalık tahmini yapan sistemlere yeni metotlar ekleyerek sınıflandırıcıların doğruluk oranlarına katkı sağlamak amaçlanmıştır. Mevcut teknikler ile hesaplanan sonuçlardaki hata miktarını azaltmak bir diğer çalışma nedeni olmuştur. Sağlık alanında bu teknolojilerin kullanım alanlarını artırabilmek ve maliyetleri aşağı çekebilmek bu çalışmanın diğer amaçlarından bazıları olmuştur.

Kalp hastalıklarının bazılarını sıralamak gerekirse:

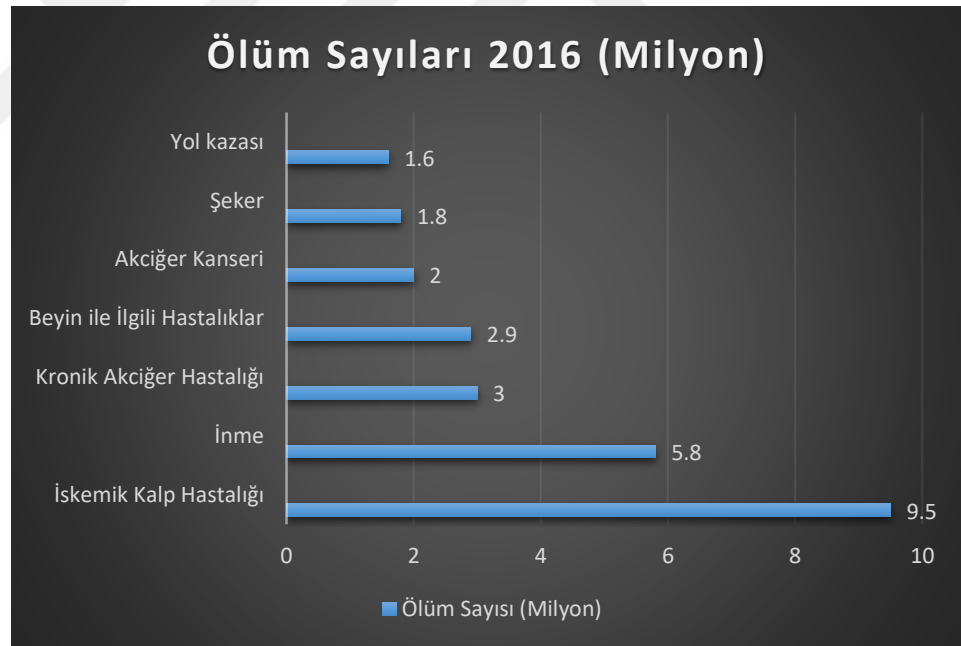
- **Koroner Kalp Hastalığı:** Kalbi besleyen damarların fonksiyonunu yitirmesidir.
- **Romatizmal Kalp Hastalığı:** Kalp romatizması (akut romatizmal ateş) beta hemolitik streptokok adlı bakteri nedeniyle oluşur. Bu bakteri halk arasında kısaca “beta mikrobu” olarak da bilinir.
- **Konjenital Kalp Hastalığı:** Kalpte, aortta doğuştan gelen bir anomalilik ile belirginleşen durumdur.
- **İnme:** Beyne giden kan akımının durması sonucu gelişen klinik bir durumdur. Beyin fonksiyonlarında bozulma, etkilenen bölgeye göre değişiklik gösterir.
- **Hiperlipidemi:** Plazmada bulunan kolesterol ve trigliserid düzeylerinin beklenen normal değerlerden yüksek olması anlamına gelir. Hiperlipidemi özellikle prematür koroner arter hastalığı riskini artırması nedeniyle önemlidir.
- **Kalp Krizi:** Kalp krizi (miyokard enfarktüsü) kalp kasının bir bölümüne yetersiz kan gelmesinden dolayı kasın işlevini yitirmesi sonucu gerçekleşir.
- **Kalp Spazmı:** Anjin olarak bilinir. Kalbe gelen kan miktarının azalması sonucu ortaya çıkan göğüs ağrısıdır.

Elektrokardiyografi, sonografi, anjiyografi, floroskopi, efor testi ve kan testleri sonuçları doktorların hastanın durumu hakkında karar verdikleri teşhis araçlarındandır. Koroner arter hastalığı ilk safhalarında hayatın olağan akışından dolayı kolay fark edilmez. Fakat bu teşhisin doktorun bireysel tecrübesine bağlı olduğu, yanlış karar verebilme ihtimalinin olduğu, bunun tedavi sürecine olumsuz etki edebileceği ve ek masraflara neden olabileceği unutulmamalıdır.

Bilindiği üzere insan sağlığı kalbin düzenli şekilde görevini yapmasına bağlıdır. Diğer organların düzgün çalışması da motor görevinden dolayı kalbe bağlıdır. DSÖ’ye göre 2016’da 57 milyon kişi vefat etmiştir. 41 milyonu bulaşıcı olmayan hastalıklardan dolayı vefat etmiştir. Bu hastalıkların içerisinde 18 milyonu kalp hastalıklarından oluşmaktadır. Kanser ise 9 milyon civarındadır. Şekerden ölüm sayısı ise 1,6 milyon kişidir. 2015’te 17,7 milyon kişi kardiyovasküler hastalıklardan vefat etmiştir (WHO, 2016). Beyin, böbrek ve diğer hayati organlar kalpten gelen materyallerle beslenir. Kalp hastalığı vücuttaki en önemli organların bir anda çökmesine neden olabilecek bir hastalıktır. Pompalanma olayı tam olarak

gerçekleşmezse organlar yeteri derecede beslenemez ve birkaç dakika içerisinde ölüm gerçekleşir. Ailede daha önce kalp hastalığı yaşandı ise, sigara tüketimi, sıkıntı vb. nedenlerden dolayı fazla ise, yüksek kan basıncına sahipseniz, kolesterol değerleriniz normalden fazla ise, spor yapamıyorsanız, nefesinizde daralmalar oluyor ve obezite belirtileriniz varsa kalp hastalığı yaşıyor olabilirsiniz.

Şekil 1.1 de görüldüğü üzere DSÖ verilerinde en sık rastlanan ölüm oranları listelenmektedir. Kalp hastalıkları ve felcin en yüksek değerlerde olduğu görülmektedir. Stres, sigara, ailede kalıtsal aktarım, felç ve benzeri durumlar tansiyon ve kolesterol rahatsızlıkları hastalıktan geç haberdar olma, yanlış teşhis, tedavi imkânı bulamama gibi durumlar maalesef ölüm oranlarını artırmaktadır. Konu ile ilgili genel tarama yapabilmek, ülkesel durumdan haberdar olmak ve bu konuyu optimum masrafla çözebilmek bir teknolojiye ihtiyacı doğurmuştur. Yapay zekâ bu teknolojik ihtiyaca çözüm olmaya aday olmuş ve her geçen gün kendisine duyulan güven ve ilgi artmıştır.



Şekil 1.1 Dünya Geneline Ölüm Oranlarında İlk 7 (WHO, 2016)

Ön işleme adımları ve Veri Madenciliği (VM) on binlerce kayıt yerine daha kompakt, daha mantıklı veri setine dönüşüm sağlayarak ve örüntüler çıkartarak yapay zekâ araçları için bilgi hazırlamaktadır. VM, çalışmada ön işleme adımlarında kullanılmıştır. Daha hızlı işlem yapılabilmesini ve daha mantıklı sonuçlar üretilebilmesini sağlamıştır.

Sağlık alanında kullanılan UCI veri merkezinde tutulan veri setleri üzerinde yapılan araştırma neticesinde kalp hastalıkları ile ilgili olan Statlog (Dua Dheeru, 2017) veri seti ve TFEBT (Krzysztof J. Cios, 2001) veri seti tercih edilmiş ve çalışmadaki veriler bu veri setlerinden elde edilmiştir. Statlog veri seti 270 kayıttan oluşmaktadır. 13 öznitelik barındırmaktadır. TFEBT veri seti 267 kayıttan ve 44 öznitelikten oluşmaktadır. Önerilen sistem Hibrit Sistem ve Sınıflandırıcı adımlarından oluşturulmuştur. Hibrit sistemde ön işlemlere tabi tutulan veri setindeki veriler yeni değerler aldıktan sonra sınıflandırıcı adımına geçmiştir. Bu kısımda hasta olup olmama durumu 2 değerden birisini alır ve bu şekilde veriler ayrılmış olmaktadır. Sınıf Nitelik Bağımlılık Maksimizasyonu (SNBM), Temel Bileşenler Analizi (TBA) ve Normalizasyondan oluşan Hibrit Sistem önerilen sistemin ön işleme kısmını oluşturmuştur. SNBM, TBA ve Normalizasyona çeşitli sırayla giren veri seti ardından sınıflandırıcılara girmiştir. Çalışmada Yapay Sinir Ağları (YSA) ve Destek Vektör Makineleri (DVM) sınıflandırıcıları oluşturmuştur. 2 farklı sınıflandırıcı seçilmesinin nedeni daha iyi sonuç verecek olan sınıflandırıcının bulunmak istenmiş olmasıdır.

SNBM; ChiMerge ve Sınıf Nitelik Kontenjans Katsayısı (SNKK) gibi ayırıklaştırma yöntemlerinden biridir. Mikro dizilerde, genetik kodlamalarda kullanılmaktadır. Kurgan ve Cios (2004) tarafından geliştirilmiştir. Bu algoritma sınıf niteliği ile ayırıklaştırılacak nitelik arasındaki bağımlılıkları ölçmeye yarayan SNBM kriterine dayanmaktadır.

TBA; yüz tanıma, resim sıkıştırma ve örüntü tanıma gibi alanlarda yaygın olarak kullanılan istatistiksel bir metottur. TBA bağımlılık yapısını yok etme ve boyut indirgeme amaçları için kullanılmaktadır. Tanıma, sınıflandırma, boyut indirgenmesi ve yorumlanmasını sağlayan, çok değişkenli bir istatistik yöntemidir. Bu yaklaşım verinin içindeki en güçlü örüntüyü bulmaya çalışır. Bu yüzden örüntü bulma tekniği olarak da kullanılabilir. Çoğunlukla verinin sahip olduğu çeşitlilik, tüm boyut takımından seçilen küçük bir boyut setiyle yakalanabilir. Verideki gürültüler, örüntülerden daha güçsüz olduklarından, boyut küçültme oranında bu gürültüler temizlenebilir.

YSA insan beynini taklit eden bir eğitim sistemine sahiptir. Sonuç ile giriş arasında en iyi tahmini verebilir. Sınıflama, çıkarım, tahmin, büyük boyutlu verilerinden mantıksal sonuç çıkarımı gibi konularda kullanılmaktadır.

Makine öğrenmesi teknikleri hastalık tahmininde de kullanılabilir. DVM son zamanlarda sınıflandırıcı olarak geliştirilmiş başarılı makine öğrenimi algoritmalarındandır. Birçok sınıflandırma uygulamasında kullanılmış ve performansı

oldukça iyidir. DVM sınıflandırma işlemini kareli en iyileme problemine dönüştürür ve bu şekilde çözer. Bu şekilde diğer algoritmalara göre daha hızlı çözüme ulaşılabilmektedir. Teknik bu özelliğinden dolayı, özellikle büyük hacimli veri setlerinde büyük avantaj sağlamaktadır. (Dinçer ve Karaoğlan, 2004)

Yapay zekâ, hastalık teşhisinde önemli bir aşamayı daha kat etmiştir. Yakın gelecekte yapay zekâ hastalık teşhisi için doktorların yerine görev yapabilecek bir konuma gelmek üzeredir. Elde edilen sonuçlar, yapay zekânın kliniklerde görev alabileceğini de göstermektedir. Bilim insanları tarafından geliştirilen yapay zekâ yazılımları, beyin damarlarının rahatsızlığının şiddetine göre gelişen felç ve bunamanın gerçekleşme olasılığını gösterebilir hale gelmiştir.

Yayımlanan makaleler; yapay zekâ teknolojisinin, doktorların acil durumlarda hastalara en iyi tedaviyi daha hızlı bir şekilde uygulamalarına yardımcı olabilecek düzeye eriştiğini gösteriyor. Her geçen gün gelişen yapay zekâ teknolojisi, doktorların en büyük yardımcısı olma yolunda hızla ilerlemektedir. Yapay zekânın, insanların ne zaman hasta olacaklarını tahmin edebileceğini düşünülmektedir. Kimilerine göre bu gelişme, yapay doktorların ayak sesleri anlamına gelmektedir.

Yapay Zekâlı sistemlerin, ilk sonuçları bile bilgisayarların tahminlerinin şaşırtıcı derecede isabetli olduğunu göstermiştir. Doktorlar hasta dosyalarındaki tüm ayrıntılara dikkat etmelerinin çok zor olduğunu dile getirmektedir. Yapay zekânın tıp uzmanlarına bu konuda asistanlık yapabileceği düşünülmektedir.

Tarihte birçok araştırmacı “tahmin etme” konusu ile ilgili farklı metotlar kullanarak araştırmalar yapmıştır. Bir dönem eski programlama dillerinde veriler sistemlere elle girilmekteydi. Girilecek olan verilerin seçimi konusundaki kararlar ise bir kişiye bağlı olmaktadır. Bu tür problemler için artık Yapay zekâ sistemleri kullanılmaya başlanmıştır.

Bu çalışmada Hibrit Sistem büyük boyutlu veri setlerinde mantıklı küçülmenin nasıl gerçekleşebileceğini göstermiştir. VM; Hibrit Sistem içerisinde önerilen sisteme katkı sağlamış sonuçların daha tutarlı ve hızlı olmasını sağlamıştır. Yapay zekâ sisteminin ise içerdiği teknik ve metotlar sayesinde çalışmanın bilim dünyasına farklı açılardan katkı sağlamasına vesile olacağı düşünülmüştür.

Tezin organizasyonu şu şekilde yapılmıştır. Birinci bölümde tez çalışmasına giriş yapılmıştır. Konunun önemi, araştırmanın amacı, kaynak kısmından ayrılan özgün yönleri giriş kısmında anlatılmıştır.

İkinci kısımda kaynak araştırması verilmiştir. Önceki çalışmaların ana fikirleri, metotları ve sonuçları bu kısımda anlatılmıştır. Önerilen sistem ile kıyaslanacak bilgiler bu kısımda aktarılmıştır.

Üçüncü kısımda materyal ve yöntem konularından bahsedilmiştir. Veri serinin nereden elde edildiği, kapsamı, boyutu, nitelikleri bu kısımda anlatılmıştır. Hibrit sistem ve sınıflandırıcılar bu kısımda anlatılmıştır. Araştırmanın düzenleme şekli uygulanan önerilen yöntemler bu kısımda detaylıca ele alınmıştır. Sınıflandırıcıların doğruluk oranlarının test edilmeside bu kısımda detaylıca ele alınmıştır. Çalışmanın detaylı anlatımı kaynak kısmındaki çalışmalardan farklılaştığı noktalar bu bölümde anlatılmıştır.

Dördüncü kısımda ise araştırma bulguları ve tartışma konuları verilmiştir. Elde edilen sonuçların aktarıldığı ve kaynak bölümündeki çalışmalarla kıyaslandığı kısımdır. Diğer çalışmalara atıf yapılmış, karşılaştırılmış ve tartışılmıştır. Hibrit sistemin diğer çalışmalardan farklılığı açıklanmıştır. Yorumlar sayısal verilerle desteklenmiştir. Elde edilen sonuçların geçerliliği ve uygulanabilirliği bu kısımda ele alınmıştır.

Beşinci bölümde ise tezin sonuç kısmı verilmiştir. Sonuç kısmı diğer bölümlerde geçen konuların bir özeti mahiyetindedir. İleride yapılacak çalışmalara, araştırmacılara ve uygulayıcılara önerilerde bulunulmuştur.

## 2. KAYNAK ARAŞTIRMASI

Çalışmada önerilen sistem ile tartışılacak ve karşılaştırılacak olan literatürdeki diğer çalışmalar ana fikirleriyle, metotlarıyla ve sonuçlarıyla birlikte bu kısımda verilmiştir.

Hasan ve arkadaşları UCI'den aldıkları 303 kayıtlı veri setiyle yapmış oldukları DVM sınıflandırıcı uygulamaları ile %81 doğruluk oranına ulaşmışlardır. Çok sınıflı sınıflandırıcıda önerilen sistemde sınıflara normal hastalar için 1 değeri, kalp hastaları için 2 değeri, anjina pektoris için 3 değeri ve kardiovasküler kalp hastası olan hastalar için 4 değeri tanımlanmıştır. Çoklu sınıflı sınıflandırıcı için Matlab uygulaması kodlanmıştır. Sistem doğruluğu için doğrusal kernel fonksiyonu kullanılmıştır. Sistem aynı eğitim seti ile test edilmiştir(Hasan ve ark., 2017).

Bhatia ve arkadaşlarının yaptığı çalışmada 2008 yılında Cleveland UCI'den almış olduğu veri setiyle %70 doğruluk oranına ulaşmışlardır. Bunu sadece 6 öznitelik kullanarak başarmışlardır. Bhatia ve arkadaşları 2008 de sundukları sayısal kodlanmış GA tabanlı kalp hastalığı sınıflandırmasında sadece önemli ve gerekli verileri almış diğer verileri elimine etmişlerdir. Cleveland kalp hastalığı veri tabanında Detrano veri setleri kullanılmıştır. 303 durumlu ve 5 sınıflı olan veri seti her sınıfta 13 öznitelik barındırdığı belirtilmiştir. 250 tanesi öğrenme için geri kalanı test için kullanılmıştır. 5'li sınıflandırıcının doğruluk oranı %72.55 olarak hesaplanmıştır. 13 öznitelikten 6 tanesi kullanılmıştır. 13 öznitelik kullanıldığında sınıflandırıcının doğruluk oranı %61.93 olmuştur. 2'li sınıflamada önerilen teknik %90.57 doğruluk oranına erişmiştir(Bhatia ve ark., 2008).

Gudadhe kalp hastalığı sınıflamada UCI'den aldığı veri setlerinde 200 örneği eğitim için kullanmış geri kalanını da test için kullanmıştır. Verileri 2 sınıfa düşürmüş ve doğruluk oranı %80.41 olarak hesaplanmıştır(Gudadhe ve ark., 2010).

Khemphila ve Boonjing, 2011 de yaptıkları çalışmada 13 özniteliği 8'e düşürmüşler ve doğruluk oranı %80 olarak bulunmuştur. Khemphila ve V. Boonjing 2011'de GB öğrenme algoritması ile UCI'den aldığı veri setinde bir sınıflama sistemi oluşturmuştur. Verinin %60'ı (n=182) eğitim için %40'ı (n= 121) test için kullanılmıştır. 13 öznitelik barındıran sistem 8 özniteliğe düşürülmüştür. Bu işlem öznitelik seçme algoritmasıyla yapılmıştır. 8 öznitelikle doğruluk oranı %80.99 olarak

hesaplanmıştır. 13 öznitelik kullanılarak yapılan çalışmada sınıflandırıcı doğruluk oranı %80.13 olarak hesaplanmıştır(Khemphila ve Boonjing, 2011).

Abushariah 2014'te otomatik olarak hastalık tahmini yapan bir program geliştirmiştir. Sistemin doğruluk oranı %85 olarak bulunmuştur. Abusharian otomatik kalp hastalığı teşhis sistemini YSA ve UBM kullanarak gerçekleştirmiştir. UCI'den alınan veri setleri kullanılmıştır. YSA ile %87.04 doğruluk oranında, UATBMÇ ile %75.93 doğruluk oranında hastalık olup olmadığı tahmin edilebilmiştir(Abushariah ve ark., 2014).

Relief F ve Rough Set tekniklerini kullanan Liu ve arkadaşları tarafından %90 doğruluk oranında kalp hastalıkları ile ilgili bir çalışma yapılmıştır. 2017'de %92.59 doğruluk oranıyla 2 alt sistem içeren bir yapıyla erişmişlerdir. Relief F ve Rough Set (RFRS) özellik çıkarımı ve C4.5 sınıflandırıcı sistemde birleştirilmiştir. UCI'den alınan 303 verinin %70'i eğitim için %30'u test için kullanılmıştır(Liu ve ark., 2017).

Juneja ve Dhingra BM modeli kullanarak hastalık tahmini yapmışlardır. Veri setini belli özelliklere göre gruplandırmışlardır. Kan basıncı, yaş, kolesterol, şeker, kalp atışı ve sigara kullanımı özniteliklerini kullanarak bir sistem geliştirmişlerdir. Sonuçlar bu mantığa göre değerlendirilmiştir. BM tabanlı yapmış oldukları program ile hasta verilerini sistemlerine dahil ederek parametrelere (if clause) göre hastalık tahminlerini değiştirmişlerdir(Juneja ve Dhingra, 2014).

Banu ve arkadaşları bulanık c-ortamalar algoritması kullanarak 13 öznitelikle bir çalışma yapmıştır. Çalışmada %92 doğruluk oranına erişmişlerdir. Yaş, cinsiyet, göğüs ağrısı, kan basıncı, kolesterol, şeker, EKG, kalp atış sayısı, anjin, EKG'deki kalp krizi noktaları, damar değişkenleri kullanılarak bu doğruluk oranı mümkün olduğu kadar yükseltilmeye çalışılmıştır(Banu, 2015).

Shahi ve Gurm yeni gelişen veri madenciliği tekniklerini kullanarak kalp hastalığı tahmini yapabilen bir çalışma önermişlerdir. DVM, KA, c4.5 algoritması, YSA, NB çalışmada kullanılan teknikler olmuştur(Shahi ve Gurm, 2017).

Al-Milli %75 doğruluk oranı veren bir YSA çözümü ile tıbbi karar destek sistemi geliştirmiştir. Amacı kalp hastalıklarındaki riski minimize etmek olan çalışma literatürdeki diğer çalışmalara göre iyi doğruluk oranı vermiştir. Sistemde 13 öznitelik kullanılmıştır. Geri yayımlı ağ kullanılarak oluşturulan yapay zekâ uygulaması ile bu sonuca ulaşılmıştır(Al-Milli, 2013).

Dilip Roy Chowdhury ve arkadaşları tarafından yenidoğan bebeklerde kalp hastalıklarını GYA ile tahmin eden bir çalışma yapılmıştır. Bu çalışmada yeni doğan bebeklerdeki kalp hastalıklarının tahmini için örüntü bulan bir sistem geliştirilmiştir. Çalışmacılar GYA algoritmasıyla sinir ağlarını eğitmek için farklı kategorilerdeki hastalık veri setlerini kullanmışlardır. Önerilen sistemin doğruluk oranı %75 olarak bulunmuştur(Chowdhury ve ark., 2011).

Milan Kumari ve arkadaşları DVM, YSA ve KA verimadenciliği algoritmaları kullanarak kardiyovasküler hastalıkları tespit etmeye çalışmışlardır. Bu çalışmaların performanslarını hassaslık, duyarlılık, doğruluk ve hata oranları faktörlerine göre analiz etmişlerdir. KA çalışmasının doğruluk oranı %81,08 olarak bulunmuştur. YSA çalışmasının doğruluk oranı %79,05 olarak bulunmuştur. DVM çalışmasının doğruluk oranı %80,06 olarak bulunmuştur. KA'nın hata oranı 0,2755, YSA'nın hata oranı 0,2248 ve DVM'nin hata oranı 0,1588 olmuştur. 3 farklı tekniğin analizine göre DVM doğruluk oranı ve hata oranına göre daha iyi performans verdiği görülmüştür(Kumari ve Godara, 2011).

Vanisree K. ve arkadaşları doğuştan gelen kalp rahatsızlıkları üzerine TKDS sistemi geliştirmişlerdir. ÇKİBSA kullanılarak sistemin altyapısı oluşturulmuştur. Değerlendirme kriterleri hastanın o anki belirtileri, tahlil sonuçları ve fiziksel muayene sonucunda oluşturulmuştur. Önerilen sistem %90 doğruluk oranına ulaşmıştır(Vanisree ve Singaraju, 2011).

Niti Guru ve arkadaşları tarafından kalp hastalıkları tahmini için bir YSA modeli önerilmiştir. Sistemde 78 kayıt içeren 13 öznitelik kullanılmıştır. GYA algoritması kullanılmış ve gözetimli ağ tercih edilmiştir. Sistem tarafından verilerdeki öznitelikler kullanılarak hastanın muhtemel hastalıkları tahmin edilebilmiştir(Guru ve ark., 2007).

Sellappan Palaniappan ve arkadaşları VM üzerinde çalışan TKDS prototipi geliştirmişlerdir. KA, NB ve YSA teknikleri kullanılmıştır. DotNet grafik arayüzü kullanılarak sistem geliştirilmiştir. Kullanılan öznitelikler ise yaş, cinsiyet, kan basıncı, kan şeker değerleri olmuştur(Palaniappan ve Awang, 2008).

Shantakumar B.Patil ve arkadaşları kalp hastalıkları tahmini için oluşturulmuş veritabanlarında örüntü arayan bir sistem geliştirmişlerdir. Sistemdeki ilk aşama madencilik işlemleri için veri ambarını ön işleme tabi tutmak olmuştur. K-ortalama kümeleme algoritması kalp hastalıkları tahmininde tercih edilmiştir. MAFLA

algoritması ile veritabanından çekilen veriler madencilik işlemlerine tabi tutulmuştur. İlave olarak hastalık tahmininde önemli rol oynayan örüntüler seçilmiştir. Etkili tahmin yapabilmek adına; ağ bulunan ve seçilen örüntüler kullanılarak eğitilmiştir(Patil ve Kumaraswamy, 2009).

Naive Bayes adlı yapay zekâ tekniği ile yapılan bir TKDS’de ise Patteraki ve Parveen olasılık ile kalp hastalığı tahmini üzerine bir çalışma yapmıştır. Konu ile ilgili olarak veri setinden mantıklı sonuçlar çıkarılmış bu teşhis konusundaki yorumlar olasılıksal olarak değerlendirilmiştir. Web arayüzü ile hastadan verilerini girmeleri istenmiştir. Veri serindeki diğer verilere bakarak bir değer üreten çalışma ile pratisyen doktorlara hastalık tahmini konusunda katkı sağlanmıştır. Ayrıca çalışma ile tedavi masrafları düşürülmüştür(Pattekari ve Parveen, 2012).

Sınıflandırma, kümeleme, regresyon, yapay zekâ, YSA, birliktelik kuralları, KA, GA, en yakın komşuluk gibi teknikleri bir arada değerlendiren Ahmed ve Hannan bir grup çalışmacı ile beraber Cleveland Clinic Foundation, Hungarian Institute of Cardiology, Medical Center ve İsviçre Üniversitesi Hastanesindeki veri setlerini kullanarak toplamda 910 veriyi kullanmış ve yüksek doğruluk oranları elde etmeye çalışmışlardır. Çalışmada KA algoritması ile %92 ve DVM algoritması ile %91 doğruluk oranına ulaşılmıştır(Ahmed ve Hannan, 2012).

VM örüntü çıkarmada, iş stratejileri belirlemede, bilimsel ispat yöntemlerinde kullanılmaktadır. Samiyya ve Sumitra tarafından tıp alanında da kullanılabileceği fikriyle yola çıkılarak bir araştırma yapılmıştır. Çalışmada yaş, cinsiyet, sigara, obezlik, depresyon, hiper tansiyon, yüksek kan kolesterolü, beslenme alışkanlığı, aile geçmişi ve fiziksel hareketsizlik gibi öznitelikler kullanılmıştır. KA, k’ncı en yakın komşuluk ve NB teknikleri kullanılarak sonuçlar karşılaştırılmıştır. Tanagra adlı uygulamada 3000 kayıt incelenmiş ve 14 nitelik hakkında sonuçlar değerlendirilmeye çalışılmıştır. NB %52.33 doğruluk oranını 609 ms’de, KA %52 doğruluk oranını 719 ms’de ve k-ncı en yakın komşuluk %45.67 doğruluk oranıyla 1000 ms’de sonuç üretmiştir. YSA ile doğruluk oranı %80 bulunmuş ve Mantıksal Regression %79 olarak görülmüştür(Sowmiya ve Sumitra, 2017).

BS adlı bir diğer teşhis sistemide Adeli ve Neshat tarafından geliştirilmiştir. Medical Center, Long Beach ve Cleveland Clinic Foundation veri setleri kullanılarak bir girdi seti oluşturulmuştur. Yaş, cinsiyet, göğüs ağrısı, kan basıncı, kolesterol, şeker, EKG, kalp atış sayısı, anjin, EKG’deki kalp krizi noktaları değişkenleri kullanılarak

mandami sonuç çıkarma metodu ile sonuç bulunmaya çalışılmıştır. %94 doğruluk oranı bulunmuştur(Adeli ve Neshat, 2010).

Waghmare tarafından yapılan bir çalışmada kalp hastalığı veri seti kullanılmıştır. Konuyu ekonomik boyutuyla ele almış ve hastalığın masraflarını düşürmek amacıyla tıp alanında bir çalışma yapılmak istenilmiştir. Quantum Neural Network'ün Novel yaklaşımı ile hastanın fiziksel, anatomik ve klinik bulguları bir araya getirilerek sonuç üretilmiştir. NICOR, National Institute for Cardiovascular outcomes Research kurumundan alınan veri seti kullanılarak değerlendirme yapılmıştır. Karmaşıklık matrisine bağlı doğruluk oranı %57 olarak bulunmuştur(Waghmare ve ark.).

Duch ve arkadaşları k-nn, DDA ve GYA algoritmalarını kullanarak bir çalışma gerçekleştirmiştir. Statlog veri setindeki doğruluk oranı GYA algoritması ile %81,3, knn algoritması ile %85,1 ve DDA algoritması ile %84,5 olarak bulunmuştur. Veri seti Medical Center, Long Beach ve Cleveland Clinic Foundation'dan alınmıştır. Çalışmadaki veri seti 303 örnekten oluşturulmuştur. 164 kayıt sağlıklı kayıtlarından geri kalan kayıtlar ise çeşitli hastalıklar barındıran hasta kayıtlarından oluşmuştur. Veri setinde 76 öznitelik olmasına karşın çalışmada 13 tanesi kullanılmıştır. 13 özneliğin 6'sı sürekli ve 4'ü de nominal değişkendir (Duch ve ark., 2001).

Sahan ve arkadaşları ÖAYBS adında bir sistem önermiştir. Bu sistem ATYBS sistemlerinde yer alan özniteliklerdeki farklılıkları gidermeye yönelik olmuştur. Sistemin k-katmanlı çapraz doğruluk oranı %82.59 bulunmuştur(Şahan ve ark., 2005).

Özşen ve Güneş 3 uzaklık kriteri - euclidean, manhattan ve hibrit - ile statlog kalp hastalığının sınıflandırılmasında basit ATYBS'de benzerlik ölçümü yapmıştır. K-katmanlı çapraz doğrulama ile doğruluk oranı %83.95 olarak bulunmuştur(Ozsen ve ark., 2009).

Polat ve Güneş tıbbi veri setlerinin ön işleminde f-puanlama özellik seçimi adında yeni bir özellik çıkarımı metodu kullanmıştır. Çalışma ATF, çekirdek f-puanlama ve EKK-DVM ile maksimum doğruluk oranı %83,7'ye ulaşmıştır(Polat ve Güneş, 2009).

Kahramanlı ve Allahverdi YSA ve BSA ile yeni bir hibrit sistem önermiştir. Önerilen sistemin doğruluk oranı %86,6 olarak bulunmuştur(Kahramanlı ve Allahverdi, 2008).

Subbulakshmi ve arkadaşları TKİBA için AHÖM adında yeni bir öğrenme metodu geliştirmişlerdir(Subbulakshmi ve ark., 2012). Sistemin mevcut sistemlere farklı bir bakış getireceği düşünülmüştür.

Karabulut ve İbrikçi tarafından KAH tahmini için bilişimsel ortalama tabanlı Dönmeli-Sıralı Orman algoritması önerilmiştir. YSA ve Levenberg-Marquardt algoritmasını ana sınıflandırıcı olarak kullanmışlardır.10 kat çapraz doğrulama ile doğrulama oranı %91.20 olarak bulunmuştur(Karabulut ve İbrikçi, 2012).

İngiltere Nottingham Üniversitesi'nden Weng ve arkadaşları, herhangi bir doktor muayenesi olmaksızın kalp krizi veya inme ihtimalini öngörebilen bir makine öğrenmesi algoritması geliştirmişlerdir. American College of Cardiology / American Heart Association (ACC / AHA) iş birliği ile yaş, kolesterol düzeyi ve kan basıncı gibi sekiz farklı faktöre dayanan; hastanın kardiovasküler riskini tahmin edebilmek adına bir dizi yönerge geliştirerek bir yapay zekâ makinesi üretilmiştir. Yapay zekâ kişinin kalp hastalıkları riskini ortalama olarak %72,8 doğruluk oranında doğru tahmin edebilmiştir. Ekip öncelikle dört farklı bilgisayar öğrenme algoritması oluşturmuştur, daha sonra sisteme İngiltere'deki 378.000 hastanın verileri girilmiştir. Sistemin iç tahmini modellerini oluşturabilmek için yaklaşık 295.000 kayıt kullanılmıştır. Kalan hasta kayıtlarını da sistemi test etmek ve gerekli düzenlemeler için kullanmışlardır. Algoritma sonuçları, %74,5 ile %76,4 doğruluk oranı göstererek, belirgin şekilde iyi bir performans göstermiştir. Test seti 83.000 hasta üzerinde kullanılan cihaz, 355'den fazla hastanın yaşamını kurtarmıştır. (Weng ve ark., 2017)

### 3. MATERYAL VE YÖNTEM

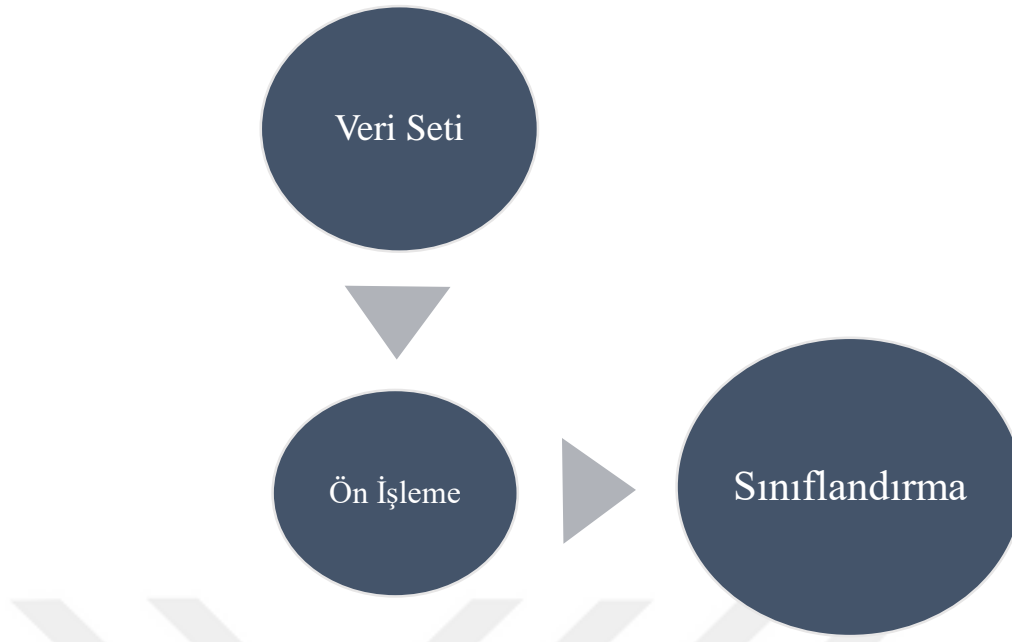
Çalışmadaki materyallerin temin şekli, sayısı, boyutu, büyüklüğü, nitelikleri, temsil ettiği alan, araştırmanın düzenlenme şekli ve uygulanmış yöntemler bu kısımda verilmiştir. Çalışmanın değerlendirilmesindeki metotta yine bu kısımda verilmiştir.

Bu çalışma için en uygun programlama dilinin hangisi olduğu konusunda araştırmalar yapılmıştır. Matematiksel işlemler üzerindeki hâkimiyeti, matrix ve vektörler ile çalışma mantığı, basit kullanım dili, arayüz tasarımında sağladığı kolaylıklar gibi faktörler göz önünde tutulduğunda Matlab programının bu çalışma için altyapı sağlayabileceği düşünülmüştür.

Matlab, matrix ve laboratory sözcüklerin ilk hecelerinin birleşmesinden oluşur. Analitik ve sayısal fonksiyonlar kullanarak, mühendislik uygulamalarına sayısal analiz yöntemlerini dahil edebilen günümüzde üniversitelerde, bilimsel çalışmalarda sık tercih edilen çok kapsamlı ve çeşitli paketlere sahip bir yazılımdır. Algoritma kurma, analiz gerçekleştirme ve mühendislik çözümlerinde grafik tabanlı çözümlerde son derece kullanışlıdır. Cleve Moler tarafından 1971'lerin sonunda matlabın temeli atılmıştır. Matlab programının tipik kullanım alanları: (Uzunoğlu, 2004)

- Algoritma geliştirmek ve kod yazmak
- Matematiksel (nümerik ve sembolik) hesaplama işlemleri
- Lineer cebir, istatistik, fourier analizi, filtreleme, optimizasyon, sayısal
- 2D ve 3D grafiklerin fonksiyona bağlı çizimi
- Modelleme, benzetim, sınıflama, VM gibi teknikler
- Sürekli ve kesikli grafik oluşturmadır.

Geliştirilen veya önerilen sistemde ön işlem süreçleri SNBM, TBA ve Normalizasyon adımlarından oluşmuştur. Sınıflandırıcılar ise YSA ve DVM'den oluşmuştur. Ön işleme süreçlerinin ardından veriler sınıflandırıcılara girmiştir. Sınıflandırıcı sonuçlarına göre hem sınıflandırıcı hemde ön işlemde kullanılan algoritmaların sonuçları ayrı ayrı değerlendirilebilmiştir.



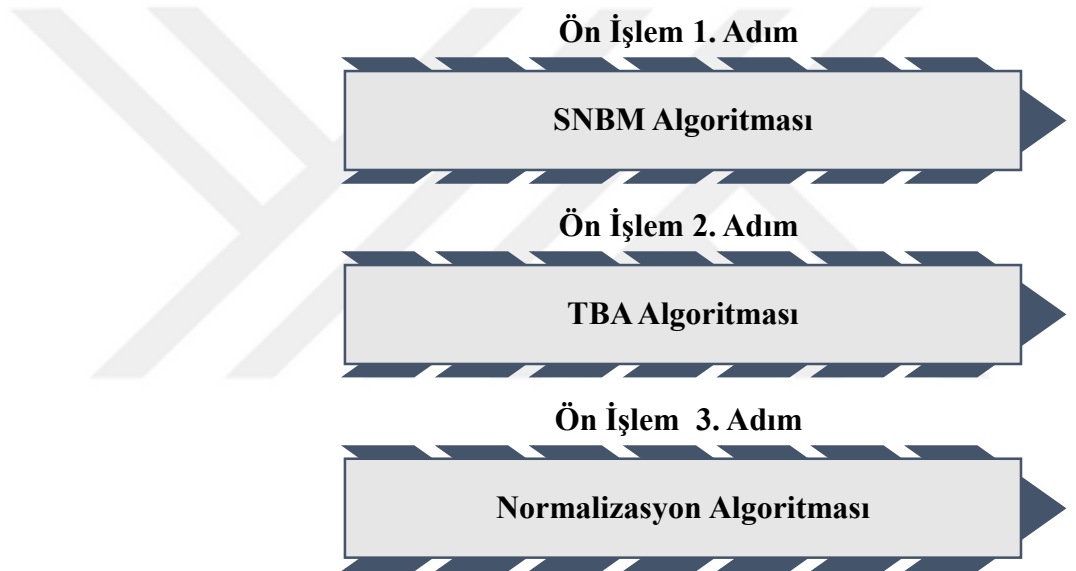
Şekil 3.1 Geliştirilen/Önerilen Sistemin Akış Şeması

Şekil 3.1’te gösterildiği gibi sistem veri seti ilk olarak ön işlemeye tabi tutulmuştur ve ön işlemeden geçen veriler daha sonra sınıflandırıcılara girilmiştir. SNBM ile veri seti öznitelik bazında yeni değerlere sahip olmuştur. Bu sayede veriler SNBM’nin yeni atadığı sayısal değerlere kavuşmuştur ve o değerler üzerinden işlem yapar hale gelmiştir. SNBM bilindiği gibi sıraladığı öznitelikler üzerinde SNBM kriterine göre işlem yapmıştır ve en küçük sınıf aralığı olacak şekilde veriler ayrılmıştır. Onlarca farklı sayı birkaç adımda SNBM’nin atadığı sınıf etiketleri ile isimlendirilmiştir. SNBM, TBA ve Normalizasyon tek başına kullanılamadığı için sınıflandırıcıya ihtiyaç duyulmuştur. YSA veya DVM sınıflandırıcılarına girilen SNBM, TBA ve Normalizasyon adımlarından geçen veriler daha yüksek doğruluk oranları bulunmasına katkı sağlamıştır.

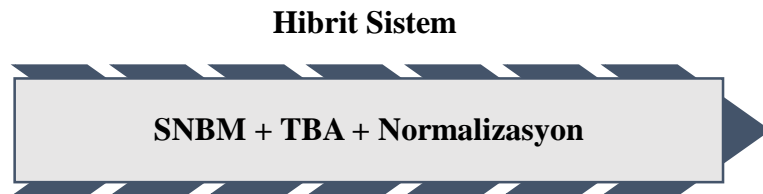
SNBM gibi bir diğer ön işleme adımı da TBA’dır. TBA veri setindeki en güçlü bağların bulunmasına yardımcı olmuştur. 13 girişi olan veri setindeki özniteliklerin birbiriyle ne kadar ilişkili olduğu TBA ile açıklığa kavuşturulmuştur. Her bir özneliğin birbirleriyle olan bağ kuvvetini bir tablo halinde sunan TBA hangi kolonların seçilip hangi kolonların iptal edileceği bilgisini de çalışmada sunmuştur. 76 adet özneliği olan kalp hastalıkları veri seti sadece 8 adet öznitelik ile temsil edilebilir hale gelmiştir.

Normalizasyon ön işleme adımı ise SNBM ve TBA uygulanmış veri setine işlem kolaylığı sağlamak için kullanılmıştır. Bu sayede işlemler normalize edilmiş şekilde sınıflandırıcıda işlenir hale gelmiştir.

Şekil 3.2’de önerilen sistemdeki ön işleme algoritmaları gösterilmiştir. SNBM ve sınıflandırıcılar kullanılarak sonuçlar bulunmuştur. Şekil 3.3’te geliştirilen Hibrit Sistemin şeması verilmiştir. SNBM, TBA ve Normalizasyon ön işlemleri çeşitli şekillerde kullanılarak bulunan sınıflandırıcı sonuçları da Araştırma Bulguları kısmında açıklanmıştır. SNBM + TBA + Normalizasyon veya TBA + SNBM gibi ön işlemler yapılarak farklı sonuçlar elde edilmiştir.

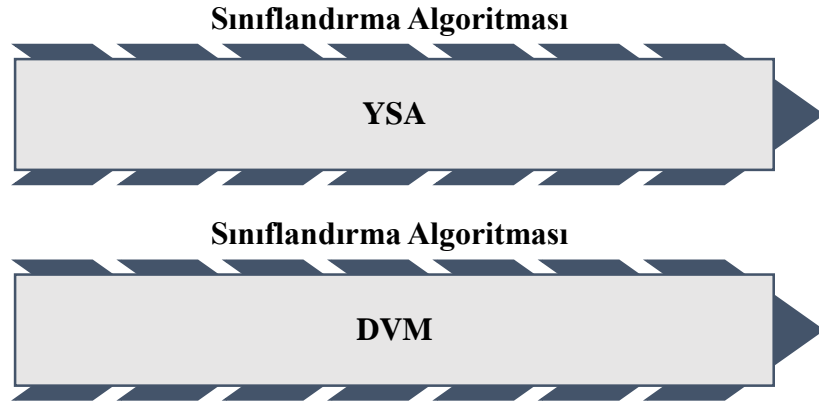


Şekil 3.2 Ön İşlem Algoritmaları



Şekil 3.3 Geliştirilen Hibrit Sistem

Şekil 3.4’te önerilen sistemdeki sınıflandırıcılar gösterilmiştir. Sınıflandırıcılar ön işlemde geçen verileri sınıflandırmıştır.

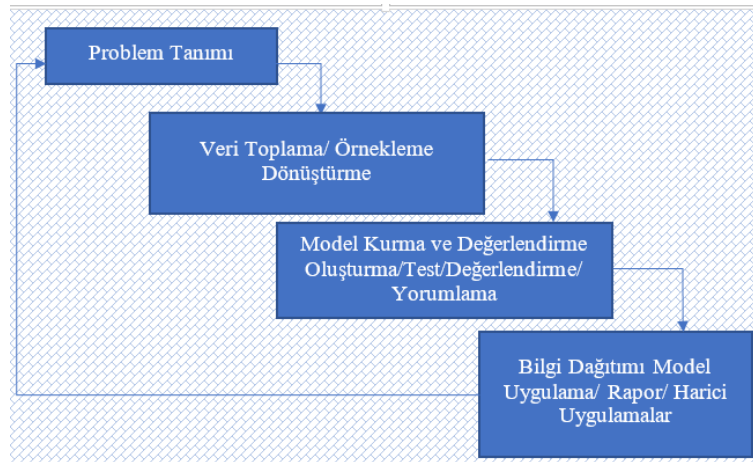


Şekil 3.4 Geliştirilen Hibrit Sistemin Girdiği Sınıflandırıcılar

### 3.1. Veri Madenciliği

Veri Madenciliği (VM) basitçe büyük miktardaki verinin içerisinde bulunan bilginin özünü çıkarma işidir. Ham veriler kullanılarak bilgiye ulaşabilmek için yöntemler kullanılır(Han ve ark., 2011).

VM, veri tabanından bilginin çıkartılmasında çok kritik bir görev üstlenir. VM diğer sektörlerde olduğu gibi; daha iyi hizmet verebilmek, kaynakları daha iyi kullanmak, imkân israfını azaltmak, doğru teşhis sayısını artırmak vb. gibi birçok amaca hizmet eder. Bilgi Keşfi Veri Tabanı, VM'nin en yoğun kullanıldığı yerdir. 4 aşamadan oluşur. Önce gereksiz veriler temizlenir. Farklı veri tabanları ortak olarak birleştirilir. Analiz teknikleri ile veri analiz süreci işletilir. Analizi tamamlanmış veriler yeni bir kütüphaneye aktarılır. VM'deki örüntüler aranmaya başlar. Aranılan örüntülerin içerisinde en başarılı olanlar bir üst basamağa aktarılır. Edinilen bilgiler kullanıcıya sunulur. Şekil 3.1.1'de VM aşamaları gösterilmiştir.



Şekil 3.1.1 VM KDD aşamaları

- **Veri Temizleme**

Uyumsuz olan ve tutarsız görülen veriler temizlenmelidir.

- **Veri Bütünleme**

Büyük veri bankaları bir bütün halinde birleştirilir.

- **Veri Seçme**

Veritabanından analiz için kullanılacak veriler alınır.

- **Veri Dönüştürme**

Özet alınırken veya toplam işlemleri yapılırken bilgi çıkarım işlemleri için uygun formlar, verilerin dönüşümü ve sağlamaştırılması yoluyla hazırlanır.

- **Veri Madenciliği**

Veri Örüntüsü oluşturmak için zekâ metotlarının kullanıldığı hassas süreçtir.

- **Örüntü Değerlendirme**

İlginç ölçümlere dayalı bilgi çıkarımını temsil eder.

- **Bilgi Sunumu**

Kullanıcıya sonuçların aktarıldığı görsel kısım.

İlk dördünde veri önışlemlere tabi tutulur. Örüntüler kullanıcı veya bilgi tabanlı olabilir. VM aşaması en önemli kısmı oluşturmaktadır. Gizli örüntüleri ortaya çıkarır. Bilgi bu kısımda keşfedilir. Büyük veri ambarlarında, bilgi depolarında, veri tabanlarında bulunan ilginç bilgilerin ortaya çıkarılması olarak bilinmektedir.

VM'nin e-ticaret, pazarlama ve perakende sektöründeki başarısı diğer alanlarda da kullanılabilmesinin yolunu açmıştır. Bu sektörlerden bir tanesi de sağlık sektörüdür. Sağlık sektörü verinin çok olduğu ama bilginin o oranda olmadığı sektörlerden bir tanesidir.

Verinin işlenmesi ve özelliklerinin, birbiriyle olan ilişkilerinin çıkarılması gerekmektedir. Hastanelerde oldukça fazla veri olmasına rağmen bu verileri toplum yararına sunabilen araçlar çok fazla değildir. Bu araçların işlemleri kısaltabilmesi için çeşitli şekillerde ilişkileri, kuralları bulabilmesi gerekmektedir.

Kalp hastalıklarının teşhisinin tahmininde KA, BİÇ, KNN, YSA ve Kümeleme tabanlı sınıflama teknikleri diğer tekniklere göre daha iyi sonuçlar verebilmektedir.

İstatistikler KA ve Bayesian Tekniklerinin GA ile en uygun öznitelik kümesi bulunduğundan sonra azaltılan veri setine uygulandıklarında daha iyi performans verdiklerini göstermiştir.

Tıbbi VM gizli kalan veya kestirimi zor olan bağlantıları bulmada çok fayda sağlamaktadır. Bu çıkarımlar, örüntüler teşhis amaçlı kullanılabilir. Ama Tıbbi veriler dağınık yerlerde, karmaşık ve büyük boyutludur. Bu veriler organize şekilde toplanmalıdır ve bir araya getirilmelidir. Bu veriler kamu yararı için organize olarak elde edildikten sonra araştırma amaçlı bilgi sistemlerine dahil edilebilir. VM kullanıcıya yönelik olarak veri içerisinde görülmemiş ve gizli örüntüleri çıkararak bir yaklaşım sunabilmektedir.

Bu araçlardan bir tanesi de tanagradır. Bu araç grafik tabanlıdır. Araştırmacıların VM tekniklerinde işlerini kolaylaştırmak için yazılmıştır. Gerçek veya sentetik verilerin incelenmesinde kullanılmaktadır. KA kullanımı kolay bir sınıflandırıcıdır. Büyük boyutlar için idealdir. Sonuçları okuması ve değerlendirmesi kolaydır.

Öznitelik bazında yapraklara erişmek sadece KA'da mümkün olmaktadır. NB'de öznitelikler arasında bir bağ olmadığı düşünülen istatistiksel bir sınıflandırıcıdır.

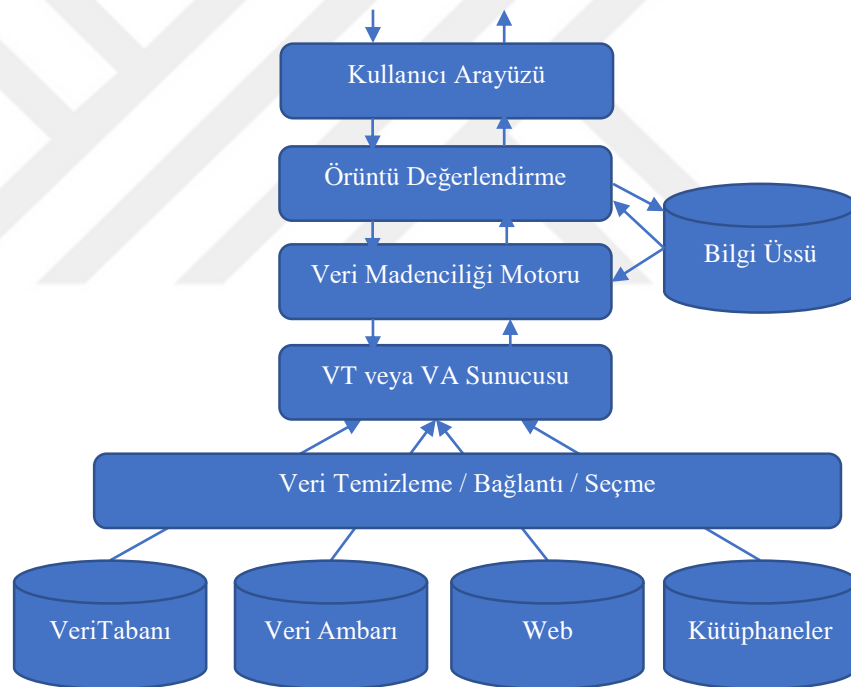
K-nn eğitim setindeki verileri en yakın komşuluk metoduna göre sınıflandırır. Örnek tabanlıdır. Makine öğrenme metodunun en kolaylarından biridir. Gürültü veya verilerde bozulmalar var ise k-nn performans olarak düşer.

GA en uygun veri boyutunu bulmak için kullanılabilir. Sınıflama gözetimli öğrenme metodudur. Kümeleme benzer özellikteki elementleri gruplamada kullanılır. Öznitelik verileri kümelemeden önce yüksek değerli olan verileri düşük değerdekileri etkilememesi için normalize edilir.

Kayıtlar arasındaki gizli bağlantıları bulabilmek için "ilişki kuralları" kullanılır. İlişki kuralları verilere uygulanmak istenirse çok fazla miktarda kural elde edilir. 4 kısıtlama şartına göre durum değerlendirilir. Tıbbi veri setinde tahmini kural bulmak 3 aşamadan oluşmaktadır. Önce gruplanmış ve sayısal veriler işlem veri setine alınır. Yukarıdaki 4 kısıt öznitelik kombinasyonları için çalıştırılır. Bulduğumuz sonuçlar eğitim ve test yaklaşımı ile değerlendirilir.

Klinik kararlar doktorun inisiyatifindedir. Dikkatli ve titizlikle yürütülmelidir. Dolayısıyla otomasyonel çözümlere yönelmek ekonomiklik, doğruluk ve hız açısından önemlidir. Çoğu hastanın birbirinden farklı yerlerde farklı şekilde kayıt altına alındığı düşünülürse bir doktorun binlerce hastanın bilgisini ve özelliğini (örüntüsünü) aklında tutması ve bunu onlarca diğer doktora aktarmasının mümkün olmadığı görülmektedir.

Hastalara sunulan hizmetin artırılması, yanlış tedavilerin önüne geçilebilmesi için otomasyon sistemleri sürekli güncellenmeli ve yeni algoritmalar ile desteklenmelidir. Gelişigüzel yapılan tahminler, doktor hataları ve gereksiz EKG, tahlil ve test gibi masraflarda bu sayede azaltılabilecek, cihazların kullanım ömründe artmış olacaktır. VM bilgi-zengini bir çevresel ortam ile klinik karar almada önemli bir yer tutmakta ve gelecek için umut vermektedir. Şekil 3.1.2’de VM’nin iş akış şeması verilmiştir.



Şekil 3.1.2 VM İşleyiş (Han ve ark., 2011)

**Bilgi Üssü:** Araştırmaya rehberlik etmesi için veya ilginç örüntüler keşfetmek için kullanılan merkez bilgi kütüphanesidir. Bu bilgi kütüphanesi kavram hiyerarşiler içerebilir, öznitelik ve öznitelik değerleri farklı soyutlamalar için kullanılabilir. Bazı kullanıcının bildiği ilginç örüntüler ilave edilebilir. Diğer kütüphane verileri kısıtlar, ölçütler ve meta veridir.

**VM Motoru:** Bu VM sisteminin temelidir. Bazı fonksiyonel görevleri barındırır. Bu fonksiyonlar şunlardır: karakterizasyon, ilişki, ilgileşim analizi, sınıflama, tahmin, kümeleme analizi, kural dışı analizi ve evrim analizidir.

**Örüntü Değerlendirme Modülü:** İlginç örüntüler bulmak için diğer modüllerle iletişime geçen bazı ilginç kısıtları kullanan bir komponenttir. İşe yaramayan örüntüleri bulup sistem dışına çıkarabilir. VM modülü ile beraber bazı görevleri yerine getirmede kullanılabilir. VM kısmına mümkün olduğu kadar örüntü çıkarma kısmını nüfuz ettirmek gerekir nedeni ise ilginç örüntüler bulabilmektir.

**Kullanıcı Arayüzü:** Kullanıcılar ve VM Sistemi arasında iletişim kurar. Bu sayede kullanıcı bir sorgu ile madencilik yapabilir. Kullanıcı sorgu ile aramayı kolaylaştırabilir ve ara havuzdaki veriler içerisinde de sorgu yapabilir. Ayrıca bu modül ile kullanıcı veri yapılarını örüntüleri arama şemalarını görebilir.

### 3.2. Sınıf Nitelik Bağımlılığı Maksimizasyonu

Sınıf-nitelik bağımlılığına bağlı olarak sonuç üreten bir diğer algoritma Lukasz Kurgan ve Krzysztof Cios tarafından geliştirilen SNBM algoritmasıdır. Bu algoritma ayırıklaştırma sonrası elde edilecek ayırık aralık sayısını otomatik olarak kendisi belirleyerek kullanıcıdan herhangi bir değer girmesini beklemez. Dolayısıyla kullanıcı etkileşimi azaltılmış olur.

SNBM algoritmasının amaçları şu şekilde sıralanır:

- a. Sınıf etiketi ve sürekli değerli nitelik arasındaki bağımlılığı en üst düzeye çıkarmak,
- b. Mümkün olan en küçük aralık sayısını elde etmek,
- c. Makul bir maliyet ile ayırıklaştırma işlemini gerçekleştirmek.

SNBM algoritması sınıf-nitelik bağımlılığını hesaplayabilmek için Quanta Matrix kullanılır. Buna göre SNBM kriteri oluşturulur. Çizelge 3.2.1'de Quanta Matrix görülmektedir. Sınıflar, aralıklar ve aralığın en büyük değerleri de tabloda yer almaktadır.

Class	Interval			Class Total
	[d <sub>0</sub> , d <sub>1</sub> ]	[d <sub>r-1</sub> , d <sub>r</sub> ]	[d <sub>n-1</sub> , d <sub>n</sub> ]	
C <sub>1</sub>	q <sub>11</sub>	q <sub>1r</sub>	q <sub>1n</sub>	M <sub>1+</sub>
C <sub>i</sub>	q <sub>i1</sub>	q <sub>ir</sub>	q <sub>in</sub>	M <sub>i+</sub>
C <sub>s</sub>	q <sub>s1</sub>	q <sub>sr</sub>	q <sub>sn</sub>	M <sub>s+</sub>
Interval Total	M <sub>+1</sub>	M <sub>+r</sub>	M <sub>+n</sub>	M

Çizelge 3.2.1 Quanta Matrix

SNBM kriterinin formülü denklem 3.1’de gösterilmektedir:

$$CAIM(C, D \setminus F) = \frac{\sum_{r=1}^n \frac{max_r^2}{M_{+r}}}{n} \quad (3.1)$$

R. sütunda bulunan, en büyük n değeri, sınıf sayısı M<sub>+r</sub>, (d<sub>r-1</sub>, d<sub>r</sub>] aralığında yer alan niteliğin sürekli değerlerinin toplam sayısını max<sub>r</sub> ile temsil eder.

Örnek sayısı M, sınıflar S, nitelik değerleri A<sub>i</sub> olmak üzere SNBM algoritması aşağıdaki gibidir:

1. Nitelik değerleri - A<sub>i</sub> - içerisindeki en büyük ve en küçük değerler hesaplanır. Bu hesaplardan elde edilen değerler Çizelge 3.2.1’de bahsi geçen aralıkların en küçük değeri e<sub>n</sub> ve en büyük değeri e<sub>0</sub> olacaktır.

2. Nitelik değerleri küçükten büyüğe doğru sıralanır ve mümkün olan tüm aralıklar hesaplanır.

3. Hesaplanan aralık değerleri ile bir aralıklar kümesi oluşturulur başlangıç SNBM değeri sıfıra eşitlenir.

4. Parametreye başlangıç değeri olarak k=1 verilir.

5. Hesaplanan aralıklar içerisine yeni geçici sınırlar eklenir.

6. Geçici eklemeler yapıldıktan sonra hesaplanan her aralık sonrası Çizelge 3.2.1 yardımı ile SNBM değerleri tekrar hesaplanır.

7. Hesaplanmış değerler arasından en yüksek olanı SNBM değeri seçilir.

8. Sıfır olarak atanan SNBM değeri yeni hesaplanan değer ile değiştirilir. Daha sonra bulunan değer en son atanan SNBM değeri ile karşılaştırılarak işlem süreçleri devam eder. Eğer bu adımda hesaplanan en yüksek değer bir önceki genel değerden küçük ise işlem sonuçlandırılır. Tersisi durumda ilerleyen adıma geçilir.

9. Maksimum değerin bulunduğu aralık bulunan aralıklar dizisinin içine eklenir. K değeri bir yükseltilecek e. adıma geçilir ve ilerleyen işlemlere devam ettirilir.

Tüm SNBM değerleri hesaplamalarının ardından ayrık aralıklar elde edilmiş olur. (Koçoğlu, 2012)

Bir örnek vermek gerekirse tümörü oluşturan genler ve SNBM ile ayrıklaştırılmış halleri SNBM kriteri ve tablo değerlerine göre hesaplandığında Çizelge 3.2.2'deki gibi gösterilebilir. En küçük olacak şekilde 1 ve 2 etiketleri kullanılmıştır. Gen1 21 değerine kadar 1 ve 21'den itibaren yüksek değerler içinde 2 olarak etiketlenmiştir. Gen 2 42 değerine kadar 1 ve 42'ten itibaren yüksek değerler için 2 olarak etiketlenmiştir. Gen 3 21 değerine kadar 1 ve 21'den itibaren yüksek değerler için 2 olarak etiketlenmiştir.

Çizelge 3.2.2 SNBM Örnek Tablosu (Yalçın Özkan, 2017)

TÜMÖR	GEN1	GEN2	GEN3	GEN1	GEN2	GEN3
A	19	35	11	1	1	1
B	26	25	26	2	1	2
B	21	42	21	2	2	2
B	44	56	28	2	2	2
A	16	32	16	1	1	1
B	17	34	17	1	1	1
B	25	50	25	2	2	2
A	10	20	10	1	1	1
A	10	20	10	1	1	1
A	17	34	17	1	1	1
B	27	54	27	2	2	2
A	60	11	12	2	1	1
B	25	43	24	2	2	2

### 3.3. Temel Bileşenler Analizi

TBA veri analizinde kullanılan büyük boyutlu verileri fazla bozulmayacak şekilde mantıklı halde küçültmeye yarayan bir yöntemdir. İlgisiz olan öznelikler bu sayede elimine edilmiş olur. Temel bileşenler verilerin içerisinde seçilir ve ilk seçim en büyük değeri taşır. Diğerleri azalan sırada sıralanır.  $X_1, X_2$  vektörleri;  $Y_1, Y_2$  vektörü olacak şekilde TBA'nın ayarladığı bir açı kadar döner. Bu sayede veriler kendi içerisinde mantıklı şekilde boyutlanmış olur. Yeni boyutlar vektörel olarak görselleştirilmiş olur. TBA birbirleriyle kesişmeyen doğrulardan oluşur. Birbirlerine 90 derecelik açı ile bağlanırlar. Birinci temel bileşen en iyi sonucu veren temel bileşendir ve ilk doğruyu oluşturur. Bir sonraki bileşen en iyi sonucu veren diğer en

önemli deęişkendir ve önceki doğruya dik çizilir. Dik doğrular en iyi bileşenleri temsil edecek şekilde sırasıyla çizilir. Bu şekilde hacim küçültülmüş ve büyük boyutlu veri görselleştirilmiş olur. TBA, veri sıkıştırma için de kullanılabilir. TBA ile hacim küçültülürken veri kaybı yaşanmamaktadır. Bu sistemin güvenilirliğini artırmaktadır.

Öznitelik azaltma yöntemi olan TBA şu şekilde formülize edilir:

M t-boyutlu veri seti olsun. n temel yön  $G_1, G_2, \dots, G_n$  olsun. Ortak olarak  $G_1, G_2, \dots, G_n$  örnek kovaryans matrisinin n tane özyöne y vektörleri olarak adlandırılır: Denklem 3.2:

$$C = \left(\frac{1}{L}\right) \sum_{k=1}^L (x_k - \bar{x})^T (x_k - \bar{x}) \quad (3.2)$$

$$x_k \in M$$

$\bar{x}$  Burada örneklerin ortalamasıdır. L örnek sayısıdır.

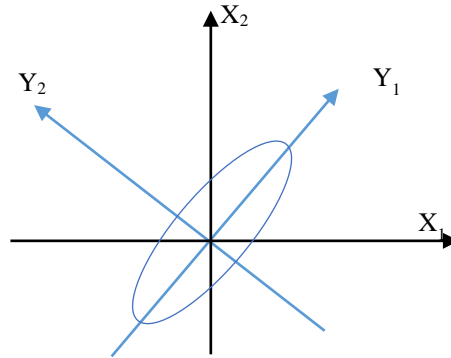
Buna baęlı olarak

$$UG_k = v_k G_k, \quad k \in 1, \dots, n$$

Burada  $v_k$  k'ncı U'nun en büyük özyöne y deęeridir. Verilen gözlem vektörünün n temel komponenti şu şekilde anlatılır.

$$Q = [q_1, q_2, q_3, \dots, q_n] = \left[ \left[ G_1^T x, G_2^T x, \dots, G_n^T x \right] = G^T x \right]$$

q x[] in en temel komponentidir. Şekil 3.3.1'de açıklandığı gibi en önemli bileşenler  $Y_1$  ve  $Y_2$  ile gösterilmiştir.



Şekil 3.3.1 TBA  $Y_1$  ve  $Y_2$  İlk İki En Önemli Bileşenler

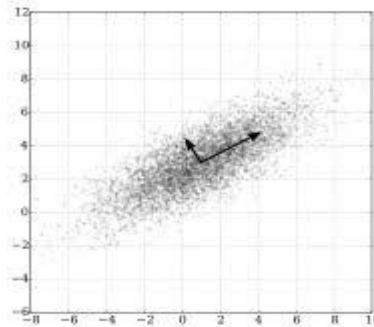
TBA (Karhunen-Loeve or K-L metodu)  $k < n$  olacak şekilde n-boyutlu ortogonal vektörleri temsil edecek k deęerlerini arar. Bu sayede büyük boyutlu veri seti küçültülmüş olur. Öznitelik sayısını azaltmak yerine onları birleştirerek sayılarını azaltır. Öznitelikler temsil edilmiş olur. TBA fark edilmeyen baęıntılarını ortaya çıkarır.

Temel prosedürü şu şekilde açıklanabilir:

- Girdi verisi normalize edilir. Her öznitelik sayıca düşmüş olur. Bu sayede büyük hacme sahip öznitelikler diğerlerini etkilemeyecektir. (Han ve ark., 2011)
- TBA normalize edilmiş girdi seti için k ortonormal vektör oluşturur. Bunlar birim vektördür ve biri diğerine diktir. Bu vektörler temel bileşenlerdir. Girdi verisi temel bileşenlerin doğrusal bir kombinasyonudur.
- Temel bileşenler kuvvet ve önemine göre azalan sırada sıralanır. Temel bileşenler yeni eksenleri oluşturur ve değişkenlerin önemini gösterir. İlk eksen en önemli veriyi, ikincisi ikinci en önemli veriyi gösterir. Böylece devam eder. Şekil 3.3.1 de  $X_1$  ve  $X_2$  merkezli yerleşmiş verilerde ilk temel bileşenler  $Y_1$  ve  $Y_2$  gösterilmiştir. Bu bilgi grup ve örüntüleri anlamada yardımcı olur.(Han ve ark., 2011)
- Sıralamada yüksek olan yukarıda olduğu için az değerde olanlar silinebilir. Orijinal veri küçültülmüş olur.

TBA hesaplaması kolaydır ve sıralanmış ve sıralanmamış verilere uygulanabilir. 2'den fazla boyut 2'ye düşürülebilir. TBA regresyon ve kümeleme analizlerinde girdi olarak kullanılabilir. Wavelet ile kıyaslandığında TBA seyrek veriler ile uğraşmada daha iyidir. Wavelet büyük boyutlarda daha iyidir.

Şekil 3.3.2'de görüldüğü gibi noktasal gösterilen veriler kümelenmiş ve 2 yön olarak en önemli bileşenler çizilmiştir.



Şekil 3.3.2 TBA Uygulanmış Veri Seti(Wikizero)

TBA'nde amaç verilerde bilgi kayıplarını azaltmak ve optimum sonuç elde etmektir. Dolayısıyla kaç adet bileşenin tercih edileceği önem arz etmektedir. Optimum sonuç için kullanılan yöntemler maddeler halinde aşağıdaki gibi açıklanabilir:

- **Durdurma kuralı:** Analizde 1 ve üzerinde değere sahip olan faktörlerdir.
- **Eğim testi:** Bu test, değer ve değişken sayısı arasındaki bağı çizimsel olarak ifade eder.
- **Önemli faktör sayısı:** Durdurma değeri 0.30'dan büyük iki veya üç değişken içeren bileşenler önemli olan değişkenlerdir.
- **Toplam varyans oranı:** 8 değişkenle ifade edilen bir veri kümesinde 8 değişkenin içerdiği kümelerde kullanılırsa değişimin tamamı görselleştirilmiş olur. Araştırmalar, toplam varyansın en az %70-80 arasında olması gerektiğini belirtmektedir.

### 3.4. Veri Normalizasyonu

Veri normalizasyonu tüm değişkenlerin belirli bir aralığa dönüştürülmesi demektir. Sinir ağlarında normalizasyon öznelikler farklı aralıklarda ise yapılır. Bunu bir niteliğin diğerine etkisini azaltmak için yapılır. İkinci olarak normalize edilmiş setler daha hızlı sonuç verir. Eğer veri setindeki veriler farklı birimlerde ise önce normalizasyona tabi tutulması gerekir.  $m^2$ , kg,  $cm^3$  gibi. Model kurmadan önce bazı standart işlemler yapılmalıdır. Filtreleme vb. işlemler ile aykırı olan veriler çıkarılmalıdır. Özneliklerin birbiriyle orantılı olması için normalize edilmesi veya standart hale getirilmesi gerekir. Değişkenler arasındaki eşitsizlik katsayılara göre düzenlenir. Normalizasyonun birim farklılığını ortadan kaldırmada kullanıldığı unutulmamalıdır.

**Asgari - Azami Normalleştirilmesinde** bir dizi içerisindeki sayısal değerler öncelikle küçükten büyüğe doğru sıralanır. En küçük değer en üstte diğerleri onu takip edecek şekilde küçükten büyüğe doğru sıralama tamamlanır. En küçük sayı 0 en büyük sayı ise 1 olacak şekilde yeni değerlerine çevrilme işlemi uygulanır. İstenilen her hangi bir dizi değeri en küçük sayıdan çıkarılır. Ardından en büyük sayıdan en küçük sayının çıkarılması ile bulunan sonuca bölünerek yeni değeri elde edilir. Bu bulunan değer 0 ile 1 arasında olup min-max (en küçük - en büyük) Normalizasyonu diğer dizi değerleride hesaplanarak tamamlanmış olur. Sıfır ve bir arasında değerler dağılmış olmaktadır. Dizi değerleri arasında benzerliğin olmadığı durumlarda verileri tek bir düzen içerisine sıralı koymak için bu yöntem tercih edilir. Standart olarak farklı terimlerle ifade edilen veriler tek bir çatı altında toplanabilmektedir. Bu değerler istenirse kıyaslama veya oranlama amaçlı da kullanılabilir.

**Standart Skor** diğerk bir yöntem olarak normalleştirme işlemlerinde tercih edilegelmiştir. Standart skorlama yönteminde ise istatistikte sık kullanılan ve verilerin birbirlerine göre durumunu ifade eden standart sapma ve ortalama değerk hesaplamalarından faydalanılır. Standart sapmaya bağılı matematiksel işlemler ile bu hesaplama yapılır. Dönüşüm sağlandıktan sonra diğerk hesaplama yöntemleri ile kıyaslama yapılabilir. Toplamların dizi uzunluğuna oranı olan değerk dizi içerisindeki sayıların karelerinin alınıp toplandıktan sonra dizi uzunluğuna bölünmesi ile elde edilen değerk kare köküne bölünmesiyle normalizasyon yapılmış olur.

### 3.5. Sınıflandırıcılar

VM’de verinin sınıflandırılması önemli bir yer tutar. Verinin içerdiği ortak özellikleri veya farklılıkları kullanarak sınıflandırma işlemi yapılır. Sınıflandırma işlemi esasen bir öğrenme algoritmasına dayanır. Bu amaçla çok sayıda algoritma geliştirilmiştir.

Sınıflandırma işlemi; veriyi kullanarak sınıflandırıcıların belirlenmesi, bu sınıflandırıcıların yeni verilere eşleştirilmesi ve bir tahminde bulunulmasını sağlar. Bu açıdan bakıldığında sınıflandırmayı hangi sınıfa ait olduğu bilinmeyen bir gözlemin sınıfının belirlenme süreci olarak tanımlamak mümkündür.

Bu amaçla kullanılan yapay zekâ teknikleri kendilerine sunulan bilgileri kategorize etmek görevini üstlenirler. Bir makine üzerinde görülen hataların sınıflandırılması buna örnek olarak verilebilir.

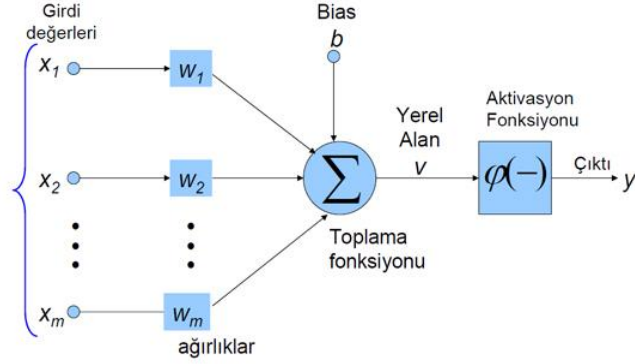
Değişik görevler için değişik sınıflandırıcı türleri vardır. Bunlar bazı projelerde hibrit olarak kullanılabilir. İstatiksel, tahmine dayalı ve mesafe uzaklık ölçümüne dayalı türleri vardır. En çok kullanılanları özetlemek gerekirse:

### 3.6. Sınıflandırıcı Türleri

#### 3.6.1. Yapay Sinir Ağları

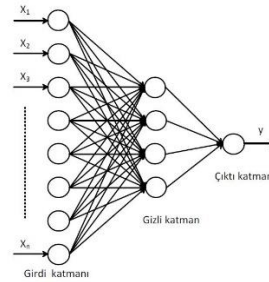
YSA insan biyolojisinden esinlenilerek tasarlanmış makine öğrenmesi tekniklerindedir. Katmanlar olarak ayrılmış ve birbirlerine bağlanmış belli sayıdaki nöronlardan oluşurlar. İleri Beslemeli Sinir Ağlarının en önemli türü ÇKA’dır. ÇKA’ler en kolay ve en sık kullanılan sinir ağları mimarisidir. Genel olarak ÇKA 3 katman içerir. Bunlar girdi katmanı, ara katman ve çıktı katmanıdır. Beyindeki sinir hücrelerine benzetilen yapay nöronlarda girdi katmanı dentritlere, ağırlıklar sinapslara, hücre gövdesi toplama fonksiyonuna ve aksonlarda çıktı fonksiyonlarına

eşleşmektedir. Toplama fonksiyonundaki bias değeri yapay sinir ağlarına sonradan eklenmiştir. Hücre gövdesinde bu değer yoktur. Dentritlerden gelen değerler sinapslarda toplanır ve hücre gövdesine iletilir. Burada yapılan işlemlerin ardından eşik değeri geçen bir sinyal üretilirse aksonlardan ilgili dentritlere tepki sinyalleri gönderilir.



Şekil 3.6.1.1 YSA Çalışma Prensibi (Adıyaman, 2007)

Şekil 3.6.1.1’de YSA’nın çalışma prensibi gösterilmiştir.  $X$  girişlerindeki değerler  $w$  ağırlıkları ile çarpılarak toplama fonksiyonuna dahil edilir bias değeri de toplama dahil edilerek aktivasyon fonksiyonuna yönlendirilir. Toplam sonucu aktive olmuş ise çıktı üretilir.



Şekil 3.6.1.2 YSA Girdi, Gizli ve Çıktı Katmanları

Şekil 3.6.1.2’de ise katmanlar arası bağlantılar görülmektedir. Girdi katmanından çıktı katmanına kadar uzanan bu ilişki katmandaki nöron sayısına göre değişiklik gösterir. Her girdi nöronu her gizli katmandaki nöronla bağlantı kurmaktadır. Gizli katman ise çıktı katmanı ile bağ kurar.

ÇKA da ara katman ve çıktı değerleri aşağıdaki denklem 3.9 ile formülle hesaplanır:

$$y_j = f\left(\sum w_{ji}x_i\right) = f(I_j) \quad (3.9)$$

f eşik değeri olarak bilinir sigmoid, hiperbolik ve açılı tabanlı fonksiyon gibi türleri vardır. GYA’da türevsel fonksiyon kullanılmalıdır.

ÇKA en çok kullanılan algoritma GYA’dır. Bir girdinin ağa uyguladığı hatayı minimize etmek için kullanılır. Geri Yayılımda önce ileri yönde bir ilerleme kaydedilir ve hata oranında geri yayılım yapılır.

Bu adımlar tolere edilebilecek seviyeye kadar devam ettirilir. GYA’da nöronlar arası ağırlık bağlantıları denklem 3.10’daki gibi gösterilir. Girdilerdeki değerler ağırlıklar ile çarpılarak tüm girdiye göre ara katman değeri hesaplanır. Bu tüm gizli katman nöronları için tekrarlanır. Gizli katmandaki değerler toplamı ile çıktı katmanındaki değer arasındaki hata farkı gizli katmandaki nöronlara yansıtılır.

$$net_j = \sum w_{ji} x_i \quad (3.10)$$

YSA çıktısı için her nöronun hatası denklem 3.11 deki gibi ölçülür:

$$E_j = y_{dj} - y_j \quad (3.11)$$

Olması gereken değer ve bulunan değer çıkartılır.

E tüm sistemdeki hata ise denklem 3.12 deki gibi formülle bulunur.

$$E = \frac{1}{2} \sum_j (y_{dj} - y_j)^2 \quad (3.12)$$

Her ağırlık hata miktarını azaltmak için ayarlanır.

Şekil tanıma gibi problemlerde hata dağıtımını gerektiğinden bu yöntemler kullanılır.

YSA’nın uygulamaları incelendiğinde YSA’nın genel olarak şu fonksiyonları gerçekleştirmek için uygulandıkları görülmektedir.

**Tahmin:** Bu amaçla kullanılan YSA ağa sunulan bilgilerden yararlanarak karşılık gelen çıktı değerini tahmin eder. Hava tahmini, borsada hisselerin değeri tahmini, döviz kurlarının tahmini gibi örnekler vermek mümkündür.

**Veri İlişkilendirme:** Bu amaçla kullanılan YSA kendilerine sunulan bilgilerin hatalı veya eksik olup olmadığını belirler. Öğrendikleri bilgiler ile eksik olan bilgileri tamamlarlar. Eksik bir resmin tamamlanması bu konuda örnek olarak verilebilir.

**Veri filtreleme:** Bu amaçla eğitilen ağlar birçok veri arasından uygun verileri belirleme görevini yerine getirirler. Telefon konuşmalarındaki gürültüleri asıl konuşmalardan ayıran ağlar bu konudaki uygulamalara örnek olarak verilebilir.

**Tanıma ve eşleştirme:** Değişik şekil ve örüntülerin tanınması, eksik karmaşık belirsiz bilgilerin işlenerek eşleştirilmesi ve tanıma fonksiyonları gerçekleştirilebilir. Kalite kontrol şemaları üzerinde şekilleri tanıyan ağ bu konuda örnek olarak verilebilir.

**Teşhis:** Bu amaçla geliştirilen ağlar sistemlerin olumsuzluklarının ortaya konulması ve problemlerin teşhis edilebilmesi işlemini yerine getirirler. Makinelerin süreçlerinin, arazi durumlarının ve hatalarının teşhis edilebilmesi buna örnek olarak verilebilir. Tıp alanında da bu tür sistemler yaygın olarak geliştirilmektedir.

**Yorumlama:** Bir olay hakkında toplanan örneklerden elde edilen ve eğitim oranı oluşturulan bilgileri kullanarak yeni olayların yorumlanması işlemleri bu kapsamda düşünülmektedir. Bir olay hakkında toplanan verilerin yorumlanarak istatistiksel dağılımlarının belirlenmesi bu konuda örnek olarak verilebilir.

**Endüstriyel:** Ürün tasarım, süreç, gaz kontrolü, otomatik rehber, kimyasal süreçler, robot görmesi, piston vb cihazların dayanıklılığı, müşteri ve portföy işlemleri, mikroçip yongası analizi, üretim planlama ve kontrolü, hatalı ürün tespiti, optimizasyon, çizelgeleme, gezgin satıcı bazı örneklerdendir.

**Finansal:** Makroekonomik tahminler, borsa benzetim çalışmaları, sahtekârlık tespiti, iflas tahminleri, kredi analizi, emlak kredilerinin yönetimi, döviz kuru tahminleri, risk analizleri bazı örnekleridir.

**Askeri:** Hedef tanıma ve takip, yeni sensörlerin analizi, radar ve görüntü sinyali işleme, sensör fizyonu, askeri uçakların uçuş yörüngelerinin belirlenmesi, optimizasyon, mayın detektörleri bu konuda örnek verilebilir.

**Sağlık:** Solunum hastalıkları teşhisi, eeg ve ekg analizleri, trasplant zamanlarının optimizasyonu, hastalıkların resimden tanınması, kardiyovasküler hastalıkların modellenmesi, tıbbi resim işleme, ctg izleme, hamile kadınların bebeklerin kalp ritmlerinin ölçülmesi, yumurtalık kanserinin önlenmesi, üroloji uygulamaları bu konunun örneklerinden bazılarını oluşturmaktadır.

**Diğer Uygulamalar:** Sigorta poliçelerinin değerlendirilmesi, uçak parçalarının hata teşhisleri, petrol ve gaz araması, bomba dedektörleri, rotalama sistemleri, resim işleme, segmentsayon yenileme, karakter, el yazısı ve imza tanıma işlemleri, şekil

sıkıştırma, vm, insani davranış sergileyen çocuk oyuncakları, genetik, ilaç endüstrisi, nanoteknoloji, robotik, baraj değişkenleri hesaplama işlemleri ve kestirimi uygulamaları örnek olarak verilebilir.

YSA matematik olarak modellenmesi mümkün olmayan veya zor olan karmaşık problemleri çok rahat modelleyerek çözebilmektedir. YSA kullanarak problemleri başarılı bir şekilde çözebilmek için problemin çok iyi modellenmesi gerekmektedir. Bu modelleme, problemi çözebilmek için sadece söz konusu olay ile ilgili örneklerin belirlenip toparlanmasına yardımcı olacaktır.

Gerçek dünyada olaylar ve olayların arkasındaki değişik faktörlerin birbirleri ile ilişkilerini ve birbirinin üzerindeki etkileri gerçek hayatta bilmek zordur. YSA bu ilişkileri otomatik olarak örneklerden öğrenirler. Kullanıcıların bu ilişkileri bilmesi ve ağa söylemesi beklenmemektedir. Geleneksel yöntemlerde bu ilişkileri belirlemek veya yok saymak gerekmektedir. Bu özellik belkide YSA'nın en önemli avantajlarından birisidir.

Benzer şekilde gerçek dünya olayları ve bu olayların arkasındaki faktörlerin birbirleri ile ilişkileri doğrusal olmaz ise bu ilişkileri modellemek çok zordur. O nedenle gerçek hayatta problemleri çözmek için bazı varsayımlar yapmak gerekmektedir. Bu da modellenen sistemin gerçek sisteme uygunluğunu azaltmakta ve gerçek sistemin davranışlarını kontrol altına almayı zorlaştırmaktadır.

Özellikle insan hayatının söz konusu olduğu yerlerde bu sebepten geleneksel sistemleri kullanmak sıkıntı yaratmaktadır. Yukarıdakinin aksine YSA için ilişkilerin doğrusal olup olmaması sorun değildir. İlişkiler örnekler üzerinden tespit edilir. Verilen örnekler üzerinden çalışan sistem gerçeğe yakın sonuçlar vermektedir.

YSA uygulamaları hem pratik hemde maliyet açısından daha hesaplıdır. Sadece örneklerin belirlenmesi basit bir program problemi çözmek için yeterli olmaktadır.

YSA zaman bakımından da verimli çalışır. Örneklerin bulunması ağların oluşturulması olayın öğrenilmesi gerçek zamanda kullanıma alınması çok kısa zamanda olabilmektedir. Ayrıca geleneksel sistemden hızlıdır.

YSA yeni bilgilerin ortaya çıkması ve ortamda bazı değişikliklerin olması durumunda yeniden eğitilebilir. Bazı ağların eğitilmesinde gerek yoktur. Kendileri öğrenerek eğitimi gerçekleştirilebilirler.

YSA paralel çalışabilmesi gerçek zamanlı kullanımını kolaylaştırmaktadır.

Kural setleri tecrübeye göre oluşturulur. Tecrübeye göre örnekler verilir ve eğitim sürdürülür. Eğitim bazen çok uzun zaman da alabilmektedir. Optimum sonuçları garanti etmez. Üretilen sonuçlar ihtimaller içinden iyi bir tanesidir denebilir. Örnekleri bulmak zor ise sonuçlarında doğruluğu tartışılabilir.

**YSA simülatörleri:** Neurosolutions, Matlab, NeurDS, Mactivation, Xerion, BrainWave, PDP++, Z Solutions, Statsoft olarak belirtilebilir.

YSA'da çözüm ararken karşılaşılan temel zorluklardan bazıları çok iyi uyum sağlama (sadece bir örneği öğrenme) ve hiç uyum sağlamama (çözümü yok gibi) problemleridir. Bu durumda varyans ve bias'tan bahsetmek gerekmektedir.

**Bias** : Öğrenme sürecinde giriş ağırlıklarının toplamının sıfır olmasını engeller.

**Varyans** : Veri setindeki değerlerin ortalamaya göre farklılığını gösterir.

**Hata** :  $Bias^2 + Varyans$

### 3.6.1.1.Eğitim Algoritması

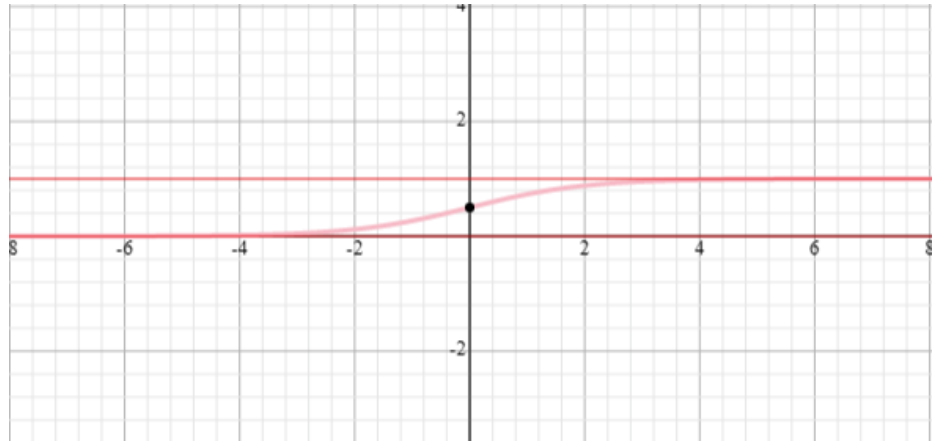
Günümüzde çok fazla eğitim algoritması bulunmaktadır. En hızlı ve kesin sonuç veren algoritmayı bilmek zordur. Çok fazla faktöre bağlıdır. Problemin karmaşıklığı eğitim setindeki nitelik sayısı ağıdaki ağırlık sayısı ve bias değerine göre değişir.

#### 3.6.1.1.1. Sigmoid fonksiyonu

Sigmoid fonksiyonu denklem aşağıdaki şekilde gösterilmektedir.

$$f(x) = 1 / (1 + e^{-x}) \quad (3.13)$$

Sigmoid fonksiyonunun ismi de fonksiyonun S eğrisine benzemesinden kaynaklanır.



Şekil 3.6.1.1.1.1 Sigmoid Fonksiyon

Şekilde görüldüğü gibi fonksiyon 0 ve 1 arasındaki y değerleri için tanımlı olmaktadır. (0,y) değerlerinden önce 0'a sonrasında ise 1'e yakınsamaktadır.

Yapay sinir ağlarında eşik değer aktivasyonu için kullanılır. Sigmoid fonksiyonu için lojistik fonksiyonu tabiri de kullanılabilir.

Doğrusal olmayışı nedeni ile YSA'da aktivasyon fonksiyonu olarak sıkça kullanılabilir. Bunun sebebi, sürekli ve türevinin alınabilir olmasıdır.

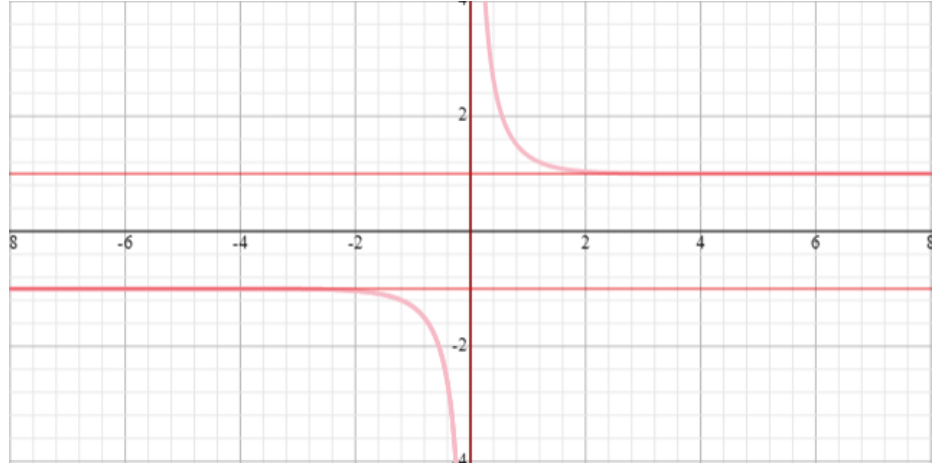
Doğrusal olmaması nedeni ile sigmoid fonksiyonu kullanılan modellerde, lineer regresyonda sıkça kullanılan MSE (Hata Karelerinin Ortalaması) gibi parametreleri kullanamayız.

Sigmoid fonksiyonu kullanılan modellerde, MSE yerine Çapraz eşik adında bir maliyet fonksiyonu kullanılır.

### 3.6.1.1.2. Tanjant Hiperbolik Fonksiyonu

Tanjant Hiperbolik Fonksiyon denklemini aşağıdaki şekilde gösterilmektedir.

$$f(x) = \frac{e^x + e^{-x}}{e^x - e^{-x}} \quad (3.14)$$



Şekil 3.6.1.1.1.2.1 Tanjant Hiperbolik Fonksiyon

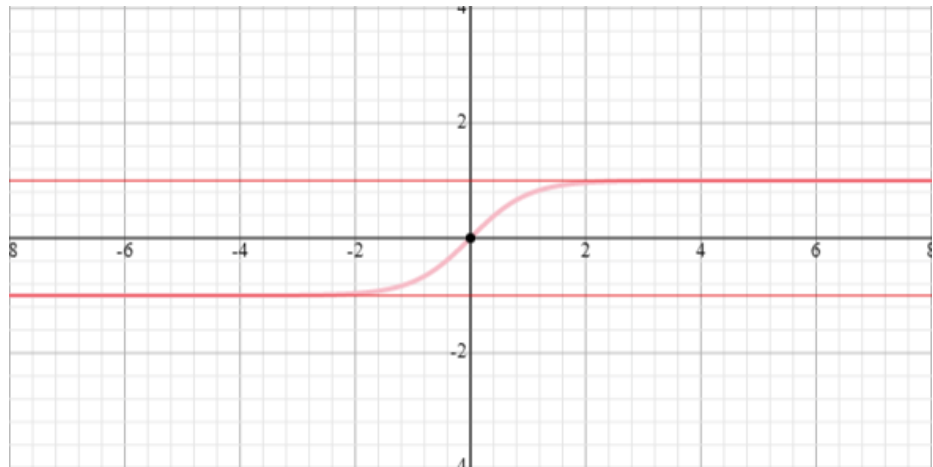
Sigmoid fonksiyona benzer. Fakat çıktı değerleri -1 ve 1 aralığındadır.

### 3.6.1.1.3. Tanjant Sigmoid Fonksiyonu

Tanjant Sigmoid Fonksiyon denklemi 3.15'teki gibi gösterilmektedir.

$$f(x) = \frac{2}{1+e^{-2x}} - 1 \quad (3.15)$$

Bu aktivasyon fonksiyonu için nöron giriş çıkış ifadesi aşağıdaki şekilde ve fonksiyonun değişimi denklem 3.15'te gösterilmektedir. Fonksiyonun dinamik değişim aralığı [-1 1] aralığıdır ve fonksiyon toplam girişe bağlı olarak bu aralıkta lineer olmayan bir değişim gösterir.



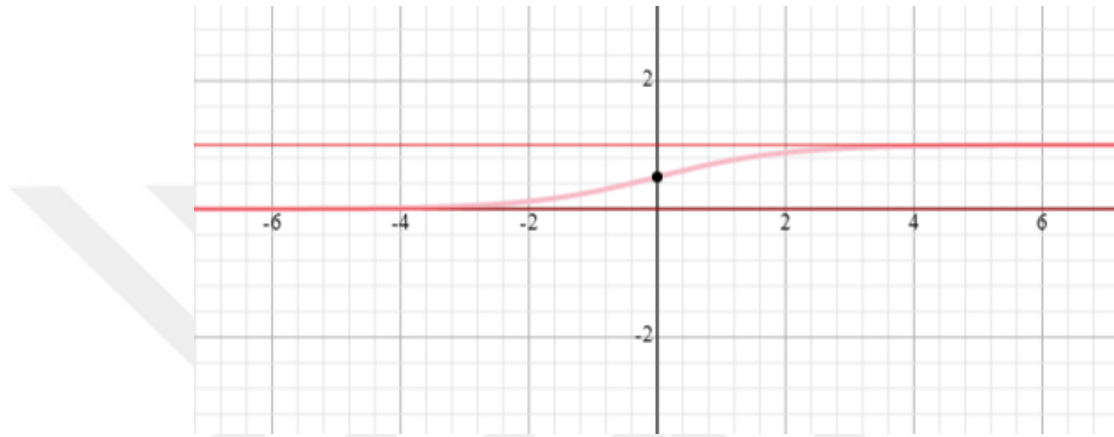
Şekil 3.6.1.1.3.1 Tanjant Sigmoid Fonksiyon

### 3.6.1.1.4. LogSig Fonksiyonu

LogSig Fonksiyon denklemini 3.16'da gibi gösterilmektedir.

$$\text{logsig}(n) = 1 / (1 + e^{-n}) \quad (3.16)$$

Logsig, logaritmik fonksiyondur. Aşağıdaki şekilde gösterildiği gibi çıktı verir. Tahmin amaçlı kullanılan geri yayılım ağ tipinde yaygın olarak kullanılmaktadır. Çünkü bu fonksiyon doğrusal olmayan veriler içinde anlamlı ve etkin sonuçlar vermektedir. [0, 1] arasında sürekli artan bir fonksiyondur.



Şekil 3.6.1.1.4.1 Logaritmik Sigmoid Fonksiyon

### 3.6.1.2. Geri Yayılım Algoritması

İleri beslemeli ve öğreticili olarak eğitilen geri yayılma; katmanlar arasında eksiksiz bağlantıların olduğu bir YSA modelidir. GYA; olması gereken değerden hesaplanan değeri - hatayı - geriye doğru - çıkıştan girişe doğru - azalttığı için bu isim verilmiştir. GYA'da verilerin sisteme dahil edildiği giriş katmanı, hesaplamaların yoğunlaştırıldığı gizli katman ve sonuçların olması gereken değerler ile kıyaslandığı çıkış katmanı bulunur. Hesaplamaların yoğunlaştırıldığı ara katman probleme göre değişiklik gösterir. Katmanlar arasındaki nöronlarda geri yayılım gerçekleşir. GYA sayesinde aktif bir yapı sergilenir. Model içerisinde YSA'da bir sinir hücrenin elde ettiği değer diğer hücrelere giriş olarak sunulur. Sinir hücrelerinin başlangıç ağırlıkları rasgele atanır. Her sinir hücresinin ağırlıklandırılmış değeri hesaplandıktan sonra çıkış katmanındaki bu sonuç olması gereken sonuç ile karşılaştırılarak mevcut hata geri yayılım ile minimize edilmeye çalışılır ve ağa yansıtılır. Belirli bir hata derecesine

ininceye kadar iterasyon işlemine devam edilir ve istenilen sonuca yaklaşılmca ağın eğitim aşaması tamamlanmış olur.

Giriş katmanı; bilindiği gibi giriş verilerinin ağı entegre edildiği kısımdır. Giriş katmanındaki sinir hücrelerindeki veri değerleri ileri beslemede gizli katmana sevk edilir. Gizli katmandaki her bir sinir hücresi, giriş değerlerini ağırlıklı olarak hesaplayarak toplam değeri bulur ve hesapladığı değerleri bir aktivasyon fonksiyonu ile dönüştürerek gizli katmana veya çıkış katmanına gönderir.

Giriş katmanında nöron sayısı, giriş sayısı kadardır ve her bir nöron bir veriyi hapseder. Daha sonra değer, hesaplama olmaksızın ilerleyen ara katmana geçer. Ara katman gizli katman olarakta bilinir ve ağı hesaplamalardaki keskinliğini artırmak için eklenmiş bağlı nöronlardan oluşur. Birden çok ara katmanlı yapılar sisteme göre kurulabilir. Ara katman sayısı ve o katmandaki nöron adedi, istenilen sonuca ve karmaşıklığa göre değişiklik gösterir.

Ara katman; giriş kısmından gelen veriyi probleme uygun şekilde işleyerek diğer bağlı bulunduğu katmana iletir. Veri sayısı çok ise ve öznelik sayısı da oldukça fazla ise ara katmanda hesaplamalar uzar.

Son katman olan çıkış katmanı; YSA modelinin en son katmanıdır. Ara katmandan alınan işlenmiş veriyi kullandığı fonksiyonla işleyerek girişe bağlı hata miktarını bulur. Son katmanındaki nöron adedi, çalışılan problemin türüne göre değişiklik arz eder. Sonuç katmanından elde edilen veriler YSA'nın problem için çıkış değerleridir.

Bahsedilen işlemlerin sonucunda üretilen edilen çıkış değerleri, hedeflenen sonuç ile mukayese edilir. YSA çıkışı ve hedeflenen çıkış arasındaki fark, ayrı ayrı hesaplanır ve hata olarak isimlendirilir. Hesaplanan hata değeri, her çıktı düğümüne giren ara katmandaki sinir hücrelerine geri aktarılır. Bu şekilde ara katmandaki değerlerin her biri toplam hatanın sadece hesaplanan bir kısmını içerir. Bu işlemler katmanlardaki düğümler toplam hatanın belirli bir kısmını içerecek şekilde giriş katmanına kadar tekrarlanır. Elde edilen hata değerleri baz alınarak, bağlantı değerleri her nöronda yeniden düzenlenir. Bu düzenleme tüm verilerin iyileştirilebileceği bir duruma ağı dönüştürülmesini sağlar.

Katmanlar arasındaki ağırlık değerleri eğitimi tamamlanmış ağıdan alınarak kullanılmak üzere yedeklenir. GYA algoritmasının uzun eğitim süresine sahip

olabilmesi bir problem teşkil edebilir. Sürenin dahada kısaltılabilmesi için hatayı en kısa sürede azaltmak adına yeni metotlar geliştirilmektedir.

### 3.6.2. Destek Vektör Makineleri

Genelleme performansının en belirgin seçiciliğe sahip olduğu sınıflandırma problemlerinde DVM ile geliştirilen makine öğrenimi algoritmaları büyük kullanım görmektedir. Performansı; eğitim için seçilen verilere, bağımlı olmayan niteliklere, model seçimine ve tercih edilen parametrelere göre değişkenlik gösterir. Problem çözümü için gereken kriterler değerlendirildiğinde, verilerden mantıklı bilginin elde edilmesi ve doğru bilgiye ulaşma, algoritmanın genelleme performansı ile doğru orantılıdır (Sevgi AYHAN).

Başka bir ifadeyle, algoritmanın genelleme performansı iyiyse elde edilen bilgi de o kadar gerçeğe yakın olacaktır. Sınıflandırma problemlerinde en başarılı makine öğrenimi algoritmalarından biri kuşkusuz DVM'dir. DVM, çoğu sınıflandırma çözümünde başarıyla sonuç vermiş ve genelleme performansı oldukça yüksek ve etkili makine öğrenimi algoritmalarından birisi olarak literatürdeki yerini almıştır.

DVM'nin en önemli avantajlarından bir tanesinde kuşkusuz, sınıflandırma problemini kareli en iyileme problemine dönüştürüp çözme işlemidir. Bu şekilde öğrenme aşamasındaki işlem sayısı azaltılmakta ve benzer algoritmalara göre daha hızlı çözüme kavuşulmaktadır. DVM, Bu nedenden dolayı, büyük boyutlu veri setlerinde büyük kolaylık sağlamaktadır. En iyileme bazlı olduğundan sınıflandırma performansı, hesaplama derecesi ve kullanılabilirlik kriterleri açısından diğer tekniklere göre daha başarılıdır. Çekirdek fonksiyonu seçimi ve parametre optimizasyonu, çeşitli veri setleri için sınıflandırma probleminin çözümüne ilişkin DVM'nin uygulanması sürecinde önemli rol oynamaktadır. DVM makine öğrenmesi algoritmaları içerisinde, risk küçültme prensibine göre çalışan dış bükey en iyilemeye dayalı bir yapı sunar. DVM, dağılımdan bağımsız öğrenme algoritmaları içerisinde de veriye ilişkin herhangi bir birleşik fonksiyonu bilgisine ihtiyaç duymaz. DVM, Li tarafından örüntü tanıma ve sınıflandırma problemlerinin çözümü için düşünülmüş ve geliştirilmiştir. DVM'in temelleri Vapnik-Chervonenkis (VC) teorisine diğer bir deyişle istatistiksel öğrenme teorisine dayanmaktadır (Li ve ark., 2009). DVM'de çözüm, sınıfları en ideal şekilde ayıracak bir düzleminin inşası sürecinden sonra elde edilir. Bir diğer bakış açısıyla, uyumsuz sınıflara ait çizgiler arasındaki mesafeyi maksimize etmektir. DVM iki sınıflı veya çok sınıflı problemlerin çözümü için geliştirilmiş makine öğrenmesi

algoritmaları içerisinde yer alır. DVM, veri setinin doğrusal olarak ayrılama durumuna göre olarak iki kısma ayrılmaktadır. DVM veri kümesinin sınıflara ayrıldığı bir düzlemde oluşmaktadır. Veri setinin bölünemediği durumlarda DVM bir kernel çekirdek fonksiyonu kullanır ve kümeyi bölünebilir ve işlem yapılabilir hale getirir. DVM açıkça ölçülebilir. Pratik problem çözümü için kullanılabilir.

Veri setleri mantıksal bölümlere ayrılır.

n-boyutlu şablon  $x$  n noktaya sahiptir,

$$x = (x_1, x_2, \dots, x_n), (x_i \in \mathbb{R} \text{ for } i = 1, 2, \dots, n)$$

Tüm şablonlar  $x$ -j setine kayıtlıdır.

$$y_j \in \{-1, +1\}$$

Eğitim setleri  $m$  ye bağlı  $T$  şablonları kendi sınıf-setleriyledir.

$$T = (\text{Hasan ve ark.}, 2017)$$

$S$  uzayında bir nokta düşünün  $x$  şablonu dahil edilmiş

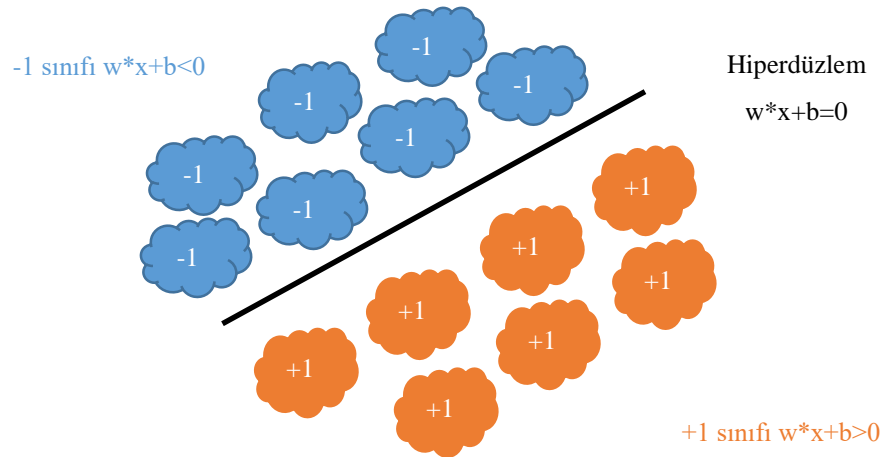
$$(x_1, x_2, \dots, x_m \in S),$$

Hiper düzlem  $S$  şu şekilde gösterilebilir.

$$\{x \in S \mid w \cdot x + b = 0\}, w \in S, b \in \mathbb{R}$$

$$w \cdot x = \sum_{i=1}^n w_i x_i \quad (3.23)$$

$w \cdot x$  noktası Şekil 3.6.2.1'deki gibi tanımlanır.



Şekil 3.6.2.1 Hiperdüzlem görünümü

Şekil 3.6.2.1’de ve denklem 3.24’te görüldüğü gibi sınıflandırıcı tüm eğitim parametrelerini sınıflar. 2 farklı bölge tanımlanmıştır. +1 olanlar  $w*x+b>0$ , -1 olanlar  $w*x+b<0$  olacak şekilde hiperdüzlemde sınıflanır.

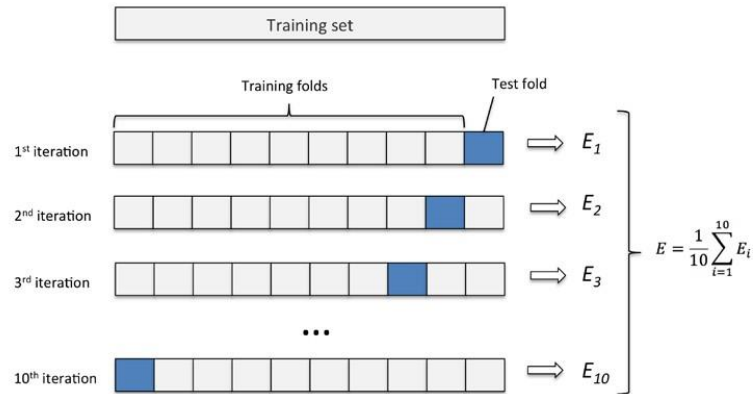
DVM eğitim süreçleri tamamlandığında üyelikleri bulmak için hazırlanmıştır. Yeni örüntüler eğitimdekilerden farklı olabilir.  $x_k$  örüntüsünün sınıfını tanımlayan eşitlik:

$$\text{sınıf}(x_k) = \begin{cases} +1 & \text{eğer } w * x_k + b > 0 \\ -1 & \text{eğer } w * x_k + b < 0 \end{cases} \quad (3.24)$$

DVM gözetimli denetimli öğrenme metotudur. Gelişigüzel olarak etiketlenmiş veriler ile başlar. Sonra hesaplamalar başlar. Etiketler gerekirse değiştirilir. Doğruluğu artırmak için değişiklikler yapılır.

### 3.6.3. K-katmanlı Çapraz-Doğrulama

K-katmanlı çapraz doğrulama test sonuçlarının daha değerli olmasını sağlar. K-katmanlı da orijinal veri k katmana ayrılır. Bir katman test için diğerleri de eğitim için kullanılır. Çapraz doğrulama her katmanın ayrı ayrı kullanılmasından elde edilir. K adet doğrulama yapılır. Bu katmanlar sadece bir kez kullanılır. Ortalama değerleri test doğruluk sonucunu vermektedir. Şekil 3.6.3.1’de gösterildiği gibi veri seti 10 parçaya ayrılmıştır.



Şekil 3.6.3.1 K-Katmanlı Çapraz Doğrulama (AKIN, 2017)

VM uygulamalarında, başvuru yöntemin başarısının mukayese edilebilmesi adına, kullanılan veri kümesi eğitim kümesi ve test kümesi olarak ikiye ayrılmaktadır. Eğitim ve test kümeleri içeriklerinin rastgele olarak doldurulması de farklı bir yöntemdir. K-katmanlı çapraz doğrulama yönteminde seçilen eğitim ve test kümeleri belirli bir sırayla değiştirilir. K-katlamalı çapraz doğrulama yönteminde öncelikle bir

k değeri - kesit - seçilir. Literatürde en çok tercih edilen k değeri 10'dur. Bu değer artarsa algoritmanın performansı düşer.

YSA problem çözümlerinde basit yaklaşım; verisetini %75'lik kısmını eğitim, %25'lik kısmını da test için ayırmaktır. Veri kısımlara ayrılırken verinin değişkenliğine bağlı olarak verisetinin eğitim ve test bölümlerinde varyans ve hatalar oluşabilmektedir. K-katmanlı çapraz doğrulama, veri setini istenilen bir k sayısı kadar eşit kısımlara paylaştırır, bu şekilde her bir kısım hem eğitim hem de test için kullanılmış olur. Bu sayede bölme ve kısımlamadan kaynaklanan sapmalar ve farklılıklar en aza inmiş olur. Modeli k adedince eğitmek, test etmek ve ayrı ayrı sonuçlardan hesaplama yapmak gibi sisteme ayrı bir yük getirir ve zaman ister. Eğitim ve test kısımları küçük olan hesaplama maliyeti az olan küçük ve orta boyutlu veriler için sorun teşkil etmesede büyük hacimli veri kümelerinde hesaplanma ve zaman yönünden maliyetli olabilir.

Modelde olabilecek çok iyi uyum sağlamadan kaçınmak ve modelin görebileceği bir veri setinde yanlış performans göstermesini engellemek içinde k kısım tercih edilebilmektedir. Veri kümesi rassal olarak k kısma bölünür. K-1 parçası eğitim için işlem görür 1 parçasıda test seti için kullanılır ve işlemler k defa tekrarlandırılır. Her bitimde elde edilen sonuçlar toplanır; ortalaması hesaplanır ve modelin performansı hakkında değerlendirme yapılır.

Çapraz doğrulama, uygulanan modelde yapılan hesaplama hatalarını daha iyi anlayabilmek için model seçiminde kullanılan bir tekniktir. Çapraz doğrulamanın bilinen mantığı, veri setlerinden doğrulama kısımları olarak bilinen gözlem örnek bölümlerini oluşturmaktır. Çapraz doğrulamanın en basit halidir elde tutma yöntemidir. Veri kümesi k alt kümeye ayrılmıştır ve elde tutma yöntemi k kez tekrarlanmalıdır. Her turda, k'ncı test kümesi olarak işleme sokulur kalan k-1 alt küme elemanları bir eğitim kümesi oluşturmak için bir arada kalır. Her k turundaki ortalama hatası hesaplanır ve sonuç sistem performansını verir.

### 3.7. Performans Değerleme

10-Katlı çapraz doğrulama metodu performans değerlendirme için kullanılmıştır. YSA algoritmasında ve DVM algoritmasında bu metot tercih edilmiştir. Bu metot ile YSA ve DVM doğruluk oranları test edilmiştir. Veri seti 10 kısma ayrılır bir parçası test için geri kalan 9 kısım ise eğitim için kullanılır. Test 10 kez tekrarlanır ve hata

oranlarının ortalaması hesaplanır. Bu metotla hata hesaplanmış olur. Tüm kayıtlar kullanıldığı için test mantıklı şekilde tamamlanmış olur.

### 3.8. Sınıflandırıcıların Değerlendirilmesi

Bu çalışmada veri setleri için sınıflandırma doğruluğu 3.17 ve 3.18'deki denklemle hesaplanır:

$$\text{doğruluk } (T) = \frac{\sum_{i=1}^N \text{değer}(t_i)}{N}, t_i \in T \quad (3.17)$$

$$\text{değer } (t_i) = \begin{cases} \text{eğer sınıflandır } (t_i) \equiv \text{doğru sınıflandırma ise} & 1 \\ \text{diğer durumlarda ise} & 0 \end{cases} \quad (3.18)$$

T test setindeki veriler olmak üzere N test örnek sayısını göstermektedir.

### 3.9. Hassaslık ve Duyarlılık Analizi

Sınıflandırma yöntemlerinin yorumlanmasında doğruluk ölçütü kullanılır. Doğruluk ölçütünün değerlendirilmesinde karışıklık matrisi kullanılmaktadır. Karışıklık matrisi bileşenleri yapılan çalışmaya göre anlamlandırılmaktadır. Bu çalışma kapsamında incelenen kalp hastalığı sınıflandırma işlemi için karışıklık matrisi bileşenleri şu şekilde tanımlanmaktadır.

- **Doğru Pozitif - DP:** Kalp hastalığına sahip kişiler kalp hastalığı var şeklinde doğru bir biçimde sınıflandırılmıştır.
- **Yanlış Pozitif - YP:** Sağlıklı kişiler yanlış olarak kalp hastalığı var şeklinde sınıflandırılmıştır.
- **Doğru Negatif - DN:** Sağlıklı kişiler doğru bir biçimde sağlıklı şekilde sınıflandırılmıştır.
- **Yanlış Negatif - YN:** Kalp hastalığına sahip kişiler yanlış bir biçimde sağlıklı şekilde sınıflandırılmıştır.

Duyarlılık Hassasiyet Analizleri için şu formüller kullanılır:

$$\text{Duyarlılık Değeri } (\%) = \frac{DN}{DN+YP} * 100 \quad (3.19)$$

$$\text{Hassasiyet Değeri } (\%) = \frac{DP}{DP+YN} * 100 \quad (3.20)$$

$$\text{Pozitif Doğruluk Değeri } (\%) = \frac{DP}{DP+YP} * 100 \quad (3.21)$$

$$\text{Negatif Doğruluk Değeri } (\%) = \frac{DN}{DN+YN} * 100 \quad (3.22)$$

Denklem 3.19’da gösterildiği gibi Duyarlılık analizi için Doğru Negatif ve Yanlış Pozitif değerleri kullanılır.

Denklem 3.20’de gösterildiği gibi Hassasiyet analizi için Doğru Pozitif ve Yanlış Negatif değerleri kullanılır.

Denklem 3.21’de gösterildiği gibi Pozitif Doğruluk Değeri analizi için Doğru Pozitive ve Yanlış Pozitif değerleri kullanılır.

Denklem 3.22’de gösterildiği gibi Negatif Doğruluk Değeri analizi için Doğru Negatif ve Yanlış Negatif değerleri kullanılır.

### 3.10. Kullanılan Veri Setleri ve Özellikleri

Kalp hastalığı teşhisine yardımcı bir sistem için geliştirdiğimiz Hibrit Sistemde ilk olarak UCI’den alınmış Statlog kalp hastalığı veri seti kullanılmıştır(Dua, 2017). Çizelge 3.10.1’de görüldüğü gibi UCI Statlog kalp hastalığı veri serinin öznitelikleri açıklamalarıyla birlikte verilmiştir. Öznitelik sütununda yer alan özniteliklerin açıklaması “Tanım” sütununda ve alabileceği değerde “Değeri” sütununda anlatılmıştır. Yaş, cinsiyet, göğüs ağrısı tipi, kan basıncı, kolesterol, kan şekeri, ekg sonucu, kalp atışı, anjin geçirip geçirmediği, ekg de st eğrisinin yönü, tıkalı damar grafisi, damar durumu gibi özniteliklere bağlı olarak yapılan anjiyografi işleminin sonucu “num” ile gösterilen kalp hastalığı teşhis sütununda ifade edilmiştir.

Çizelge 3.10.1 Kalp Hastalığı Veri Setinin Öznitelik Değerleri

Öznitelik	Tanım	Değeri
Age	Yaş	Sayısal
Sex	Cinsiyet	1 erkek; 0 bayan
Cp	Göğüs ağrısı tipi	1 tipik anjin, 2 tipik olmayan anjin, 3 anjin dışı ağrı, 4 belirtisiz
Trestbps	Kan basıncı (mmHg)	Sayısal
Chol	Kolesterol (mg/dl)	Sayısal
Fbs	Kan şekeri	120 mg/dl, 1 evet, hayır 0
Restecg	EKG sonuçları	0 Normal, 1 ST-T dalgası anormal, 2 LV Belirli sol ventrikül hypertrophy si mevcut Estes' kriteri
Thalach	Kalp atışı en fazla	Sayısal
Exang	Anjin yaşadımı?	1 evet, 0 hayır
Oldpeak	Kalp krizi geçirme durumu	Sayısal

Slope	ST değerinin yönü nedir?	1 yukarı, 2 yatay, 3 aşağı
Ca	Büyük damar sayısı floroskopi ile renklenmiş	Sayısal
Thal	Talyum sintigrafisi kusurları	3 normal,
	Radyopak bir maddenin enjeksiyonundan sonra	6 düzeltilemez,
	x-ışınları ile kan damarlarının incelenmesi	7 düzeltilebilir
Num	Kalp hastalığının teşhisi	Eğer ana damarlarda %50 den fazla damar genişliği daraldı ise, 0 yok, 1 var.

Kalp hastalığı teşhisi için geliştirdiğimiz Hibrit Sistem TFEBT (Krzysztof J. Cios, 2001) veri setinde de denenmiştir. Bu veri setinde tıkalı damar veya kalbin hasta olan kısmı TFEBT grafiklerine bakılarak uzmanlar tarafından çözümlenmektedir. TFEBT’de vücuda enjekte edilen radyoaktif maddelerden yansıyan ışımlar, ışımları farkedebilen bir cihaz tarafından alınır ve bilgisayar üzerinde işlenerek kalp hastalığı tahminleri gerçekleştirilir. Tıkalı veya kan akışı zayıf olan hasta yapılar aynı bölgede yer alan normal yapılara bakılarak farkedilmiş olur. Bu amaç için hazırlanmış veri seti de UCI’den alınmıştır (Dua, 2017). Bu veri seti makine öğrenmesi kategorisi içerisinde yer almaktadır; test ve eğitim olarak iki kısımdan oluşmaktadır. Veri setinde toplamda 267 adet hasta kaydı bulunmaktadır. Bu kayıtlardan 80 tanesi eğitim amaçlı hazırlanmıştır. Kayıtlardan 187 tanesi de test için oluşturulmuştur. Çizelge 3.10.2’de görüldüğü üzere veri seti 44 öznitelikten oluşturulmuştur. İlk öznitelik TFEBT resmine bağlı hastalık sonucunu göstermektedir. F1R-F22S arası öznitelikler resmin bir parçasının aldığı ikili değerden (0 veya 1) birini temsil eder.

Çizelge 3.10.2 TFEBT Öznitelikleri

Öznitelik	Sayısal	Teşhis Sonucu
1	Sayısal	0,1 Normal veya Değil
2	F1R:	Bölge 1 de dinlenme
3	F1S:	Bölge 1 de stres
4	F2R:	Bölge 2 de dinlenme
5	F2S:	Bölge 2 de stres
6	F3R:	Bölge 3 te dinlenme
7	F3S:	Bölge 3 te stres
8	F4R:	Bölge 4 te dinlenme
9	F4S:	Bölge 4 te stres
10	F5R:	Bölge 5 te dinlenme
11	F5S:	Bölge 5 te stres
12	F6R:	Bölge 6 da dinlenme
13	F6S:	Bölge 6 da stres
14	F7R:	Bölge 7 de dinlenme

15	F7S:	Sayısal	Bölge 7 de stres
16	F8R:	Sayısal	Bölge 8 de dinlenme
17	F8S:	Sayısal	Bölge 8 de stres
18	F9R:	Sayısal	Bölge 9 da dinlenme
19	F9S:	Sayısal	Bölge 9 da stres
20	F10R:	Sayısal	Bölge 10 da dinlenme
21	F10S:	Sayısal	Bölge 10 da stres
22	F11R:	Sayısal	Bölge 11 de dinlenme
23	F11S:	Sayısal	Bölge 11 de stres
24	F12R:	Sayısal	Bölge 12 de dinlenme
25	F12S:	Sayısal	Bölge 12 de stres
26	F13R:	Sayısal	Bölge 13 de dinlenme
27	F13S:	Sayısal	Bölge 13 de stres
28	F14R:	Sayısal	Bölge 14 de dinlenme
29	F14S:	Sayısal	Bölge 14 de stres
30	F15R:	Sayısal	Bölge 15 de dinlenme
31	F15S:	Sayısal	Bölge 15 de stres
32	F16R:	Sayısal	Bölge 16 da dinlenme
33	F16S:	Sayısal	Bölge 16 da stres
34	F17R:	Sayısal	Bölge 17 de dinlenme
35	F17S:	Sayısal	Bölge 17 de stres
36	F18R:	Sayısal	Bölge 18 de dinlenme
37	F18S:	Sayısal	Bölge 18 de stres
38	F19R:	Sayısal	Bölge 19 da dinlenme
39	F19S:	Sayısal	Bölge 19 da stres
40	F20R:	Sayısal	Bölge 20 de dinlenme
41	F20S:	Sayısal	Bölge 20 de stres
42	F21R:	Sayısal	Bölge 21 de dinlenme
43	F21S:	Sayısal	Bölge 21 de stres
44	F22R:	Sayısal	Bölge 22 de dinlenme
45	F22S:	Sayısal	Bölge 22 de stres

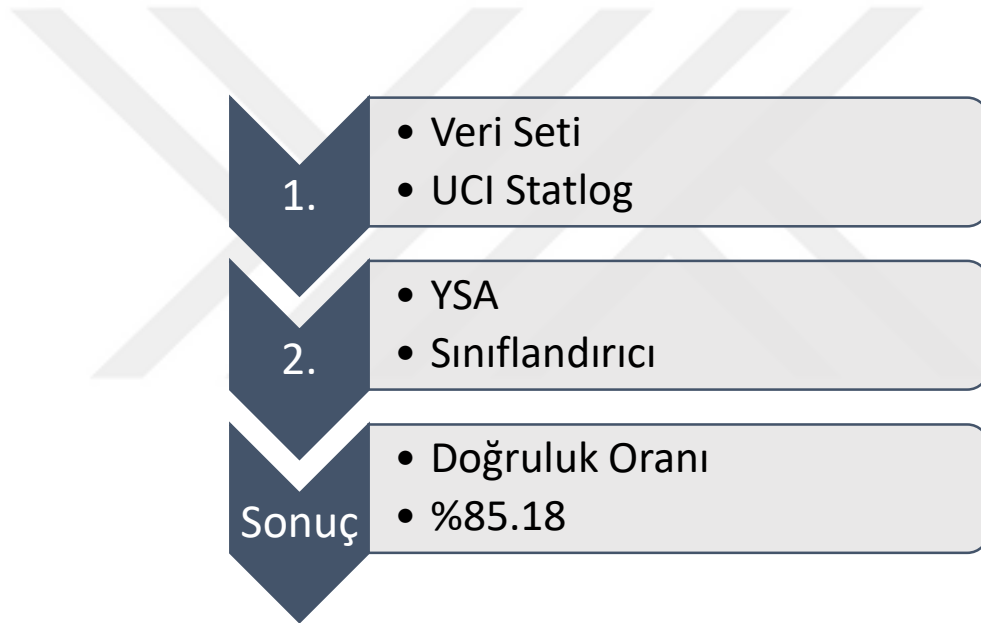
#### 4. ARAŞTIRMA BULGULARI VE TARTIŞMA

Bu bölümde tez çalışmasında elde edilen sonuçlar verilmiştir. Yapılan diğer çalışmalara atıflar yapılmış ve önerilen sistem ile diğer çalışma sonuçları tartışılmıştır. Hibrit sistemin diğer çalışmalara göre üstün yönleri de bu kısımda aktarılmıştır. Farklı sonuçların muhtemel nedenleri yine bu kısımda aktarılmıştır. Elde edilen sonuçların geçerliliği ve uygulanabilirliği yine bu başlık altında anlatılmıştır.

Ham veri setimizin YSA ile DVM ile SNBM + YSA ile SNBM + DVM ile Hibrit Sistem + YSA ile ve Hibrit Sistem + DVM ile bulunan doğruluk oranları tablolar halinde verilmiştir. Çalışmada Intel(R) Core(TM) i5-3210M CPU @ 2.50 GHz makinede işlemler gerçekleştirilmiştir.

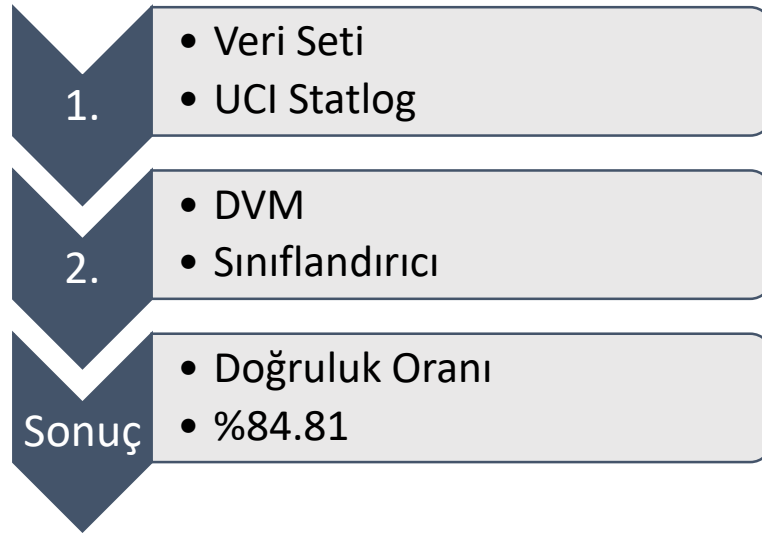
Tez konusu kalp krizi ve felçle sonuçlanabilecek kalp hastalıklarını tahmin edebilen tıbbi karar destek sistemini hayata geçirebilmek olmuştur. Bu amaç için planlanan hedefe ulaşabilmek adına sağlık alanında çalışılabilecek veri setleri araştırılmıştır. California Irvine University (UCI) veri ambarından çalışmada kullanılmak üzere Statlog kalp hastalığı ve TFEBT veri setleri alınmıştır. Araştırmacılarında tercih ettiği bu veri setlerindeki öznitelikler çalışmanın gereksinimlerini karşıladığı için seçilmiştir. Veri setlerine ayrıklaştırma, TBA ve Normalizasyon adımları çeşitli kombinasyonlar ile uygulanmıştır. En son ön işleme adımından sonra veriler sınıflandırıcılara yüklenmiştir. Matlab ile belli bir konfigürasyona göre adım adım işletilen programda Statlog veri setindeki veriler için Hibrit Sistem (SNBM + PCA + Normalizasyon) ve YSA sonucumuzun doğruluk oranı %88.89 olarak bulunmuştur. Sadece YSA ile yapılan çalışmanın doğruluk oranı ise %85 olarak bulunmuştur. İkinci sınıflandırıcı olan DVM sınıflandırıcısı ise hibrit sistemden gelen verileri işleyerek %87.77 doğruluk oranına ulaşmıştır. Sadece DVM ile yapılan çalışma sonucundaki doğruluk oranı ise %84 olarak bulunmuştur. İkinci veri seti (TFEBT) ile yapılan çalışmada ise TBA + SNBM ve Normalizasyon algoritmalarının ardından bulunan doğruluk oranı %89.17 olarak bulunmuştur. TBA + SNBM + Normalizasyon ve YSA ile bulunan doğruluk oranı ise %89.91 olarak bulunmuştur. Veri seti ön işleme sonucunda YSA ile sınıflandırıldığında doğruluk oranı yüzde 88.23 olmuştur. Veri seti TBA + SNBM'den sonra YSA sınıflandırıcısına yüklendiğinde doğruluk oranı yüzde 92.51 olmuştur. Sonuç 10-katlı çapraz doğrulama ile test edilmiştir. Veri seti ayrıca Matlab programında DVM ile de sınıflandırılmıştır.

DVM ile yapılan sınıflandırmada doğruluk oranı yüzde 75.27 bulunmuştur. Veri seti TBA, SNBM ve Normalizasyon adımlarından sonra DVM sınıflandırıcısına yüklendiğinde doğruluk oranı yüzde 89.17 olarak bulunmuştur. Veri seti YSA ile sınıflandırıldığında (gizli katmandaki düğüm sayısı 14 ve iterasyon sayısı ise 43) ise doğruluk oranı %79.40 olarak bulunmuştur. Hibrit Sistem ve YSA ile sınıflandırıldığında (gizli katmandaki düğüm sayısı 14 ve iterasyon sayısı ise 43) ise doğruluk oranı %89.91 olarak bulunmuştur. Sonuç 10 kat çapraz doğrulama ile test edilmiştir. Görüldüğü gibi Hibrit Sistem YSA ve DVM sınıflandırıcılarının sonuçlarına olumlu katkı sağlamıştır. Geliştirilen sistemde Veri Seti (VS)'nin sadece YSA sınıflandırıcısı ile bulunan doğruluk oranı Şekil 4.1'de gösterilmiştir.



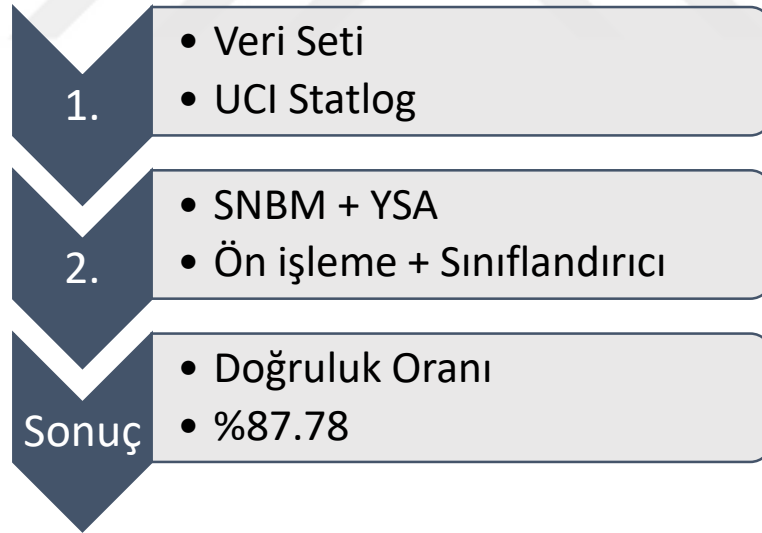
Şekil 4.1 Veri Seti'nin YSA sınıflandırıcısı ile sonucu

Geliştirilen sistemde VS'nin sadece DVM sınıflandırıcısı ile bulunan doğruluk oranı Şekil 4.2'de gösterilmiştir.



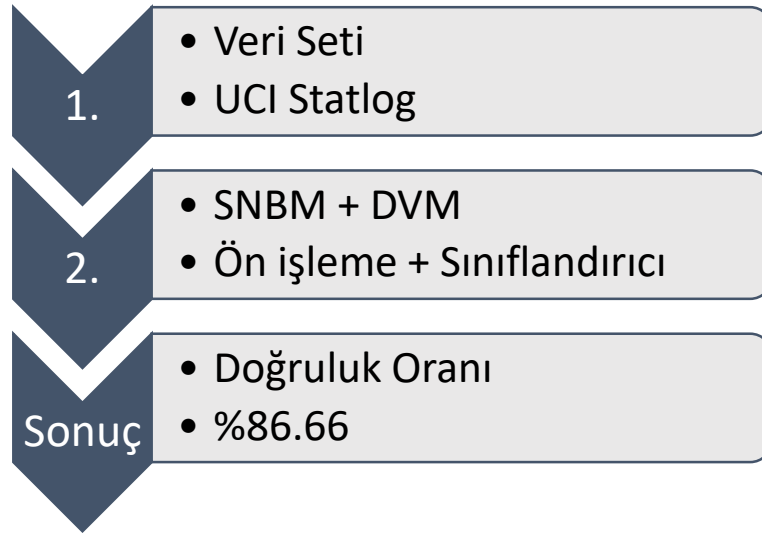
Şekil 4.2 Veri Seti'nin DVM sınıflandırıcısı ile sonucu

Geliştirilen sistemde VS'nin SNBM ön işleminin ardından YSA sınıflandırıcısı ile bulunan doğruluk oranı Şekil 4.3'de gösterilmiştir.



Şekil 4.3 Veri Seti'nin SNBM + YSA sınıflandırıcısı ile sonucu

Geliştirilen sistemde VS'nin SNBM ön işleminin ardından DVM sınıflandırıcısı ile bulunan doğruluk oranı Şekil 4.4'te gösterilmiştir



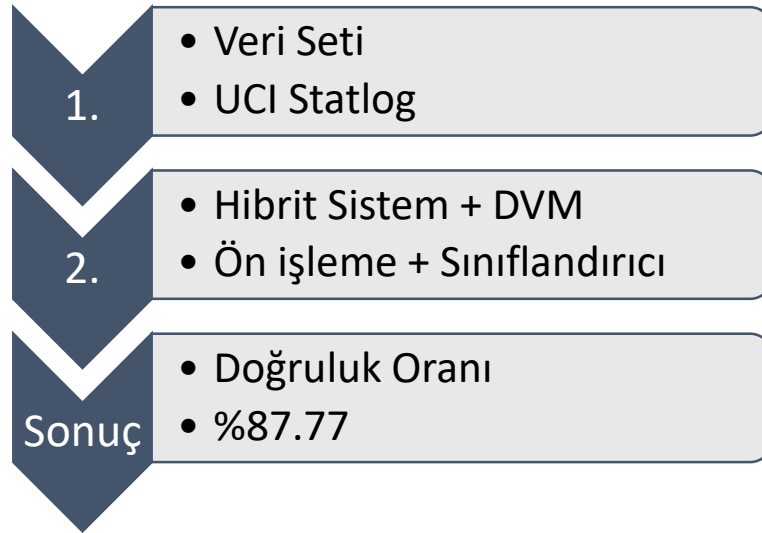
Şekil 4.4 Veri Seti'nin SNBM + DVM sınıflandırıcısı ile sonucu

Geliştirilen sistemde VS'nin Hibrit Sistemden sonra YSA sınıflandırıcısına yüklendikten sonraki doğruluk oranı Şekil 4.5'de gösterilmiştir



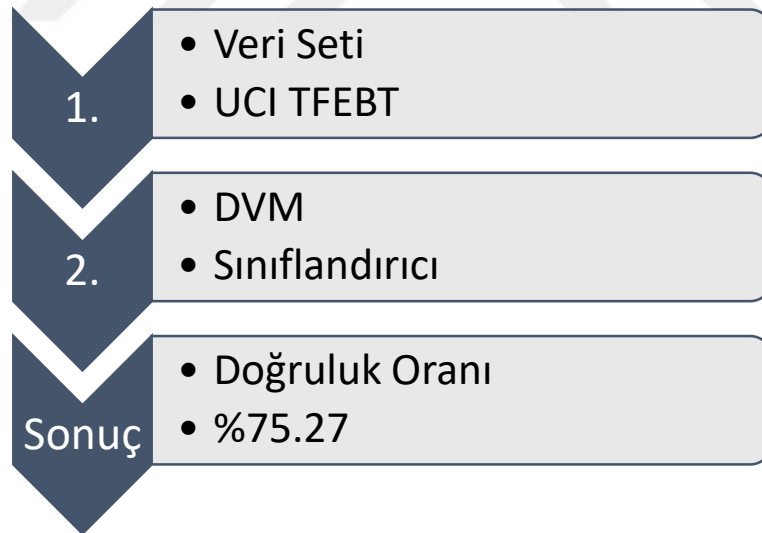
Şekil 4.5 Veri Seti'nin Hibrit Sistem + YSA ile sonucu

Geliştirilen sistemde VS'nin Hibrit Sistem'den sonra DVM sınıflandırıcısına yüklendikten sonraki doğruluk oranı Şekil 4.6'da gösterilmiştir.



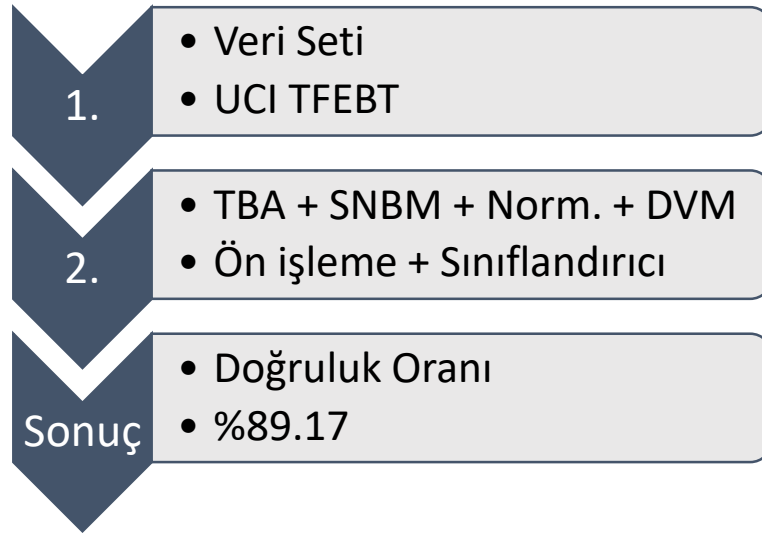
Şekil 4.6 Veri Seti'nin Hibrit Sistem + DVM ile sonucu

Geliştirilen sistemde TFEBT veri setinin DVM sınıflandırıcısına yüklendikten sonraki doğruluk oranı Şekil 4.7'te gösterilmiştir.



Şekil 4.7 Veri Seti'nin DVM ile sonucu

Geliştirilen sistemde TFEBT veri setinin Hibrit Sistem'den sonra DVM sınıflandırıcısına yüklendikten sonraki doğruluk oranı Şekil 4.8'de gösterilmiştir.



Şekil 4.8 Veri Seti'nin TBA + SNBM + Norm. ve DVM ile sonucu

Geliştirilen sistemde TFEBT veri setinin Hibrit Sistem'den sonra YSA sınıflandırıcısına yüklendikten sonraki doğruluk oranı Şekil 4.9'da gösterilmiştir.



Şekil 4.9 Veri Seti'nin Hibrit Sistem + YSA ile sonucu

Çizelge 4.1 DVM (Statlog) Parametre Değerleri

Parametreler	Değer
Metot	Destek Vektör Makineleri (DVM)
Optimizasyon algoritması	SMO
Değerlendirme metodu	k-katmanlı çapraz doğrulama (10-katmanlı CV)
kN	500
Kernel Fonksiyonu	Doğrusal
TolKKT	1.0000'e-003
Maksimum Adım	15000
Kernel Önbellek Limiti	5000
Başlangıç Değeri	Rastsal

Çizelge 4.1'de DVM metoduna ait parametre değerleri verilmiştir. Sıralı Minimal En İyileme (SMO) optimizasyon algoritması kullanılmıştır. K-katmanlı çapraz doğrulama değerlendirme metodu kullanılmıştır. Kernel fonksiyonu doğrusaldır. Başlangıç değerleri rastsaldır.

Çizelge 4.2 YSA ve SNBM+YSA (Statlog) Parametre Değerleri

MATLAB Mimari Başlık	YSA	SNBM + YSA
Katman Sayısı	3	3
Girdi nöron sayısı	13	13
Gizli katmandaki düğüm sayısı	14	14
Çıktı nöron sayısı	1	1
Başlangıç değeri ve Bias değeri	Rastsal	
Aktivasyon Fonksiyonları	Tansig-fonksiyon (gizli katman)	Tansig-fonksiyon (çıktı katmanı)
Öğrenme fonksiyonu	Momentumun açısız azalması metodu	
Hedef	0,000001	

Çizelge 4.2'de YSA sınıflandırıcısının matlab ortamındaki parametreleri görülmektedir. Katman sayısı 3'tür. Girdi nöron sayısı 13 öznelikten dolayı 13 olmuştur. Gizli katmandaki düğüm sayısı 14 olduğu zaman en iyi sonucu vermiştir. Çıktı katmanı 1 hücreden oluşmuştur. Başlangıç ve bias değerleri rastsaldır. Aktivasyon fonksiyonu tanjant sigmoit fonksiyondur.

Çizelge 4.3 YSA ve Hibrit + YSA (TFEBT) Parametre Değerleri

MATLAB Mimari Başlık	YSA	Hibrit + YSA
Katman Sayısı	3	3
Girdi nöron sayısı	44	44
Gizli katmandaki düğüm sayısı	18	11
Çıktı nöron sayısı	1	1
Başlangıç değeri ve Bias değeri	Rastsal	
Aktivasyon Fonksiyonları	Tansig-fonksiyon (gizli katman)	Tansig-fonksiyon (çıktı katmanı)
Öğrenme fonksiyonu	Momentumun açışal azalması metodu	
Hedef	0,000001	

Çizelge 4.3’de YSA sınıflandırıcısının matlab ortamındaki parametreleri gözükmektedir. Katman sayısı 3’tür. Girdi nöron sayısı 44 öznelikten dolayı 44 olmuştur. Gizli katmandaki düğüm sayısı 14 ve iterasyon sayısı 43 olduğu zaman en iyi sonucu vermiştir. Çıktı katmanı 1 nörondan oluşmuştur. Başlangıç ve bias değerleri rastsaldır. Aktivasyon fonksiyonu tanjant sigmoit fonksiyondur. Performans kriterlerimizdeki artış ve azalışlar veri setimizin farklı algoritmalar ile değişikliğe maruz kaldığını göstermektedir. Belirli bir düzenden yeni düzenlerine geçtikleri için ayrı ayrı değerlendirilmelidir. Yeni düzenlerine göre hesaplandığında kendi içlerinde yapılan yeni hesaplamaları içermektedirler. Haliyle eski verilerin yerine yeni veriler geçmiştir. Veri seti değiştiği içinde performans kriterleri yeni değerler almıştır.

Çizelge 4.4 DVM ve SNBM+DVM (Statlog) Sonuç Değerleri

Performans Kriteri	DVM	SNBM + DVM
Doğruluk Oranı (%)	84.81	86.66
Hassaslık Oranı (%)	80.61	87.05
Özgüllük Oranı (%)	88.37	86.64
Pozitif Doğruluk Değer Oranı (%)	85.40	85.14
Negatif Doğruluk Değer Oranı (%)	84.29	89.46

Çizelge 4.4’de DVM ve SNBM + DVM sonuçları gösterilmiştir. Doğruluk oranı %84.81’den %86.66’ya çıkmıştır. Hassaslık oranı %80.61’den %87.05’e yükselmiştir. Özgüllük oranı %88.37’den %86.64’a düşmüştür. Pozitif tahmini değer oranı %85.40’dan %85.14’e düşmüştür. Negatif tahmini değer oranı %84.29’dan %89.46’a yükselmiştir. Bu sonuçlara göre SNBM’nin olumlu katkısı olduğu gözükmektedir.

Çizelge 4.5 YSA ve SNBM+YSA (Statlog) Sonuç Değerleri

Performans Kriteri	YSA	SNBM + YSA
Doğruluk Oranı (%)	85.18	87.78
Hassaslık Oranı (%)	88.31	86.39
Özgüllük Oranı (%)	83.24	90.59
Pozitif Doğruluk Değer Oranı (%)	85.29	86.76
Negatif Doğruluk Değer Oranı (%)	84.50	88.36

Çizelge 4.5’de YSA ve SNBM + YSA sonuçları gösterilmiştir. Doğruluk oranı %85.18’den %87.78’e çıkmıştır. Hassaslık oranı %88.31’den %86.39’a düşmüştür. Özgüllük oranı %83.24’den %90.59’a çıkmıştır. Pozitif tahmini değer oranı %85.29’dan %86.76’a çıkmıştır. Negatif tahmini değer oranı %84.50’den %88.36’a çıkmıştır. Bu sonuçlara göre SNBM’nin olumlu katkısı olduğu gözükmemektedir.

Çizelge 4.6 TBA + DVM (Statlog) Sonuç Değerleri

Performans Kriteri	DVM	TBA + DVM
Doğruluk Oranı (%)	84.81	85.18
Hassaslık Oranı (%)	80.61	87.71
Özgüllük Oranı (%)	88.37	82.38
Pozitif Doğruluk Değer Oranı (%)	85.40	85.62
Negatif Doğruluk Değer Oranı (%)	84.29	85.26

Çizelge 4.6’da DVM ve TBA + DVM sonuçları gösterilmiştir. Doğruluk oranı %84.81’den %85.18’e çıkmıştır. Hassaslık oranı %80.61’den %87.71’e yükselmiştir. Özgüllük oranı %88.37’den %82.38’e düşmüştür. Pozitif tahmini değer oranı %85.40’dan %85.62’ye çıkmıştır. Negatif tahmini değer oranı %84.29’dan %85.26’a çıkmıştır. Bu sonuçlara göre TBA’nın olumlu katkısı olduğu gözükmemektedir.

Çizelge 4.7 TBA + YSA (Statlog) Sonuç Değerleri

Performans Kriteri	YSA	TBA + YSA
Doğruluk Oranı (%)	85.18	86.66
Hassaslık Oranı (%)	88.31	91.15
Özgüllük Oranı (%)	83.24	81.73
Pozitif Doğruluk Değer Oranı (%)	85.29	86.54
Negatif Doğruluk Değer Oranı (%)	84.50	87.87

Çizelge 4.7’de YSA ve TBA + YSA sonuçları gösterilmiştir. Doğruluk oranı %85.18’den %86.66’e çıkmıştır. Hassaslık oranı %88.31’den %91.15’e yükselmiştir. Özgüllük oranı %83.24’ten %81.73’e düşmüştür. Pozitif tahmini değer oranı %85.29’dan %86.54’e çıkmıştır. Negatif tahmini değer oranı %84.50’den %87.87’ye çıkmıştır. Bu sonuçlara göre TBA’nın olumlu katkısı olduğu gözükmektedir.

Çizelge 4.8 Hibrit Sistem + DVM (Statlog) Sonuç Değerleri

Performans Kriteri	YSA	Hibrit Sistem + DVM
Doğruluk Oranı (%)	85.18	87.77
Hassaslık Oranı (%)	88.31	89.22
Özgüllük Oranı (%)	83.24	85.22
Pozitif Doğruluk Değer Oranı (%)	85.29	86.83
Negatif Doğruluk Değer Oranı (%)	84.50	89.34

Çizelge 4.8’de YSA ve Hibrit Sistem + DVM doğruluk oranı gösterilmiştir. Doğruluk oranı %85.18’den %87.77’e çıkmıştır. Hassaslık oranı %88.31’den %89.22’e çıkmıştır. Özgüllük oranı %83.24’den %85.22’e düşmüştür. Pozitif tahmini değer oranı %85.29’dan %86.83’e düşmüştür. Negatif tahmini değer oranı %84.50’dan %89.34’a çıkmıştır. Hibrit Sistem sınıflandırıcı sonucumuza katkı sağlamıştır.

Çizelge 4.9 Hibrit Sistem + YSA (Statlog) Sonuç Değerleri

Performans Kriteri	YSA	Hibrit Sistem + YSA
Doğruluk Oranı (%)	85.18	88.89
Hassaslık Oranı (%)	88.31	91.21
Özgüllük Oranı (%)	83.24	85.79
Pozitif Doğruluk Değer Oranı (%)	85.29	88.77
Negatif Doğruluk Değer Oranı (%)	84.50	89.07

Çizelge 4.9’de YSA ve Hibrit Sistem + YSA sonuçları gösterilmiştir. Doğruluk oranı %85.18’den %88.89’a çıkmıştır. Hassaslık oranı %88.31’den %91.21’e çıkmıştır. Özgüllük oranı %83.24’den %85.79’a çıkmıştır. Pozitif tahmini değer oranı %85.29’dan %88.77’e çıkmıştır. Negatif tahmini değer oranı %84.50’den %89.07’a çıkmıştır. 17’nci iterasyonda bu sonuca erişilmiştir. Hibrit Sistem sınıflandırıcı sonucumuza katkı sağlamıştır.

TFEBT veri seti ile yapılan çalışmalar aşağıda sunulmuştur. Kısaca TFEBT veri seti 44 öznitelik barındırmaktadır. 267 hasta kaydı içermektedir. Öznitelikler sayısal değer alır ve sonuç 0 veya 1 ikili değere sahip olmaktadır.

Çizelge 4.10 DVM ve Normalizasyon + DVM (TFEBT) Sonuç Değerleri

Performans Kriteri	DVM	Norm. + DVM
Doğruluk Oranı (%)	75.27	75.27
Hassaslık Oranı (%)	71.00	76.45
Özgüllük Oranı (%)	76.45	71.00
Pozitif Doğruluk Değer Oranı (%)	45.68	45.68
Negatif Doğruluk Değer Oranı (%)	91.05	91.05

Çizelge 4.10'da DVM ve Normalizasyon + DVM sonuçları gösterilmiştir. DVM ile bulunan doğru oranı %75.27 olmuştur. Normalizasyon ile bulunan doğruluk oranı da %75.27 olarak bulunmuştur. Hassaslık oranı %71.00'dan %76.45'e yükselmiştir. Özgüllük oranı %76.45'den %71.00'e düşmüştür. Pozitif tahmini değer oranı %45.68 olarak kalmıştır. Negatif tahmini değer oranı da %91.05 olarak kalmıştır. Normalizasyon işlemi pozitif doğruluk değer oranını ve negatif doğruluk değer oranını değiştirmesede diğer değerleri değiştirmiştir.

Çizelge 4.11 DVM ve TBA + Norm.+ DVM (TFEBT) Sonuç Değerleri

Performans Kriteri	DVM	TBA + Norm. + DVM
Doğruluk Oranı (%)	75.27	78.30
Hassaslık Oranı (%)	71.00	80.17
Özgüllük Oranı (%)	76.45	72.33
Pozitif Doğruluk Değer Oranı (%)	45.68	45.27
Negatif Doğruluk Değer Oranı (%)	91.05	92.31

Çizelge 4.11'de DVM ve TBA + Normalizasyon sonuçları gösterilmiştir. DVM ile bulunan doğru oranı %75.27 olmuştur. TBA ve Normalizasyon ile bulunan doğruluk oranı %78.30 olarak bulunmuştur. Hassaslık oranı %71.00'den %80.17'ye çıkmıştır. Özgüllük oranı %76.45'den %72.33'e düşmüştür. Pozitif tahmini değer oranı %45.68'den %45.27'ye düşmüştür. Negatif tahmini değer oranı %91.05'den %92.31'e yükselmiştir. TBA ve Normalizasyon algoritmalarının sınıflandırıcı sonucumuza katkı sağlamış olduğu görülmüştür.

Çizelge 4.12 DVM ve TBA + SNBM + Norm.+ DVM (TFEBT) Sonuç Değerleri

Performans Kriteri	DVM	TBA + SNBM + Norm. + DVM
Doğruluk Oranı (%)	75.27	89.17
Hassaslık Oranı (%)	71.00	90.58
Özgüllük Oranı (%)	76.45	83.66
Pozitif Doğruluk Değer Oranı (%)	45.68	72.53
Negatif Doğruluk Değer Oranı (%)	91.05	95.70

Çizelge 4.12’de DVM ve TBA + SNBM + Normalizasyon sonuçları gösterilmiştir. DVM ile bulunan doğruluk oranı %75.27 olmuştur. TBA + SNBM + Normalizasyon ile bulunan doğruluk oranı %89.17 olarak bulunmuştur. Hassaslık oranı %71.00’den %90.58’e yükselmiştir. Özgüllük oranı %76.45’den %83.66’a yükselmiştir. Pozitif tahmini değer oranı %45.68’den %72.53’e yükselmiştir. Negatif tahmini değer oranı %91.05’den %95.70’e yükselmiştir. TBA, SNBM ve Normalizasyon algoritmalarının sınıflandırıcı sonucumuza katkı sağlamış olduğu görülmüştür.

Çizelge 4.13 DVM ve SNBM + TBA + Norm.+ DVM (TFEBT) Sonuç Değerleri

Performans Kriteri	DVM	TBA + SNBM + Norm. + DVM
Doğruluk Oranı (%)	75.27	83.98
Hassaslık Oranı (%)	71.00	84.91
Özgüllük Oranı (%)	76.45	80.66
Pozitif Doğruluk Değer Oranı (%)	45.68	61.88
Negatif Doğruluk Değer Oranı (%)	91.05	94.05

Çizelge 4.13’de DVM ve SNBM + TBA + Normalizasyon sonuçları gösterilmiştir. DVM ile bulunan doğruluk oranı %75.27 olmuştur. SNBM + TBA + Normalizasyon ile bulunan doğruluk oranı %83.98 olarak bulunmuştur. Hassaslık oranı %71.00’den %84.91’e yükselmiştir. Özgüllük oranı %76.45’den %80.66’ya çıkmıştır. Pozitif tahmini değer oranı %45.68’den %61.88’e çıkmıştır. Negatif tahmini değer oranı %91.05’den %94.05’e yükselmiştir. SNBM + TBA ve Normalizasyon algoritmalarının sınıflandırıcı sonucumuza katkı sağlamış olduğu görülmüştür.

Çizelge 4.14 YSA ve TBA + SNBM + Norm.+ YSA (TFEBT) Sonuç Değerleri

Performans Kriteri	YSA	TBA + SNBM + Norm. + YSA
Doğruluk Oranı (%)	79.43	89.91
Hassaslık Oranı (%)	87.27	94.30
Özgüllük Oranı (%)	50.33	72.66
Pozitif Doğruluk Değer Oranı (%)	87.43	93.24
Negatif Doğruluk Değer Oranı (%)	48.03	48.05

Çizelge 4.14'te YSA ve TBA + SNBM + Normalizasyon sonuçları gösterilmiştir. YSA ile bulunan doğruluk oranı %79.43 olmuştur. TBA + SNBM + Normalizasyon ile bulunan doğruluk oranı %89.91 olarak bulunmuştur. Hassaslık oranı %87.27'den %94.30'e yükselmiştir. Özgüllük oranı %50.33'den %72.66'ya çıkmıştır. Pozitif tahmini değer oranı %87.43'den %93.24'e çıkmıştır. Negatif tahmini değer oranı %48.03'den %48.05'e yükselmiştir. TBA + SNBM ve Normalizasyon algoritmalarının sınıflandırıcı sonucumuza katkı sağlamış olduğu görülmüştür.

Çizelge 4.15'te ve çizelge 4.16'da önerilen sistem ile literatürdeki diğer sonuçlar kıyaslanmıştır. Statlog veriseti için (çizelge 4.15); Duch, Sahan ve arkadaşları, Özsen ve Güneş, Kahramanlı ve Allahverdi, Tian ve arkadaşları, Polat ve Güneş, Subbulakshmi, Karabulut, Ahmet ve arkadaşları'nın yapmış oldukları çalışma sonuçları önerilen hibrit sistem ile kıyaslanmıştır. Önerilen sistem diğer çoğu çalışma sonuçlarından daha iyi sonuç vermiştir. Duch, Sahan, Ozsen, Kahramanlı, Polat, Subbulakshmi çalışmacılarından daha iyi sonuç alınmıştır. Karabulut'tan düşük bulunmasının nedenini Levenberg-Marquardt algoritmasının hatayı daha iyi minimize etmesine ve ikinci dereceden türev almasına bağlanmıştır. TFEBT veri seti için de (çizelge 4.16); Andy, Michael Fernando, Kristanto Halim, Gadiyanto Sanjaya, Roya Asadi, Norwati Mustapha, Nasir Sulaiman, Nematollaah Shiri, Jose George ve Rajeev Kumaraswamy çalışmacılarının bulmuş oldukları sonuçlar önerilen sistem sonuçları ile kıyaslanmıştır. Önerilen sistemin çoğu algoritmadan daha iyi sonuç verdiği görülmüştür.

Araştırmacıların UCI statlog veri seti ile ve TFEBT veri seti ile yaptığı çalışmalar aşağıda listelenmiştir. Çalışmalar sonucunda buldukları doğruluk oranları Hibrit sistemde kullanılan aynı veri seti doğruluk oranları ile kıyaslanmıştır. Sonuçlar analiz edildiğinde Hibrit sistem ile bulduğumuz doğruluk oranları araştırmacıların

bulduğu doğruluk oranları ile kıyaslandığında Hibrit sistemin sonuç olarak iyi performans ve doğruluk oranları elde ettiği görülmüştür.

Çizelge 4.15 Statlog Veri Seti İçin Araştırmacıların Detaylı Mukayese Değerleri

Araştırmacı	Algoritma	Doğruluk Oranı (%)
Duch(Duch ve ark., 2001).	k-NN, k=28, 7 Özellik (10-katmanlı CV)	84.6–85.6
	k-NN, k=28, Manhattan (10-katmanlı CV)	82.2–83.4
	FSM, 27 fuzzy-rules	82
	SSV, 3 rules	8.2–83.4
Sahan et al(Şahan ve ark., 2005).	AWATYBS (10-katmanlı CV)	82.59
Ozsen ve Gunes(Ozsen ve ark., 2009).	ATYBS algoritması	83.95
	Hybrid benzerlik Ölçüm (10-katmanlı CV)	
Kahramanli ve Allahverdi(Kahramanli ve Allahverdi, 2008).	Hibrit-sistem kullanılarak YSA ve FSA (10-katmanlı CV)	86.8
	CC-EBFNN	82.45
Tian J. Et(Tian ve ark., 2009). Polat ve Gunes(Polat ve Güneş, 2009).	LS-DVM Sınıflandırıcı Özellik çıkarmadan (Eğitim: 50 % - test: 50 %)	80
	LM YSA's özellik çıkarmadan	71.11
	F-score özellik çıkarma ve LS-DVM Sınıflandırıcı	77.78
	F-score özellik çıkarımı ve YSA's	77.61
	Doğrusal Kernel F-score özellik çıkarımı	
	LS-DVM Sınıflandırıcı (eğitim: 50% - test: 50%)	80
	Doğrusal Kernel ile F-score özellik seçimi	83.7
	YSA's (eğitim: 50% - test: 50%)	
	Kernel RBF ve F-score özellik çıkarımı	80
	LS-DVM	
	Doğrusal Kernel F-score özellik çıkarımı	76.3
	YSA (eğitim: 50% - test: 50%)	
Subbulakshmi(Subbulakshmi ve ark., 2012).	ELM	87.5
Karabulut E.M(Karabulut ve İbrikçi, 2012).	Sıralı Dönme Ormanı, Levenberg-Marquardt (10 - katmanlı CV)	91.2

Önerilen Yöntem (Statlog)	DVM	84.81
Önerilen Yöntem (Statlog)	YSA	85.18
Önerilen Yöntem (Statlog)	SNBM + DVM	86.66
Önerilen Yöntem (Statlog)	Hibrit Sistem + DVM	87.77
Önerilen Yöntem (Statlog)	SNBM + YSA	87.78
Önerilen Yöntem (Statlog)	Hibrit Sistem + YSA	88.89

Çizelge 4.16 TFEBT Veri Seti İçin Araştırmacıların Detaylı Mukayese Değerleri

Araştırmacı	Algoritma	Doğruluk Oranı (%)
Andy, Michael Fernando,	GA-KNN	85.7
Kristanto Halim, Gadiyanto Sanjaya (Andy ve ark., 2013)	GA-SVM	86.5
	GA-DT	84.2
	GA-LDA	85
Roya Asadi, Norwati Mustapha	New SMFFNN by using WLA	94
Nasir Sulaiman, Nematollaah Shiri (Asadi ve ark., 2009)	SBPN	79
	BPN by using PCA	75.1
	K-NN (K=1)	72.1
	CLIP3	77
	CLIP4	77
	RBF	91.4
Jose George , Rajeev Kumaraswamy (George ve Kumaraswamy, 2008)	Shannon	74.9
	Morlet	92.5
	Sinc	91.4
	Sinc-RBF	93
Önerilen Yöntem (TFEBT)	YSA	79.43
Önerilen Yöntem (TFEBT)	DVM	75.3
Önerilen Yöntem (TFEBT)	TBA + SNBM + Norm. + DVM	89.17
Önerilen Yöntem (TFEBT)	TBA + SNBM + Norm. + YSA	89.91

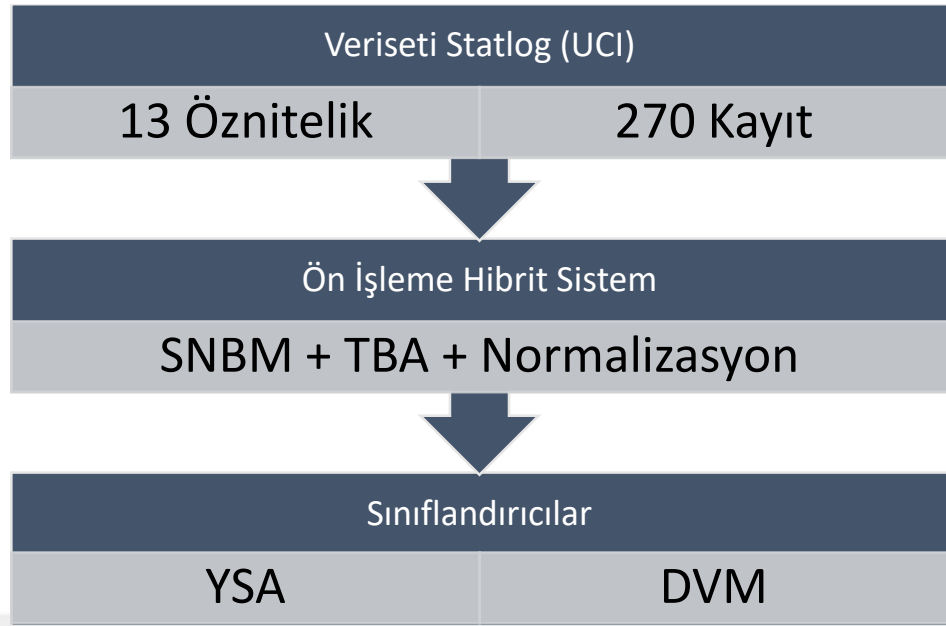
## 5. SONUÇLAR VE ÖNERİLER

Tez çalışmasında kalp krizi riskini tespit etmek için UCI Makine Öğrenmesi Deposu'ndan alınan kalp hastalığı veri kümeleri incelenmiştir. Bu veri kümeleri 4 ayrı veri tabanından oluşmaktadır. Bunlar; Statlog, Cleveland, Hungary, Switzerland'dır. Tüm veri setleri aynı değişkenleri içermektedir ve aynı formatta oluşturulmuştur. Statlog veri kümesi 76 değişken içermektedir; ancak, bu veri kümesi ile yapılan çalışmalar, bu değişkenlerden, 13 tanesinin en önemli verileri barındırdığını göstermiştir. Statlog veri kümesi 270 hasta kaydından oluşmaktadır. 13 öznitelikten oluşan veri seti çizelge 5.1'de ki alanlardan oluşmuştur. Çalışmada kullanılan diğer veri seti olan TFEBT veri seti ise 267 kayıttan oluşmaktadır ve 44 öznitelikten oluşturulmuştur. Her öznitelik sayısal değer almaktadır.

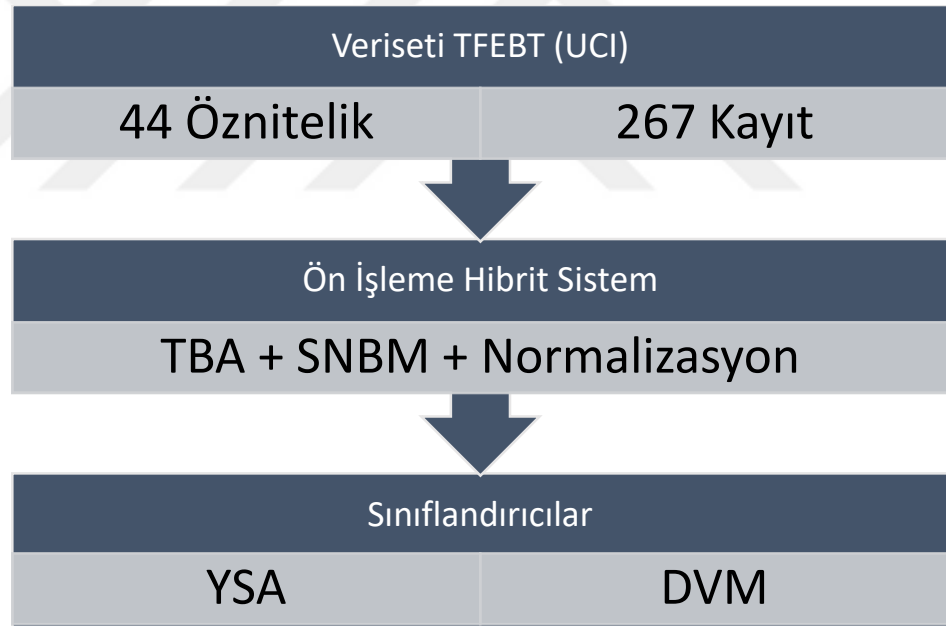
Çizelge 5.1 UCI Kalp Hastalığı Statlog  
Veri Seti Öznitelikleri

1	Yaş	Sayısal
2	Cinsiyet	Sayısal
3	Göğüs ağrısı	Sayısal
4	Kan basıncı	Sayısal
5	Kolesterol	Sayısal
6	Şeker	Sayısal
7	EKG	Sayısal
8	Kalp atışı sayısı	Sayısal
9	Anjin varlığı	Sayısal
10	Kalp krizi varlığı	Sayısal
11	ST eğim yönü	Sayısal
12	Sintigrafi sonucu	Sayısal
13	Tıkalı damar durumu	Sayısal
14	Anjiyografi sonucu	Sayısal

Şekil 5.2 ve Şekil 5.3'de görüldüğü gibi önerilen sistemin aşamaları verilmiştir. UCI'den alınan verisetleri Hibrit Sistem ardından sınıflandırıcılara iletilmektedir. İki farklı sınıflandırıcı ile verisetine en uygun sınıflandırıcı anlaşılmış olmaktadır.



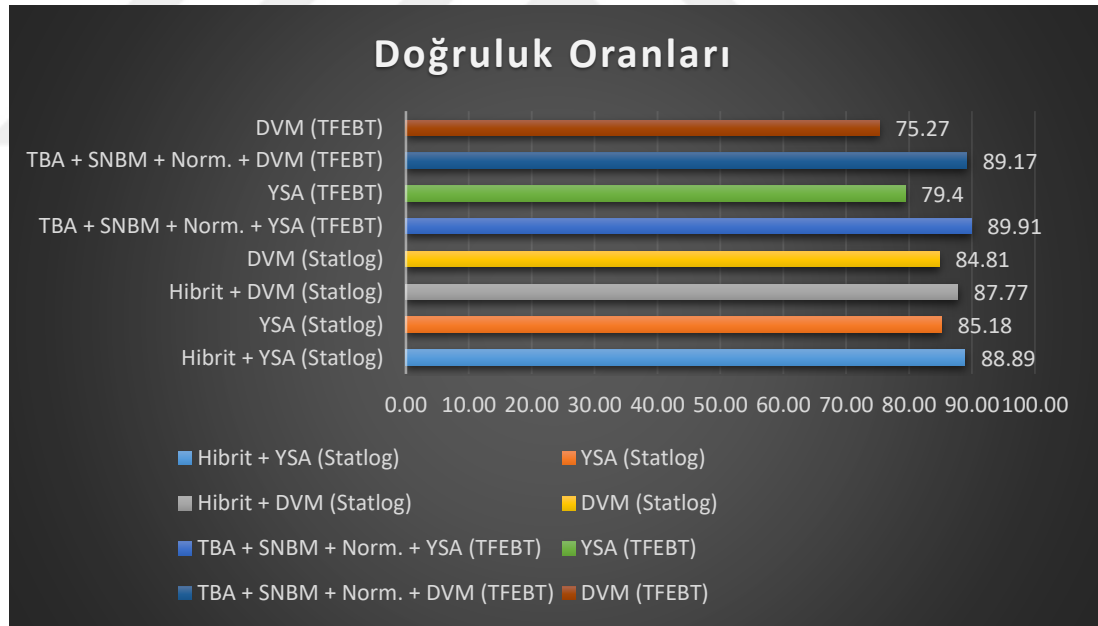
Şekil 5.2 Önerilen Sistem'in İş Akışı Statlog Veri Seti ile



Şekil 5.3 Önerilen Sistem'in İş Akışı TFEBT Veri Seti ile

Yukarıda belirtildiği gibi SNBM, TBA ve Normalizasyon algoritmaları veri setlerinde ön işleme yapmıştır. Ayrıklaştırma işlemi sonucunda yeni değerler oluşmuştur. Bu sayede büyük boyutlu veriler kendi içerisinde gruplanmış, yeniden etiketlenmiş ve birbirleriyle olan ilişkilerine göre yeni duruma sahip olmuşlardır. Belli aralıklara bölünen ayırık değerler bölünen sınıf sayısı kadar etiketlenmiş işlem hızı da

arttırılmıştır. UCI'den alınmış veri setlerine SNBM, TBA ve Normalizasyon algoritmaları ile, sınıflandırıcıya girmeden önce ön işleme yaptırılmıştır. Çalışmada Hibrit sistemin çeşitli versiyonları da denenmiştir. Oluşturulan yeni veri seti ayrı ayrı sınıflandırıcılara sokularak hangi sınıflandırıcıda daha iyi sonuç vereceği yorumlanmak istenmiştir. Sonuçların çizelge 4.14'te ve çizelge 4.15'te belirtildiği gibi umut verici olduğu görülmüştür. Statlog veri seti ile Hibrit ve YSA uygulanmış sistemin doğruluk oranı %88.89 olarak bulunmuştur. Hibrit ve DVM uygulanmış sistemin doğruluk oranı ise %87.77 olarak bulunmuştur. TFEBT veri seti ile yapılan çalışma sonucunda ise TBA + SNBM + Normalizasyon ve DVM algoritmasının doğruluk oranı %89.17 olarak bulunmuştur. TBA + SNBM + Normalizasyon ve YSA ile bulunan doğruluk oranı ise %89.91 olarak bulunmuştur. DVM algoritmasının doğruluk oranı ise %75.27 olmuştur. YSA sınıflandırıcısı ile bulunan doğruluk oranı ise 79.4 olarak bulunmuştur. TBA + SNBM + Normalizasyon ve DVM ile bulunan doğruluk oranlarının ve TBA + SNBM + Normalizasyon ve YSA ile bulunan doğruluk oranlarının doktor tahminlerinden oldukça iyi olduğu görülmüştür.



Şekil 5.4 Çalışma Sonuçları

Şekil 5.4'te DVM ve YSA sonuçlarına Hibrit Sistemin sınıflandırma başarısını arttırdığı görülmüştür. TFEBT veri seti için TBA + SNBM + Normalizasyon + YSA'lı doğruluk sonucu diğer doğruluk sonuçlarını geride bırakmıştır. Ayrıca şekil 5.4'te üzerinde DVM ve TBA + SNBM + Normalizasyon + DVM sonuçları da gösterilmiştir. Karabulut'un yaptığı çalışmadaki doğruluk oranının önerilen sistemden yüzde olarak daha iyi oluşu; Sıralı Dönme Ormanı ve Levenberg-Marquardt algoritmasının

ikinci dereceden türev almış olmasına ve doğruluk oranının bu metotla daha kesin olarak hesaplanmış olabileceği düşünülmüştür. Önerilen sistemin diğer sistemlerden olan üstünlükleri ise önışleme yönteminin sınıflayıcıyı daha etkili bir eğitim yaptırıp tahmin gücünü arttırmasına, yapay zekâ tekniğinde seçilen GYA algoritmalarına, gizli katman sayısına, gizli katmandaki yapay sinir sayısına, eğitim algoritmalarına ve Hibrit Sisteme bağlanmıştır.

Veri madenciliği algoritmaları KAH'nın belirttiği risk faktörlerinin belirlenmesinde önemli katkı sağlamıştır. Bu çalışmada KAH hastalığının riskinin belirlenmesinde GYA yöntemini içeren YSA sınıflandırıcı ve 2 sonuçlu DVM sınıflandırma yöntemi geliştirilen bir Hibrit Sistem ile kullanılmıştır. Sınıflandırma algoritmaları; doğruluk, hassasiyet, gibi performans metrikleri kullanılarak karşılaştırılmıştır.

Veri madenciliği sınıflandırma algoritmaları doğruluk açısından incelendiğinde, Statlog veri seti için en iyi sonuç %88.89 doğruluk oranı ile Hibrit Sistem ile entegre edilen GYA'lı YSA sınıflandırma yönteminden elde edilmiştir.

Veri madenciliği sınıflandırma algoritmaları doğruluk açısından incelendiğinde, TFEBT veri seti için en iyi sonuç %89.91 doğruluk oranı ile TBA + SNBM + Norm. ile entegre edilen YSA sınıflandırma yönteminden elde edilmiştir.

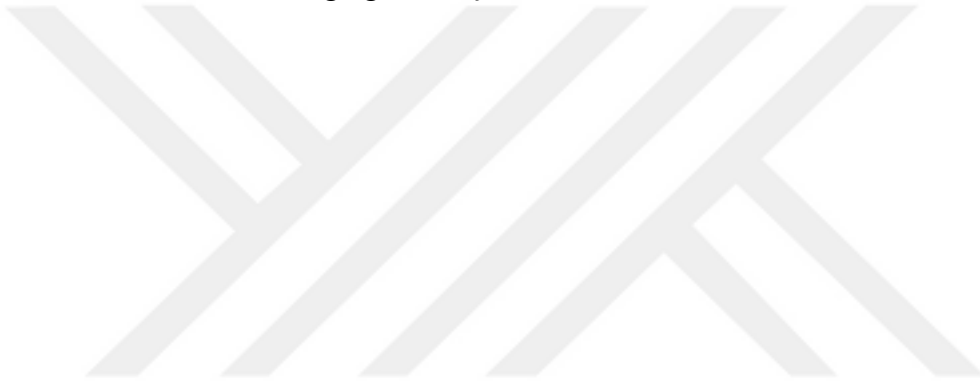
Sonuçlar olumlu olduğu için uzman hekimler ve araştırmacılara teşhislerinde ve araştırma sonuçlarında yorumlama ve kıyaslama imkânı sunabilmiştir. Bu nedenlerle uygulamalarda hangi metodun tercih edileceği uygulamaya bağlı olarak belirlenmelidir. Bu çalışmanın sonuçlarının KAH, Felç, Damar Tıkanıklığı vb. hastalık şüphesi ile kliniğe başvuran hastaların tanı ve tedavi sürecinde kardiyoloji alanında çalışan uzmanların klinik kararlarına rehberlik edeceği düşünülmektedir.

Ayrıca, geliştirilen model ile tıbbi hatalar, gereksiz uygulamalar, teşhis ve tedavi maliyetleri azaltılabileceği için, bu sayede hasta güvenliği ve yaşam kalitesi artırımı hedeflenmiş olacaktır.

Gelecek çalışmalarda optimizasyon algoritmaları kullanılarak veri kümesindeki değişkenler daha detaylı incelenerek ön işleme ve sınıflandırma algoritmaları uygulanabilir. Ayrıca modellerin uygulanmasında Doğru Pozitif ve Yanlış Negatif oranlarının sonuçları da dikkate alınarak hastalık tahmini sonuçları ile ilgili yeni değerlendirmeler yapılabilir.

Bu tezde kullanılan SNBM, YSA, TBA, DVM, Normalizasyon vb. algoritmalar deęiştirilerek veya yeni geliřtirilecek algoritmalar (algoritmanın yeni versiyonu gibi) sisteme dahil edilerek gelecekte bu risk tespitindeki doęruluk oranlarını daha yukarıya çıkarmak amaçlanacaktır. Sürekli güncellenen veritabanı üzerinde sistem belli periyotlarda çalışarak yeni modeller üretebilir. Hastalara kendilerini takip edebilecekleri bir arayüz geliřtirilebilir. Saęlık Bakanlıęı'na baęlı birimlerde bir arayüz tanımlanabilir.

Kalp hastalık teřhisi için oluřturulan klinik karar destek sistemlerinin çok verimli olduęu, çeřitli problemlere cevap verebildięi, sonuç doęruluk deęerlerinin hibrit yöntemlerle daha da geliřtirilebileceęi anlařılmış ve çalışmaların ilerisi için umut verici olduęu görülmüřtür.



## KAYNAKLAR

- Abushariah, M. A., Alqudah, A. A., Adwan, O. Y. ve Yousef, R. M., 2014, Automatic heart disease diagnosis system based on artificial neural network (ANN) and adaptive neuro-fuzzy inference systems (ANFIS) approaches, *Journal of software engineering and applications*, 7 (12), 1055.
- Adeli, A. ve Neshat, M., 2010, A fuzzy expert system for heart disease diagnosis, *Proceedings of International Multi Conference of Engineers and Computer Scientists, Hong Kong*.
- Adıyaman, F., 2007, Talep tahmininde yapay sinir ağlarının kullanılması, *Fen Bilimleri Enstitüsü*.
- Ahmed, A. ve Hannan, S. A., 2012, Data Mining Techniques to Find Out Heart Diseases: An Overview, *International Journal of Innovative Technology and Exploring Engineering*, 1 (4), 18-23.
- AKIN, Çağrı. Emre., 2017, Çapraz Doğrulama, <http://cagrimreakin.com/veri-bilimi/k-fold-cross-validation-1.html>.
- Al-Milli, N., 2013, Backpropagation neural network for prediction of heart disease, *Journal of theoretical and applied information Technology*, 56 (1), 131-135.
- Andy, M. F., Halim, K. ve Sanjaya, G., 2013, Optimization features using GA-SVM approach, *Breast Cancer*, 569 (569), 30.
- Asadi, R., Mustapha, N., Sulaiman, N. ve Shiri, N., 2009, New supervised multi layer feed forward neural network model to accelerate classification with high accuracy, *European Journal of Scientific Research*, 33 (1), 163-178.
- Banu, G. P., M & H Bousal, J & Mca, Jamala, 2015, Predicting Heart Attack using Fuzzy C Means Clustering Algorithm., *International Journal of latest Trends in Engineering and Technology*. , 5.
- Bhatia, S., Prakash, P. ve Pillai, G., 2008, SVM based decision support system for heart disease classification with integer-coded genetic algorithm to select critical features, *Proceedings of the world congress on engineering and computer science*, 34-38.
- Chowdhury, D. R., Chatterjee, M. ve Samanta, R., 2011, An artificial neural network model for neonatal disease diagnosis, *International Journal of Artificial Intelligence and Expert Systems (IJAE)*, 2 (3), 96-106.
- Diñçer, B. T. ve Karaođlan, B., 2004, Sentence boundary detection in Turkish, *International Conference on Advances in Information Systems*, 255-262.
- Dua Dheeru, G. C., 2017, UCI Machine Learning Repository.
- Duch, W., Adamczak, R. ve Grabczewski, K., 2001, A new methodology of extraction, optimization and application of crisp and fuzzy logical rules, *IEEE Transactions on Neural Networks*, 12 (2), 277-306.
- George, J. ve Kumaraswamy, R., 2008, A Hybrid Wavelet Kernel Construction for Support Vector Machine Classification, *DMIN*, 96-101.
- Gudadhe, M., Wankhade, K. ve Dongre, S., 2010, Decision support system for heart disease based on support vector machine and artificial neural network, *Computer and Communication Technology (ICCCT), 2010 International Conference on*, 741-745.
- Guru, N., Dahiya, A. ve Rajpal, N., 2007, Decision support system for heart disease diagnosis using neural network, *Delhi Business Review*, 8 (1), 99-101.
- Han, J., Pei, J. ve Kamber, M., 2011, Data mining: concepts and techniques, Elsevier, p.
- Hasan, T. T., Jasim, M. H. ve Hashim, I. A., 2017, Heart Disease Diagnosis System based on Multi-Layer Perceptron neural network and Support Vector Machine.
- Juneja, U. ve Dhingra, D., 2014, Multi Parametric Approach Using Fuzzification on Heart Disease Analysis.
- Kahramanli, H. ve Allahverdi, N., 2008, Design of a hybrid system for the diabetes and heart diseases, *Expert systems with applications*, 35 (1-2), 82-89.
- Karabulut, E. M. ve İbrikçi, T., 2012, Effective diagnosis of coronary artery disease using the rotation forest ensemble method, *Journal of medical systems*, 36 (5), 3011-3018.
- Khemphila, A. ve Boonjing, V., 2011, Heart disease classification using neural network and feature selection, *Systems Engineering (ICSEng), 2011 21st International Conference on*, 406-409.
- Koçođlu, F., 2012, Veri Madenciliginde Veri Ayırklařtırma Yöntemlerinin Karřılařtırılması ve Bir Uygulama. , *İstanbul Üniversitesi. Fen Bilimleri Enstitüsü*.
- Krzysztof J. Cios, L. A. K., 2001, {UCI} Machine Learning Repository.
- Kumari, M. ve Godara, S., 2011, Comparative study of data mining classification methods in cardiovascular disease prediction 1.
- Li, S., Li, H., Li, M., Shyr, Y., Xie, L. ve Li, Y., 2009, Improved prediction of lysine acetylation by support vector machines, *Protein and peptide letters*, 16 (8), 977-983.
- Liu, X., Wang, X., Su, Q., Zhang, M., Zhu, Y., Wang, Q. ve Wang, Q., 2017, A hybrid classification system for heart disease diagnosis based on the rfrs method, *Computational and mathematical methods in medicine*, 2017.
- Ozsen, S., Gunes, S., Kara, S. ve Latifoglu, F., 2009, Use of kernel functions in artificial immune systems for the nonlinear classification problems, *IEEE Transactions on Information Technology in Biomedicine*, 13 (4), 621-628.

- Palaniappan, S. ve Awang, R., 2008, Intelligent heart disease prediction system using data mining techniques, *Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on*, 108-115.
- Patil, S. B. ve Kumaraswamy, Y., 2009, Intelligent and effective heart attack prediction system using data mining and artificial neural network, *European Journal of Scientific Research*, 31 (4), 642-656.
- Pattekari, S. A. ve Parveen, A., 2012, Prediction system for heart disease using Naïve Bayes, *International Journal of Advanced Computer and Mathematical Sciences*, 3 (3), 290-294.
- Polat, K. ve Güneş, S., 2009, A new feature selection method on classification of medical datasets: Kernel F-score feature selection, *Expert Systems with Applications*, 36 (7), 10367-10373.
- Şahan, S., Polat, K., Kodaz, H. ve Güneş, S., 2005, The medical applications of attribute weighted artificial immune system (AWAIS): diagnosis of heart and diabetes diseases, *International Conference on Artificial Immune Systems*, 456-468.
- Sevgi AYHAN, Ş. E., 2014, Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi *dergipark.org.tr*.
- Shahi, M. ve Gurm, E. R. K., 2017, Heart Disease Prediction System Using Data Mining Techniques-A Review, *Heart Disease*, 3 (4).
- Sowmiya, C. ve Sumitra, P., 2017, Analytical study of heart disease diagnosis using classification techniques, *Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2017 IEEE International Conference*, 1-5.
- Subbulakshmi, C., Deepa, S. ve Malathi, N., 2012, Extreme learning machine for two category data classification, *Advanced Communication Control and Computing Technologies (ICACCCT), 2012 IEEE International Conference on*, 458-461.
- Tian, J., Li, M. ve Chen, F., 2009, A hybrid classification algorithm based on coevolutionary EBFNN and domain covering method, *Neural Computing and Applications*, 18 (3), 293-308.
- Uzunoğlu, Mehmet, 2004, Her Yönüyle Matlab.
- Vanisree, K. ve Singaraju, J., 2011, Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks, *International Journal of Computer Applications (0975–8887) Volume*.
- Waghmare, A., Verma, N., Gaur, A. ve Kannan, R. J., Department of Computing Science & Engineering, VIT Chennai.
- Weng, S. F., Reys, J., Kai, J., Garibaldi, J. M. ve Qureshi, N., 2017, Can machine-learning improve cardiovascular risk prediction using routine clinical data?, *PloS one*, 12 (4), e0174944.
- WHO, 2016, The top 10 causes of death, <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>.
- Wikizero, Análisis de componentes principales, [https://es.0wikipedia.org/wiki/Analysis\\_de\\_componentes\\_principales](https://es.0wikipedia.org/wiki/Analysis_de_componentes_principales).
- Yalçın Özkan, Çiğdem Selçukcan Erol, 2017, Biyoenformatik DNA Mikrodizi Veri Madenciliği (978-605-4220-89-2).

## ÖZGEÇMİŞ

### KİŞİSEL BİLGİLER

**Adı Soyadı:** Mehmet Emin ÇİFCİ  
**Uyruğu:** T.C.  
**Doğum Yeri ve Tarihi:** Konya / 01.05.1981  
**Telefon:** 0534 295 0 294  
**e-mail:** [mehmetemin.ciftci@saglik.gov.tr](mailto:mehmetemin.ciftci@saglik.gov.tr); [emin.eposta@gmail.com](mailto:emin.eposta@gmail.com)



### EĞİTİM

Derece	Adı, İlçe, İl	Yıl
Lise	Erbil Kuru Süper Lisesi, Selçuklu, Konya	1999
Üniversite	Selçuk Üniversitesi Bilgisayar Mühendisliği, Selçuklu, Konya	2005

### İŞ DENEYİMLERİ

Yıl	Kurum	Görevi
2013 - devam	Konya Numune Hastanesi	Mühendis
2008 - 2010	Türkiye Halk Bankası	Uzman Yardımcısı

**UZMANLIK ALANI** Yapay Sinir Ağları, Mikro İşlemci Yazılımı, Ağ Tasarımı, Web Programlama, Nesne Yönelimli Uygulama Geliştirme, Veri Madenciliği, Veritabanı Tasarımı ve Güncelleme, Nesnel Komponent Geliştirme, Sunucu / Sistem Kurulumu ve Tasarımı, Elektronik Devre Tasarımı ve Analizi, SQL Raporlama ve Analiz.

**YABANCI DİLLER** İngilizce, Çince

### BELİRTMEK İSTEDİĞİNİZ DİĞER ÖZELLİKLER

Dağcılık, kitap okuma ve düşünme, sinema, tiyatro, teknoloji, spor (futbol, basketbol, kondisyon) ilgi alanlarını oluşturur.

1. Yayın: Yüksek Lisans Tezi "Yeni Hibrit Sistemle Güçlendirilmiş Yapay Zekâ ile Hastalık Tahmini, IJESC Uluslararası Dergisi Cilt 8 Basım No 6, June 2018