T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

DOKTORA TEZİ

DİYABETLİ HASTALARDA GLİKOZ YOĞUNLUĞU'NUN TAHMİN EDİLMESİ İÇİN YAPAY ZEKA TABANLI BİR KARAR DESTEK SİSTEMİ TASARIMI

Caner OKUTKAN

DANIŞMAN
Prof. Dr. Şakir ESNAF

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Programı

İSTANBUL-2021

Bu çalışma, 27.07.2021 tarihinde aşağıdaki jüri tarafından Endüstri Mühendisliği Anabilim Dalı, Endüstri Mühendisliği Programında Doktora tezi olarak kabul edilmiştir.

Tez Jürisi

Prof. Dr. Şakir ESNAF(Danışman)
İstanbul Üniversitesi-Cerrahpaşa
Mühendislik Fakültesi

Prof. Dr. Şükrü Alp BARAY
İstanbul Üniversitesi-Cerrahpaşa
Mühendislik Fakültesi

Prof. Dr. Hür Bersam BOLAT
İstanbul Teknik Üniversitesi
İşletme Fakültesi

Prof. Dr. Mehmet Mutlu YENİSEY
İstanbul Üniversitesi-Cerrahpaşa
Mühendislik Fakültesi

Prof. Dr. Ferhan ÇEBİİ
İstanbul Teknik Üniversitesi
İşletme Fakültesi

ÖNSÖZ

Bu tez çalışması boyunca çalışmaya liderlik eden ve bana her zaman destek olan tez danışmanım sayın Prof. Dr. Şakir Esnaf'a, tez izleme komitesi üyelerimiz sayın Prof. Dr. Alp Baray'a ve sayın Prof. Dr. Hür Bersam Bolat'a şükranlarımı sunarım.

Ayrıca tez çalışmaları sırasında benimle fikir alışverişlerinde bulunan Doğuş Üniversitesi öğretim üyeleri Ömer Çelepçıkay ile Kıvanç Onan'a ve Matlab programı kodlamalarında bana yardımcı olan arkadaşlarım Egemen Berki Çimen ile Zahid Gürbüz'e teşekkürlerimi sunarım.

Son olarak bu süreçte beni motive eden ablam Cansev Karakuş'a ve eşim Rukiye Okutkan'a göstermiş oldukları desteklerden dolayı teşekkürü borç bilirim.

Temmuz 2021 Caner Okutkan

İÇİNDEKİLER

Sayfa No

ÖNSÖZ ..... **i**
İÇİNDEKİLER ..... **ii**
ŞEKİL LİSTESİ ..... **iv**
TABLO LİSTESİ ..... **v**
SİMGE VE KISALTMA LİSTESİ ..... **ix**
ÖZET ..... **x**
SUMMARY ..... **xii**
1. GİRİŞ **1**
2. GENEL KISIMLAR **4**
2.1. DİYABET HASTALIĞI 4
2.2. DİYABET'İN SINIFLANDIRILMASI 5
2.3. DİYABET'İN TEDAVİSİ VE EĞİTİMİ 6
2.3.1. İlaç Kullanımı Olmaksızın Uygulanabilecek Tedavi Yöntemleri 6
2.3.2. İlaç Kullanarak Uygulanabilecek Tedavi Yöntemleri 7
2.4. TÜRKİYE VE DÜNYADA DİYABET 9
2.5. DİYABET TAHMİNİ İLE İLGİLİ LİTERATÜR ARAŞTIRMASI 10
2.5.1. Tahmin Tanımı ve Tahmin Yöntemlerinin Sınıflandırılması 10
2.5.2. Matematiksel Modeller Kullanılarak Yapılan Diyabet Tahmini Çalışmaları 10
2.5.3. Zaman Serileri Modelleri Kullanılarak Yapılan Diyabet Tahmini Çalışmaları 17
2.5.4. Karar Ağacı Algoritmaları Kullanılarak Yapılan Diyabet Tahmini Çalışmaları 21
2.5.5. Mobil Uygulamalar Oluşturarak Yapılan Diyabet Tahmini Çalışmaları 22
2.5.6. Bulanık Mantık Uygulamaları Kullanılarak Yapılan Diyabet Tahmini Çalışmaları 24
2.5.7. Yapay Sinir Ağları Kullanılarak Yapılan Diyabet Tahmini Çalışmaları 26
2.5.8. Destek Vektör Makineleri Kullanılarak Yapılan Diyabet Tahmini Çalışmaları 32
2.5.9. Rastgele Orman Algoritmaları Kullanılarak Yapılan Diyabet Tahmini Çalışmaları 35
2.5.10. Hibrit Modeller Oluşturarak Yapılan Diyabet Tahmini Çalışmaları 37

3. MALZEME VE YÖNTEM 44
3.1. YAPAY SİNİR AĞLARI 44
3.1.1. Yapay Sinir Ağları Yöntemi Ortaya Çıkışı ve Zaman İçerisindeki Gelişimi 44
3.1.2. Yapay Sinir Ağları Modeli Oluşturma Araçları ve Temel Elemanları 46
3.1.3. Yapay Sinir Ağları Sınıflandırılması 52
3.1.4. Yapay Sinir Ağlarının Eğitim Yöntemleri 53
3.1.5. Yapay Sinir Ağları Model Yöntemlerinin Diğer Yöntemlerle Karşılaştırılması ve Değerlendirilmesi 55
3.2. DESTEK VEKTÖR MAKİNALARI 56
3.2.1. Destek Vektör Makinaları Yöntemi Ortaya Çıkışı ve Zaman İçerisindeki Gelişimi 56
3.2.2. Destek Vektör Makinaları Mimarisi ve Temel Elemanları 57
3.2.3. Destek Vektör Makinaları'nın Sınıflandırılması 58
3.2.4. Destek Vektör Makinaları Yöntemlerinin Diğer Yöntemlerle Karşılaştırılması ve Değerlendirilmesi 64
3.3. RASTGELE ORMAN ALGORİTMASI 65
3.3.1. Rastgele Orman Algoritmaları Yöntemi Ortaya Çıkışı ve Zaman İçerisindeki Gelişimi 66
3.3.2. Rastgele Orman Algoritmaları Temel Yapısı ve Uygulama Aşamaları 67
3.3.3. Rastgele Orman Algoritmaları Yöntemlerinin Diğer Yöntemlerle Karşılaştırılması ve Değerlendirilmesi 73
3.4. PERFORMANS ÖLÇÜTLERİ 74
3.5. R STUDIO 76
4. BULGULAR 78
4.1. VERİ SETİ DÜZENLEMESİ 79
4.2. ALTERNATİF ÇÖZÜM YÖNTEMLERİNİN UYGULANMASI 83
4.2.1. Senaryo 1 Uygulama 83
4.2.2. Senaryo 2 Uygulama 92
4.2.3. Senaryo 3 Uygulama 97
4.2.4. Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözümler 106
4.2.5. Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözümler 117
4.2.6. Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler 123
5. TARTIŞMA VE SONUÇ 134
KAYNAKLAR 140
ÖZGEÇMİŞ 153

ŞEKİL LİSTESİ

Sayfa No

Şekil 3.1: Biyolojik sinir hücresi (Maltarollo ve diğ., 2013)……..……43

Şekil 3.2: Yıl bazlı yapay sinir ağlarından bahsedilen çalışmaların sayısal değişimi (www.webofknowledge.com, 2019)……..…..…….....45

Şekil 3.3: Biyolojik sinir ağ model yapısı (Maltarollo ve diğ., 2013)……..……...45

Şekil 3.4: YSA elemanları (Öztemel, 2006)…….…….....……46

Şekil 3.5: Lineer transfer fonksiyonu (Çayıroğlu, 2015)……..…….……..48

Şekil 3.6: Logaritmik sigmoid fonksiyonu (Çayıroğlu, 2015)…..……...49

Şekil 3.7: Tanjant-sigmoid fonksiyonu (Çayıroğlu, 2015)…..……49

Şekil 3.8: İleri beslemeli sinir ağ modeli (Öztemel, 2006)……..…….....51

Şekil 3.9: Geri beslemeli sinir ağ modeli (Kabalcı, 2015)…...……51

Şekil 3.10: Optimum ayırıcı düzlemler (Cherkassky, 1998)……..…….……55

Şekil 3.11: Yıl bazlı destek vektör makinalarından bahsedilen çalışmaların sayısal değişimi (www.webofknowledge.com, 2019)……....…….….....55

Şekil 3.12: Destek vektör makinaları temel örnek mimarisi (Güraksın, 2015)……....56

Şekil 3.13: Hata dilim değeri kayıp fonksiyonu (Moura ve diğ., 2011)……...57

Şekil 3.14: Ayırıcı düzlem ve sınıflar (Kabaoğlu, 2010)…….....58

Şekil 3.15: Verilerin doğrusal olarak ayrılmama durumu (Özkan, 2008)……..…..59

Şekil 3.16: Doğrusal olmayan sınıflandırma (Özkan, 2008)…….……...60

Şekil 3.17: Lineer olmayan verilerin üç boyutlu uzaya taşınarak sınıflandırılması (Ayhan ve Erdoğmuş, 2014)…….....……60

Şekil 3.18: Yıl bazlı rastgele orman algoritmalarından bahsedilen çalışmaların sayısal değişimi (www.webofknowledge.com, 2019)…….....64

Şekil 3.19: Önyükleme örnek yapısı (Ayas, 2014)……...……..…..……67

Şekil 4.1: Çalışmanın bütünsel akış diyagramı……...……..…..……....75

TABLO LİSTESİ

Sayfa No

Tablo 2.1: İnsülin tipleri, özellikleri ve etki süreleri (Sağlık Bakanlığı, 2015)…….....8

Tablo 2.2: Diyabet Atlası 2019 verileri (www.diabetesatlas.org, 2019)……...9

Tablo 2.3: Matematiksel modeller kurularak yapılan çalışmaların analizi……..……...15

Tablo 2.4: Zaman serileri kullanılarak yapılan çalışmaların analizi……....20

Tablo 2.5: Yapay sinir ağları kullanılarak yapılan çalışmaların analizi……...30

Tablo 2.6: Destek vektör makinaları kullanılarak yapılan çalışmaların analizi……...33

Tablo 2.7: Rastgele orman algoritmaları kullanılarak 2010 yılı sonrasında yapılan bazı çalışmaların analizi…….....……36

Tablo 2.8: Hibrit modeller oluşturularak yapılan çalışmaların analizi…….....41

Tablo 3.1: 1975 yılı öncesi yapay sinir ağları gelişimi (Elmas, 2007)…….……...44

Tablo 3.2: Biyolojik sinir ağ yapısı ile yapay sinir ağları yapısı karşılaştırılması (Sarıoğlu, 2013).....46

Tablo 3.3: Örnek model yapısı çözüm sonuçları…….....73

Tablo 4.1: İnsülin ile ilgili kodlar ve açıklamaları…….....76

Tablo 4.2: Kandaki glikoz yoğunluğu ölçüm sonuçları ile ilgili kodlar ve açıklamaları…….77

Tablo 4.3: Egzersiz / Etkinlik ile İlgili kodlar ve açıklamaları……77

Tablo 4.4: Örnek veri seti yapısı.....…….....78

Tablo 4.5: Senaryo 1'de yer alan girdi ile çıktı değişkenleri ve açıklamaları..……....79

Tablo 4.6: Senaryo 1'de yer alan Veri Seti 1 çalışma yapısı kesiti görünümü……....80

Tablo 4.7: Senaryo 1'de yer alan Veri Seti 1 özellikleri…….....82

Tablo 4.8: Senaryo 1-Veri Seti 1 yöntemlerin performans sonuçları……..82

Tablo 4.9: Senaryo 1'de yer alan Veri Seti 2 özellikleri……..83

Tablo 4.10: Senaryo 1-Veri Seti 2 yöntemlerin performans sonuçları……..……..83

Tablo 4.11: Senaryo 1'de yer alan Veri Seti 3 özellikleri……84

Tablo 4.12: Senaryo 1-Veri Seti 3 yöntemlerin performans sonuçları……84

Tablo 4.13: Senaryo 1'de yer alan Veri Seti 4 özellikleri……....85

Tablo 4.14: Senaryo 1-Veri Seti 4 yöntemlerin performans sonuçları……....85

Tablo 4.15: Senaryo 1'de yer alan Veri Seti 5 özellikleri……86

Tablo 4.16: Senaryo 1-Veri Seti 5 yöntemlerin performans sonuçları……86

Tablo 4.17: Senaryo 1'de yer alan Veri Seti 6 özellikleri……87

Tablo 4.18: Senaryo 1-Veri Seti 6 yöntemlerin performans sonuçları.....……..87

Tablo 4.19: Senaryo 1-Tüm veri setlerinin performans sonuçları…….....88

Tablo 4.20: Senaryo 2'de yer alan girdi ile çıktı değişkenleri açıklamaları.....……88

Tablo 4.21: Senaryo 2'de yer alan veri seti çalışma yapısı görünümü.....89

Tablo 4.22: Senaryo 2-Veri Seti 1 yöntemlerin performans sonuçları……91

Tablo 4.23: Senaryo 2-Veri Seti 2 yöntemlerin performans sonuçları……91

Tablo 4.24: Senaryo 2-Veri Seti 3 yöntemlerin performans sonuçları……92

Tablo 4.25: Senaryo 2-Veri Seti 4 yöntemlerin performans sonuçları……92

Tablo 4.26: Senaryo 2-Veri Seti 5 yöntemlerin performans sonuçları……92

Tablo 4.27: Senaryo 2-Veri Seti 6 yöntemlerin performans sonuçları..……..93

Tablo 4.28: Senaryo 2-Tüm veri setlerinin ortalama performans sonuçları……93

Tablo 4.29: Senaryo 3'de yer alan girdi ile çıktı değişkenleri açıklamaları ...……94

Tablo 4.30: Senaryo 3'de yer alan Veri Seti 1 özellikleri……96

Tablo 4.31: Senaryo 3-Veri Seti 1 yöntemlerin performans sonuçları……96

Tablo 4.32: Senaryo 3'de yer alan Veri Seti 2 özellikleri……97

Tablo 4.33: Senaryo 3-Veri Seti 2 yöntemlerin performans sonuçları……....97

Tablo 4.34: Senaryo 3'de yer alan Veri Seti 3 özellikleri……98

Tablo 4.35: Senaryo 3-Veri Seti 3 yöntemlerin performans sonuçları……98

Tablo 4.36: Senaryo 3'de yer alan Veri Seti 4 özellikleri……99

Tablo 4.37: Senaryo 3-Veri Seti 4 yöntemlerin performans sonuçları……99

Tablo 4.38: Senaryo 3'de yer alan Veri Seti 5 özellikleri……..100

Tablo 4.39: Senaryo 3-Veri Seti 5 yöntemlerin performans sonuçları……..100

Tablo 4.40: Senaryo 3'de yer alan Veri Seti 6 özellikleri……..101

Tablo 4.41: Senaryo 3-Veri Seti 6 yöntemlerin performans sonuçları…….....101

Tablo 4.42: Senaryo 3-Tüm veri setlerinin ortalama performans sonuçları……..102

Tablo 4.43: Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözüm'de mutlak sapma değerleri.....112

Tablo 4.44: Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözüm'de mutlak yüzde hata ortalama değerleri…….113

Tablo 4.45: Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözüm'de yer alan veri seti çalışma yapısı kesit görünümü……....117

Tablo 4.46: Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözüm'e göre hibrit yöntemlerin mutlak sapma değerleri……...…….……..118

Tablo 4.47: Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözüm'e göre hibrit yöntemlerin mutlak yüzde hata ortalama değerleri……...…….118

Tablo 4.48: Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e göre ortalama mutlak sapma değerleri…..…….124

Tablo 4.49: Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e göre mutlak yüzde hata ortalama değerleri……..125

Tablo 4.50: Senaryolara göre yöntemlerin tahmin başarıları…….....125

Tablo 4.51: Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e göre analiz edilen Test Veri Seti 7 özellikleri…….126

Tablo 4.52: Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e göre analiz edilen Test Veri Seti 8 özellikleri……..…….....126

Tablo 4.53: Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e göre analiz edilen Test Veri Seti 9 özellikleri…….127

Tablo 4.54: Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e göre analiz edilen Test Veri Seti 10 özellikleri..…….128

Tablo 4.55: Test verilerine uygulanan hibrit yöntemlerin mutlak sapma değerleri…….....128

Tablo 4.56: Test verilerine uygulanan hibrit yöntemlerin mutlak yüzde hata ortalama değerleri.....129

SİMGE VE KISALTMA LİSTESİ

Simgeler **Açıklama**

N : Rastgele Orman Algoritmaları İçin Orjinal Eğitim Verisinin Boyutu
n : Rastgele Orman Algoritmaları İçin Orjinal Veri Setindeki Gözlem Sayısı
m : Rastgele Orman Algoritmaları İçin Veri Setinden Seçilen Örneklem
Ŧ : Rastgele Orman Algoritmaları İçin Test Verisi (Önyükleme Örnekleri)
Ĺ : Rastgele Orman Algoritmaları İçin Eğitim Verisi (Önyükleme Örnekleri)
Ĺ **(k)** : Rastgele Orman Algoritmaları İçin Alt Veri Setleri (Alt Eğitim Setleri)
ßi : Rastgele Orman Algoritmaları İçin Önyükleme Örnekleri, (ß1, ß2, ß3 v.b.)
Ćİ : Rastgele Orman Algoritmaları İçin Önyükleme Örnekleri Sınıflandırıcılar
Ɵ : Rastgele Orman Algoritmaları İçin Önyükleme Tekrarları
C̊ : Rastgele Orman Algoritmaları İçin Final Sınıflandırıcı
y : Rastgele Orman Algoritmaları İçin Sınıf Etiketi
x : Rastgele Orman Algoritmaları İçin Girdi Verisi
Є : Destek Vektör Makinaları İçin Hata Terimi
C : Destek Vektör Makinaları İçin Düzenleme Faktörü
Xi : Yapay Sinir Ağları İçin Girdi Değişkenleri
Wi : Yapay Sinir Ağları İçin Ağırlıklar
Yi : Yapay Sinir Ağları İçin Çıktı değişkenleri

Kısaltmalar **Açıklama**

DVM : Destek Vektör Makinaları
ROA : Rastgele Orman Algoritması
YSA : Yapay Sinir Ağları
YZ : Yapay Zeka

ÖZET

DİYABETLİ HASTALARDA GLİKOZ YOĞUNLUĞU'NUN TAHMİN EDİLMESİ İÇİN YAPAY ZEKA TABANLI BİR KARAR DESTEK SİSTEMİ TASARIMI

DOKTORA TEZİ

Caner OKUTKAN

İstanbul Üniversitesi-Cerrahpaşa

Lisansüstü Eğitim Enstitüsü

Endüstri Mühendisliği Anabilim Dalı

Danışman : Prof. Dr. Şakir ESNAF

Diyabetli hastalar sürekli kan şekeri seviyelerini takip etmektedirler. Diyabet hastalarının tedavilerini kolaylaştırmak amacıyla, tedavi için gerekli kan şekeri konsantrasyonunu tahmin etmek için yapay zeka tabanlı bir karar destek sistemi tasarlamak bu tezin amacıdır. Bu doğrultuda, tez kapsamında diyabetik hastaların sağlık durumlarını, yaşam tarzlarını takip etmek, kandaki glikoz konsantrasyonlarına göre tedavi süreçleri geliştirmek ve uygulanan tedavilere göre tahminler yapmak amaçlanmaktadır.

Bu çalışmada veri kümeleri için bağımsız değişkenler kullanılarak veri madenciliği yöntemleri uygulanmıştır. Kandaki glikoz konsantrasyonunu tahmin etmek için yapay zeka tabanlı karar destek sisteminin tasarlanmasında yapay sinir ağları, rastgele orman algoritması ve destek vektör makineleri kullanılarak hibrit modeller oluşturulmuştur. Çalışmanın amacı, yüksek performanslı hibrit modelleri kullanarak glikoz seviyesi tahmini için yapay zeka tabanlı uzman karar destek sistemi oluşturmaktır. Bu tezi literatürde farklı konumlandırılması düşünülen durum belirtilen 3 yöntemin birarada kullanılması ve 27 adet farklı hibrit çözüm algoritması elde edilmesidir. Bu algoritmalardan elde edilen en iyi performansa sahip hibrit modellerin seçilmesi stratejisinin uygulanması düşünülmüştür. Bu 3 yöntemin bir arada hibrit olarak çözüme girdiği bu durum literatürde görülmeyen bir durum olduğu gözlemlenmiş ve buna göre çözüm şekillenmiştir.
Temmuz 2021, 162 sayfa.

Anahtar kelimeler: Yapay Sinir Ağları, Hibrit Modeller, Rastgele Orman Algoritması, Diyabet, Destek Vektör Makinası, Kan şekeri tahmini.

SUMMARY

THE DESIGN OF AN ARTIFICIAL INTELLIGENCE BASED DECISION SUPPORT SYSTEM FOR PREDICTION OF GLUCOSE CONCENTRATION IN DIABETES PATIENTS

Ph.D. THESIS

Caner OKUTKAN

Istanbul University-Cerrahpasa

Institute of Graduate Studies

Department of Industrial Engineering

Supervisor : Prof. Dr. Şakir ESNAF

Patients with diabetes are constantly monitoring their blood sugar levels. The aim of this thesis is to design an artificial intelligence-based decision support system to predict the blood glucose concentration required for treatment in order to facilitate the treatment of diabetic patients. In this direction, within the scope of the thesis, it is aimed to follow the health status and lifestyle of diabetic patients, to develop treatment processes according to glucose concentrations in the blood and to make predictions according to the treatments applied.

Data mining methods by using independent variables for the data sets are applied in this study. In order to estimate the glucose concentration in the blood, hybrid models were created by using artificial neural networks, random forest algorithm and support vector machines in the formation of artificial intelligence based decision support system. In study's aim is create an artificial intelligence-based expert decision support system for glucose level prediction using the highest performance hybrid models. The situation that is thought to be positioned differently in this thesis in the literature is the use of 3 methods together and obtaining 27 different hybrid solution algorithms. The strategy of selecting the best performing hybrid models obtained from these algorithms is considered. It has been observed that this situation, in which these 3 methods are solved together as a hybrid, is not seen in the literature and the solution has been formed accordingly

July 2021, 162 pages.

Keywords: Artificial Neural Networks, Hybrid Models, Random Forest Algorithm, Diabetes, Support Vector Machine, Prediction Blood Glucose Level.

1. GİRİŞ

Bu tezin amacı; diyabet hastalığının teşhisi ve tedavisinde yapay zeka tabanlı bir karar destek sisteminin tasarlanmasıdır. Diyabet hastalarının sağlık durumlarının ve yaşam tarzlarının takip edilerek yeni tedavi yöntemleri uygulamak ve uygulanan tedavilere göre ileriye yönelik tahminler sağlanması amaçlanmıştır.

Uluslararası Diyabet Federasyonu'na göre (2019), 400 milyondan daha fazla sayıda diyabet hastasının olduğu tespit edilmiştir. Yapılan araştırmalar diyabet sebebiyle 1 dakika başına 10 kişinin öldüğünü göstermiştir (UDF, 2019). Hastalığın manevi boyutu ile beraber maddi boyutu da ciddi anlamda önemli olduğu görülmüştür. Diyabet Atlası 2015 verilerine göre dünyada yapılan sağlık giderlerinin yaklaşık % 12'lik kısmı (yaklaşık 673 milyar $) diyabet hastalığı teşhis ve tedavisi için kullanılmıştır. İleriye yönelik yapılan tahminlerde, diyabet hasta oranlarının daha da artacağı görülmüştür.

IDF 2040 yılı tahminlerine göre, her 10 yetişkinden 1 tanesinin diyabet hastası olacağı (642 milyon kişi) öngörülmüştür. Hastalığın maddi boyutları düşünüldüğünde, küresel sağlık harcamalarının 2040 yılı için diyabet hastalığının payının da 802 milyar $'ı aşacağı öngörülmüştür. Bu anlamda diyabet hastalığı teşhisi ve tedavisi üzerine çalışılmasının önemli olduğu sonucuna varılmıştır.

Bu araştırmada diyabet hastalığı teşhisi ve tedavisinde önemli olan kandaki glikoz yoğunluğunun tahmin edilmesi için yapay zeka sistemlerinin kullanıldığı bir model önerisi sunulmuştur. Bu amaçla bu çalışmada tasarlanması amaçlanan yapay zeka tabanlı karar destek sistemleri, karar verme sürecinde yer alan analizleri desteklemek üzere hazırlanmıştır. Tasarlanacak yapay zeka tabanlı karar destek modeli; girilecek bilgilere göre ihtiyaç duyulan sonuçlar üretmiş ve bununla birlikte yeni girilecek veriler ile sistem karar noktaları kendisini sürekli güncel tutmuştur.

Çalışmada elde edilmiş veri setlerine yönelik değişkenler kullanılarak veri madenciliği yöntemleri uygulanmıştır. Yapay zeka tabanlı karar destek sistemleri oluşturmada; yapay sinir ağı, rastgele orman algoritması ve destek vektör makina algoritmalarına dayalı yöntemlerinin kullanılarak yeni hibrit modellerin oluşturulması, oluşturulan yeni hibrit modellerin

performanslarının karşılaştırmalarının yapılması, en yüksek performansa sahip hibrit modeli kullanarak yapay zeka tabanlı uzman karar destek sistemi oluşturulması amaçlanmıştır. Çalışmada tasarlanacak yapay zeka tabanlı karar destek sisteminin üç seviyeden oluşması planlanmıştır. İlk seviye gerekli ve yeterli sayıda verilerin toplandığı veritabanından oluşmuştur. İkinci seviyede hibrit modeller kurulmuş ve çözümleri yapılmıştır. Üçüncü seviyede ise oluşturulan hibrit modeller arasından seçim yapılmış ve hangi hibrit modellerin kulllanılacağı ile ilgili kararlar verilmiştir. Performansı en yüksek hibrit çözüm modelleri tespit edilmiştir.

Çalışmanın kapsamı çerçevesinde ele alınan şeker hastalığı teşhisi ve tedavisi kan şekerinin ölçülmesine göre düzenlenmiştir. Tıp literatüründe çeşitli kan şeker ölçüm zamanları vardır. Örneğin; yemek öncesinde uzun süreli aç kalma durumundan sonra açlık durumunda kan şekeri ölçümü yapılabilir, yemek sonrasında tokluk durumunda kan şekeri ölçümü yapılabilir veya Hemoglabin A1c (glikatlı hemoglobin) testleri yapılabilir. Bu testler sonucunda elde edilen kandaki glikoz değerlerine göre şeker hastalarına tedavi yöntemleri uygulanır. Tedavi yöntemleri arasında beslenme alışkanlıklarının düzenlenmesi, viral yoldan alınan ilaç tedavisi, insülin enjeksiyonu başlıca hastalık sürecinde kullanılan yöntemlerdendir. Bu tedavi yöntemlerinin kullanımında ve kullanım miktarlarının tespiti için hastanın kanındaki glikoz miktarı açlık kan şekeri, tokluk kan şekeri veya HbA1c testleri aracılığıyla sağlanmaktadır. Çalışmada hastanın aldığı besinler, harcamış olduğu aktivite, yaş, cinsiyet, hastalık türü değişkenlerine göre kan şekerinin ölçülmesinde hastaların rahatsızlık duyduğu, acı ve ağrı çektiği düzenli olarak kan verilmesine gerek duyulmaksızın kandaki glikoz miktarının tahminlenmesi ve buna göre teşhis-tedavi yapılması hedeflenmiştir.

Literatürde tahmin yapmada yapay zeka yöntemlerinin kullanıldığı çok fazla sayıda çalışmaya rastlanmıştır. Bununla birlikte yapılmış çalışmalar içerisinde, bu çalışmada kullanılmış destek vektör makinaları, rastgele orman algoritması ve yapay sinir ağları yöntemlerinin kullanıldığı araştırmaların da yer aldığı görülmüştür. Ancak bu üçünün bir arada kullanıldığı bu tip bir hibrit araştırmaya rastlanmamıştır, bu yönüyle özgünlük anlamında literatüre katkı sağlayacağı düşünülmüştür.

Bu çalışmada, 3 temel yöntemin bir arada kullanılarak hibrit model oluşturulması, çalışmayı literatürdeki diğer çalışmalardan ayırmıştır. Literatür tarandığında tahmin problemleri çözüm

yöntemlerinin, ekseriyetle ayrı ayrı çözüm yöntemleri olarak uygulandığı görülmüştür. Şöyle ki sadece yapay sinir ağı kullanılarak yapılan çalışmalar, sadece destek vektör makinaları kullanılarak yapılmış çalışmalar veya sadece rastgele orman algoritması kullanılarak yapılan çalışmalar gibi. Bu yöntemlerden en son bulunanı olan rastgele orman algoritmaları 20 yıl önce önerilmiş ve bu sebeple diğer çözüm yöntemlerine göre bu alanda yapılan çalışma sayısı yapay sinir ağları ve destek vektör yöntemlerine göre daha az olduğu görülmüştür.

Gelişen teknolojiyle beraber bu yöntemlerin birarada kullanılarak yöntemlerin hangisinin daha iyi sonuç verdiği ile ilgili çalışmalarda son yıllarda ortaya çıkmıştır. Şöyle ki tespit edilen bir soruna yönelik hem YSA ve hem de ROA uygulanmış. Veya hem DVM ve hem YSA uygulanmış. Bu 2 yöntem karşılaştırılmış, hangi yöntemin daha iyi performans verdiği tespit edilmiştir.

Akademik çalışma sayısındaki artış araştırmacılarda artık 3 veya daha fazla yöntemin kullanıldığı modelleri oluşturmaya itmiştir. Burada farklı çözüm arayışlarına girilmiştir. Hibrit olarak 3 yöntemin birarada kullanıldığı çalışmalar gözlenmiştir. Burada sorunun çözümünde bazı çalışmalarda da bu 3 yöntem ikili çözüm olarak birarada kullanılmış ve farklı ikili hibrit çözümler oluşturulmuştur. Örneğin; DVM-ROA, YSA-DVM, YSA-ROA v.b. yöntemler bir arada kullanılmıştır.

Bu çalışmayı literatürde farklı konumlandırılması düşünülen durum 3 yöntemin birarada kullanılarak 27 adet farklı hibrit çözüm algoritması elde edilmesi ve bu algoritmalardan elde edilen en iyi performansa sahip hibrit modellerin seçilmesi stratejinin uygulanması olduğu düşünülmüştür. Bu 3 yöntemin bir arada hibrit olarak çözüme girdiği bu durum literatürde görülmeyen bir durum olduğu gözlemlenmiş ve buna göre çözüm şekillenmiştir. Bununla birlikte diyabet hastaları için kan şekeri ölçümünün devamlı olarak yapılması gerekliliği ve yaşam kalitesini olumsuz etkileyen zorlu bir süreç olduğu bilinmektedir. Sağlık alanındaki bu soruna genel itibarıyla sağlıkçılar perspektifinden yaklaşılırken, burada girdi ve çıktılar düzenlenerek mühendislik bakış açısı ile yapay zeka yaklaşımlarından yararlanılmaya çalışılmış ve farklı bir perspektiften çözüme gidilmeye çalışılmıştır.

2. GENEL KISIMLAR

2.1. DİYABET HASTALIĞI

Türkiye Endokrinoloji ve Metabolizma Derneği (2019), tıp literatürü adıyla Diabetes Mellitus olarak tabir edilen halk arasında bilinen adıyla şeker hastalığını, vücudun insulin noksanlığı veya insülin kaynaklı problemler sebebiyle organizmlardaki temel besin öğelerinden (proteinlerden, karbonhidratlardan ve yağlardan) yeterli derecede verim sağlanamaması dolayısıyla sürekli olarak kontroller gerektiren ve kalıcı olan bir hastalık olarak tanımlamıştır.

Diyabet hastalığı, sebep olduğu diğer patolojiler ve pahalı tedavisi sebebiyle önemi gün geçtikçe artan bir hastalık haline gelmiştir (Olgun, 2015).

Amerikan Diyabet Derneği'ne (2019) göre insan vücudu için temel enerji kaynağı olan glikozun, dolaşımdaki miktarının sabit bir aralıkta tutulması önemlidir. Her geçen yıl yapılan araştırmalar ile kan glikoz referans aralıkları güncellenmektedir.

Sağlıklı kişilerde açlık kan şekeri (AKŞ) seviyesi desilitrede 100 miligram seviyesinden daha düşük olmalıdır. (10 desilitre 1 litreye tekabül etmektedir.) Bu seviye desilitrede 100 miligram ile 126 miligram arasında bir seviyesinde olur ise bu durum hastada diyabet riski olduğunu gösterir. Eğer açlık kan şekeri desilitrede 126 miligram ve üzerinde ölçümlenirse hastaya diyabet ön tanısı konulabilir (Ulate ve diğ., 2008).

Şeker hastalığı diğer bir ifadeyle diyabet, vücudun insulin üretememesi veya üretmeye çalıştığı insulin hormonunun vücut tarafından ihtiyaç duyulan miktarın aşağısında olmasından kaynaklı vücutta bozukluklara yol açmaktadır. İnsülin, pankreas ismindeki organımızdan salgılanan bir hormondur (Türkiye Diyabet Vakfı, 2019).

Diyabet hastaları kan şekeri düşüklüğü kaynaklı hipoglisemi ve hızlı kan şekeri yüksekliğinden kaynaklı hiperglisemi problemleri ile karşı karşıya gelebilir. Tip 1 ve tip 2 ilaç kullanan diyabet hastalarında hem hipoglisemik, hem de hiperglisemik ataklar görülebilmektedir (Türkiye Diyabet Vakfı, 2019).

Hipoglisemi; vücut içi titreme, soğuk terleme, öfke, dalgınlık, yorgunluk, açlık, kalp çarpıntısı, baş dönmesi-ağrısı, endişe, görme kaybı gibi ciddi semptomlar ile kendisini tanımlar. Bu problemlerden hipoglisemiden kaçınmak için uzun süre aç kalınmamalıdır. Yapılan araştırmalarda hipoglisemi'nin sebebinin kandaki şeker yoğunluğunun desilitrede 70 miligram seviyesinin altına inmesinden kaynaklı olduğu bulunmuştur. Eğer uzun süre aç kalınıyorsa hipoglisemiye mahal vermemek için biraz şeker ve meyve suyu takviyesi mutlaka yapılmalıdır (Terzihan, 2020).

Hiperglisemi, kan şekerinin yükselmesi anlamına gelmektedir. Halsizlik durumu, sebepsiz karın ağrısı, idrara sık çıkma, ağızda meydana gelen kuruluk hissi ve çokça su içme isteği, nefes alış-verişinde zorlanma, idrarda yüksek miktarda ketona ve şekere rastlanması v.b. durumlara neden olabilir. Hiperglisemi kan şekerinin birden yükseldiği durumlarda meydana gelir. İnsülin kullanılmasına rağmen meydana geliyorsa sebepleri; gerektiği kadar insulin kullanmamak, bozulmuş insulin ilacı kullanmak, hassasiyeti problemli bölgeye insulin yapmak, çok fazla besin tüketimi ve fiziksel aktivite yapmamak olabilir. Hiperglisemi'yi önlemek için hasta kan şekeri takibini düzenli olarak yapmalı ve önlemini almalıdır (Çağlayan, 2020).

2.2. DİYABET'İN SINIFLANDIRILMASI

Amerikan Diyabet Derneği (2013), diyabet hastalığını 4 ayrı sınıfa ayırmıştır. Birinci sınıf; Tip 1 diyabettir. Tip 1 diyabet hastaları, genç yaştaki hastalardır. Pankreasta insulin hormonunu üreten β hücrelerinin hasar görmesi sonucunda insulin üretiminin olmaması veya çok az üretiminden kaynaklı olarak ortaya çıktığı tespit edilmiştir. Tahminlere göre Tip 1 diyabet hasta sayısı, 20 milyon ile 45 milyon arasında bir değerdedir. Bu sınıfta yer alan hastaların büyük kısmında, çocukluk dönemi ve ergenlik dönemlerinde hastalıkların belirtilerinin ortaya çıktığı gözlemlenmiştir. Ciddi akut komplikasyonlardan; hipoglisemik ataklar ve diyabetik ketoasidoz rahatsızlıklarının oranla fazlaca görüldüğü diyabet tipidir (Taşkaya, 2014).

İkinci sınıf; Tip 2 diyabettir. Tip 2 diyabette vücutta pankreas salgı bezleri içerisinde konumlanmış Langerhans adacıkları içerisinde yer alan β hücreleri insulin üretebilme kapasitesine sahiptir. Fakat ilginç bir şekilde vücut kendi ürettiği bu insulin hormonuna tepki

vermektedir. Bundan dolayı vücut, insulini yeterli ve verimli olarak kullanamamaktadır. Diyabet tiplerinden, tip 2 diyabet, en çok rastlanan diyabet tipidir. Dünyada ki diyabet hastalarının minimum % 90'ı tip 2 diyabet hastasıdı (Taşkaya, 2014).

Üçüncü sınıf; gebelik diyabetidir. Gebelik diyabeti; hamile kadınlarda ortaya çıkmaktadır. Yapılan doğum ile beraber diyabet hastalığı semptomları sona ermektedir. Kalıcı bir şeker hastalığına yol açmadığı düşünülmekte olmasına karşın Harris ve diğ. (2012) yaptıkları çalışmada gebelik diyabeti olan kadınlarda, hastalık teşhisinden sonra geçen 25 yıllık sürede tip 2 diyabet hastalığının görülme olasılığının diğer kadınlara göre daha fazla olduğunu tespit etmiştir. Bunu önlemek için veya risk oranını azaltmak için gebe kadınlara düzenli olarak spor yapmaları önerilmektedir.

Dördüncü ve son sınıf; tasniflenememiş farklı diyabet türlerini içermektedir. Bazı endikasyonlar sonucunda β hücrelerinde meydana gelen hasarlar bu noktada insulin hormonunun üretimine mani olmaktadır. Mide hastalıkları, pankreas hastalıkları, çeşitli kimyasallara maruz kalma, kullanılan ilaçların yan etkileri v.b. durumlar hastalarda diyabet hastalıklarına yol açabilir (Amerikan Diyabet Derneği, 2013).

2.3. DİYABET'İN TEDAVİSİ VE EĞİTİMİ

Her hastalıkta doğruluğunun kanıtlandığı gibi diyabet hastalığında da tedavi aşaması, öncelikle hastalığın teşhis edilmesi ile beraber başlamaktadır. Bazı kaynaklara göre 4500 yıllık olan diyabet hastalığının, ilk zamanlar idrar içerisindeki şeker kontrolü yapılarak teşhisi sağlanmaktaydı. Teknolojinin gelişmesi ile beraber hastalık teşhisi ve tedavi yöntemlerinde de gelişmeler yaşanmıştır. Şu an diyabet hastalarına konulacak teşhis ve uygulanacak tedavi aşamalarında kan şekeri ölçümleri yapılarak süreç yürütülmektedir. Kan şekeri ölçümü için hastadan kan almak gereklidir. Alınan kan içerisindeki, kan şekeri yoğunluğu analiz edilmektedir (Sacks, 2011). Kan şekeri ölçümü temel anlamda 3 farklı türde yapılabildiği yukarıda çalışma kısımlarında belirtilmişti.

2.3.1. İlaç Kullanımı Olmaksızın Uygulanabilecek Tedavi Yöntemleri

Bu yöntemleri 3 sınıfa ayırmak mümkündür.

2.3.1.1. Bireysel Olarak Kan Şekeri Takipi

Diyabet hastalarının sağlıklı bir yaşam sürdürmeleri için kanlarındaki glikoz seviyelerinin izlenmesi ve takip edilmesi büyük önem arz etmektedir (Boas ve diğ., 2012). Bilinç sahibi şeker hastaları kaliteli bir yaşam standartı yakalamak için belirli aralıklarla kandaki glikoz konsantrasyonu ile ilgili ölçümler yapmalı, bu ölçümleri kayıt etmeli ve gerektiği vakitlerde ölçüm sonuçlarını doktorlarla paylaşmalıdır (Türk Diyabet Vakfı, 2018). İşin özü olarak her şeker hastası birey aynı zamanda kendi doktoru, kendi hemşiresi olmalıdır.

2.3.1.2. Düzenli ve Sağlıklı Beslenerek Yaşam Sürdürme

İlaç kullanan diyabet hastaları dahil olmak üzere, şeker toleransı yüksek kişiler ve ileride şeker hastası olabilme ihtimali olan tüm bireyler sağlıklı beslenmelidir. Ani kan şekeri artışına sebep olabilecek yüksek yağ ve karbonhidrat ihtiva eden besinlerden uzak durulmalıdır (Türk Diyabet Vakfı, 2018).

2.3.1.3. Egzersiz Yapma

Yapılan araştırmalarda, fiziksel aktivitelerin kan şekerinin düzenlenmesinde önemli bir etkiye sahip olduğu ortaya konulmuştur. Spor veya egzersiz faaliyetleri sayesinde vücuttaki enerji tüketimi artar, vücuda girmiş olan gıdaların kas dokuları tarafından kullanılması sağlanarak kan şekerinin yükselmesinin önüne geçilmiş olunur (Türk Diyabet Vakfı, 2018).

2.3.2. İlaç Kullanarak Uygulanabilecek Tedavi Yöntemleri

2.3.2.1. Oral Tedavi Yöntemleri

Diyabet tipleri; insüline bağımlı olan diyabet ve insüline bağımlı olmayan diyabet şeklinde tasniflemeye tabii tutulmuştur. Bu noktada insüline bağımlı olmayan diyabetler antidiyabetik ilaçlar kullanılarak da tedavi edilmektedir. Sağlıklı ve doğru diyetler uygulanarak beslenilmesine rağmen kan şekerinin kontrol edilemediği durumlarda gerekli ilaçlar alınmalıdır (Atmaca ve Ecemiş, 2012).

Murphy ve Holder (2000) tarafından yapılan çalışmada antidiyabetik ilaçların sınıflandırılmasında 4 temel ayrım yapmıştır. Şu hususa dikkat edilmelidir ki dünya üzerinde şu an insulin hormonu görevi gören hiçbir ilaç ağız yoluyla alınamamaktadır. Antidiyabetik ilaçlar şu şekilde sınıflandırılmıştır; insulin hormonunun salgılanmasını arttıran ilaçlar, insulin duyarlılığını arttıran ilaçlar, glikozun absorbe edilmesini geciktiren ilaçlar ve yeni nesil ajan ilaçlardır.

2.3.2.2. İnsülin Tipleri ve Özellikleri

Vücudumuz'un ürettiği insulin; bazal ve bolus olmak üzere 2 temel tipte yer almaktadır. Bazal insulin, bireyin yemek yemediği zamanlarda gerekli olan insülin dozaj miktarını sağlamaktadır. Bolus insulin ise, yemek durumundan sonra lazım olan yüksek miktardaki insülin üretimini sağlamaktadır. Bazal insulin ve bolus insulin üretim zamanları farklılık göstermektedir (Olgun, 2015).

Tablo 2.1: İnsülin tipleri, özellikleri ve etki süreleri (Sağlık Bakanlığı, 2015).

İnsülin Özelliği	İnsülin'in Etkisini Göstermeye Başladığı Süre	İnsülin Tipi	İnsülin Etki Süresi
Kısa Etkili (Regular Tip İnsülin)	15 dk.- 45 dk. arası / 30 dk.- 1 saat arası	Bolus İnsülin	5 saat – 8 saat arası / 6 saat – 8 saat arası / 4 saat – 6 saat arası
Orta Etkili (NPH Tip İnsülin)	1 saat - 2 saat arası / 1 saat- 3 saat arası	Bazal İnsülin	10 saat – 14 saat arası / 12 saat – 16 saat arası / 18 saat – 24 saat arası
Uzun Etkili (Ultralente Tip insulin)	3 saat – 4 saat arası / 2 saat – 5 saat arası	Bazal İnsülin	24 saat - 30 saat arası / 30 saatten daha uzun süre
Hızlı Etkili	15 dakikadan daha az sürede	Bolus İnsülin	3,5 saat – 4,5 saat arası / 3 saat – 5 saat

İlaç sektöründe üretilen ve piyasada olan ilaçların üretim ve hazırlanış yöntemlerindeki patent ve üretim teknolojileri, etken bakteri, hammadde v.b. farklılıklarından ötürü farklı isim ve tipteki ilaçların gösterdiği başlangıç sürelerinde, etki sürelerinde v.b. bazı farklılıklar tabloda da görülmektedir (Sağlık Bakanlığı, 2015).

2.4. TÜRKİYE VE DÜNYADA DİYABET

Araştırmalar, Türkiye de yaşayan insanlardan nüfusa oranlı diyabetli hasta ortalamalarının, Avrupa ve Dünya geneli ile karşılaştırıldıklarında oldukça ileride olduğunu ortaya koymuştur. Nitekim, Türk Diyabet Vakfı Temel Yılmaz (2016) tarafından açıklanan bilgilere göre Türkiyede ki diyabet hastalığı oranı Avrupa'nın 3 katı, Dünya nufusunun da 2 katı düzeyinde gerçekleşmiştir (Türk Diyabet Vakfı).

Uluslararası Diyabet Federasyonu Diyabet Atlası verilerine göre Dünya genelinde yapılan sağlık harcamalarının yaklaşık %12'lik kısmı ve Türkiye'nin sağlık alanındaki yapmış olduğu harcamaların da yaklaşık % 22'lik kısmı da diyabet teşhis-tedavisi türevleri ile ilgilidir. Bu harcamaların ilgili kullanım yerleri, diyabet ilaçları ve diyabet hastalığına bağlı meydana gelen sağlık sorunlarının giderilmesi amacıyla kullanılan teşhis ve tedavi yöntemleridir (Uluslararası Diyabet Federasyonu, 2019).

Amerikan Diyabet Derneği (ADA) 'nin 2011 yılında hazırlamış olduğu raporda; 2010 yılında 300 milyonun altında olan diyabet hastası sayısının, 2030 yılında 450 milyona yaklaşacağı öngörüsünde bulunmuştur. Fakat 2020 için, Uluslararası Diyabet Federasyonu (IDF) 'na göre 463 milyon diyabet hastası insanının günümüzde yaşamını sürdürmeye çalıştığı bilgisi ortaya konulmuştur. Bu noktada tahminlerin ötesinde bir hızla hasta sayısı artmıştır ve artmaya devam etmektedir. Eş zamanlı olarak şu yorumda yapılabilir, gelişen teknoloji ile beraber diyabet hastaları eskisi gibi çabuk ölmüyor, yaşamlarını sürdürebiliyorlar. Her 2 yorumda da diyabet hasta sayısının arttığı gerçeği değişmediğinden, nüfusu artan diyabet hastalarının kaliteli bir yaşam standartı yakalanması için diyabet alanında daha fazla çalışma yapılması gerekmektedir.

Diyabet hastalığı ve bu hastalığa bağlı olarak gelişen diğer kronik hastalıklar ile ilgili gelişmelerin etki ettiği insan sayısı düşünüldüğünde teknoloji, mühendislik ve tıp alanında multidisipliner çalışmaların yapılması büyük önem arz etmektedir. Bu sebeplerden ötürü çalışmalar incelendiğinde hem sağlık bilgileri, hem mühendislik yöntemleri ve hem de teknolojik ilerlemeler takip edilerek çözümlerin oluşturulduğu görülmüştür. Bu tezde önerilen çalışma, donanım ve yazılım alanlarındaki gelişmelerden yararlanılarak oluşturulmuş ve çalışma sonucunda insanlığa yararlı olabilecek, hastalık kontrolünün daha kolay ve hızlı yapılmasının sağlayabilecek analizler yapılmaya çalışılmıştır.

Tablo 2.2: Diyabet Atlası 2019 verileri (www.diabetesatlas.org, 2019).

Sınıf	Diyabet Hasta Yaş Aralıkları	Diyabet Hastası Oranı
1 (Çocuklar ve ergenler)	0-19 yaş aralığı	% 0,20
2 (Çalışma yaşı grubu)	20-64 yaş aralığı	% 72,00
3 (Yaşlı İnsanlar)	>65 yaş üzeri nüfus	% 27,80

2019, IDF Diyabet Atlası'na göre diyabet hastalarının büyük kısmı 20-64 yaş aralığında meydana gelmiştir. Yaklaşık 352 milyon kişinin bu aralıkta olduğu tahmin edilmiştir. Bu durum aynı zamanda toplumların gelişmesi ve ilerlemesi için bir tehdit unsuru olarak değerlendirilmiştir. Sebebi, bu yaş aralığındaki genç nüfus denilen insanların toplumların enerji-güç potansiyelini göstermesi olarak yorumlanmıştır (www.diabetesatlas.org, 2019).

2.5. DİYABET TAHMİNİ İLE İLGİLİ LİTERATÜR ARAŞTIRMASI

2.5.1. Tahmin Tanımı ve Tahmin Yöntemlerinin Sınıflandırılması

Türk Dil Kurumu tahmini; akla, sezgiye ve bazı verilere dayanarak olabilecek bir şeyi, bir olayı önceden kestirme veya kestirim olarak tanımlamıştır. Tahminlerin amacı gelecekte farklı durumlarda karşılaşılabilecek olaylara karşı eldeki verileri kullanarak hesaplamalar yaparak, öngörülerde bulunmak şeklinde de ifade edilebilir.

Çalışmanın bu bölümünde literatürde yer alan çalışmalar kronolojik sıralamayla gözden geçirilmiş, özellikle çalışmada kullanılan hibrit modelleri oluşturan yöntemler ile ilgili çalışmalar irdelenmiş ve tablolar ile gösterilmiştir. Literatürde yer alan diyabet tahmini ile ilgili çalışmalar alt başlıklara ayrılmıştır. Yazarlar, yıl, tahmin zamanı, yazılım, hasta sayıları ve gözlem zamanları değişkenleri çalışmaları tasniflemede kullanılmıştır. Bunlarla birlikte çalışmalarda kullanılan girdi değişkenleri kendi içerisinde 4 seçeneğe ayrılmıştır. Bu seçenekler glikoz, insulin dozaj, karbonhidrat ve diğer girdi değişkenleri olarak belirlenmiştir. Çalışma da aynı zamanda performans ölçüm kriterleri de incelenmiş ve bu kriterler de değerlendirilmiştir.

2.5.2. Matematiksel Modeller Kullanılarak Yapılan Diyabet Tahmini Çalışmaları

Bergman ve diğ. (1979) literatürde yazılı olarak yer alan ilk çalışmalardan matematiksel minimal model metodunu ortaya koymuştur. Çalışmada bir organizmada pankreas organının

cevap verebilme yeteneği ve insulin duyarlılığının kantitatif tahminini yapmak amaçlanmıştır. Bu amaçla farklı insulin dozları enjekte edilen köpeklerden alınan veriler çalışmada kullanılmış ve karşılaştırmalar yapılmıştır. Bu model oldukça az sayıda değişken içerdiği için minimal model olarak isimlendirilmiştir. Model 3 adet farklı diferansiyel eşitlikten oluşmuştur. Bu diferansiyel eşitlik denklemleri ile plazma glikoz yoğunluğu, plazma insulin yoğunluğu ve doku insulin yoğunluğunu hesaplamak amaçlanmıştur. İnsülin ölçüm zamanları girdi, glikoz yoğunlukları çıktı olarak kullanılmıştır. 3 yıl sonra Bergman ve diğ., (1982) minimal model yaklaşımını 8 adet erkek ve 10 adet kadın bireylerden oluşan gerçek hastalardan elde edilen verileri kullanarak analiz etmişlerdir.

Lehmann ve Deutch (1992) yaptıkları çalışmada insulin etkileşimi ve karbonhidrat tüketimine göre kandaki glikoz miktarlarının tahmin edilmesini amaçlamışlardır. Çalışmada gerçek klinik verileri kullanılmıştır. Model ile bilgisayar destekli simülasyon aracılığıyla 24 saatlik tahminler yapılmıştır.

Bremer ve Gough (1999) kan şekeri dinamiklerinin rastgele olmadığını ve kan şekeri değerlerinin, en azından yakın gelecekte, sıklıkla örneklendiği durumlarda önceki değerlerden tahmin edilebileceğini göstermiştir. Çalışmada çeşitli glisemi durumlarını temsil eden 22 adet veri seti literatürden elde edilmiştir. Burada hem diyabet hastası kişilerden alınan ölçüm sonuçları, hem de diyabet hastası olmayan kişilerden alınan ölçüm sonuçları kullanılmıştır. Çalışmada yapılan otokorelasyon fonksiyonu tahmini ile, prosesin farklı zamanlardaki ayrı ayrı ölçümleri arasındaki bağımlılığın istatistiksel ölçümü sağlanmıştır. Oluşturulmuş modellerde 0-10 saatlik periyotlarda proses ölçümleri gözlemlenmiş, verilerin ikinci yarısı olarak ifade edilmiş 10-20 saatlik periyotlarda tahminler yapılmıştır.

Hovorka ve diğ. (2004), tip 1 diyabetli hastalarda glikoz yoğunluğunun tahmin edilmesi için lineer olmayan model kullanarak tahmin yapmayı amaçlamışlardır. Oluşturulan nonlineer modelde Bayesian analiz yaklaşımı kullanılmış, modelin ihtiyaçları model parametreleri ve model sabitleri olmak üzere olarak 2 kısıma ayrılmıştır. Çalışma, yapılan deneyler ile doğrulanmış fizyolojik bir model olarak kabul edilmiştir. Bu model sistemi; glikoz alt sistemi, insulin alt sistemi ve insulin etki alt sistemi olmak üzere 3 alt sistemden oluşmuştur. Çalışmada gerçek hastalardan elde edilen 15 adet veri kullanılmıştır. Tahmin aralıkları 15

dakika ile 240 dakika arasında değişmekte olduğu görülmüştür. Performans kriteri olarak hata kareler ortalaması değerleri kullanılmıştır.

Yamaguchi ve diğ. (2005), yaptıkları çalışmada tip 1 diyabet hastalarının kan şekeri düzeylerinin tahmin edilmesinde veri madenciliği üzerinden çalışmışlar ve tepki yüzeyleri metodolojisi kullanmışlardır. Burada tepki yüzey metadolojisi diğer ifadesiyle yanıt yüzey metadolojjisinde açıklayıcı değişkenler ile yanıt değişkenleri arasındaki ilişki kurulmaya çalışılmıştır. Çalışmada 5 ay yapılan gözlemler sonucunda oluşturulan veri setleri ile glikoz düzeyi, metobolik hızı, besin alımı ve fiziksel durum değişkenleri kullanılmıştır.
Islam ve diğ. (2007), yaptıkları çalışmada glikoz yoğunluğunu, insulin miktarlarını ve vücut içerisinde dolaşan kan şekerinin hacminin dağılımını kullanarak matematiksel modeler oluşturmuştur. Tip 1 ve tip 2 diyabet hastalarının vücutlarının besin alımına etkisini doğru bir şekilde tahmin etmeleri ve tehlikeli hipoglisemik ve hiperglisemik durumlardan kaçınmalarına yardımcı olmak çalışmada amaçlanmıştır. Oluşturulmuş algoritma, öğün alımından hemen sonra öngörülen en yüksek kan şekeri seviyesine dayalı olarak insülin dozajını tahmin etmesini amaçlamıştır. İnsan vücudu ile uğraşıldığı için, her organın nasıl davranacağını modellemek oldukça zordur, bu nedenle her bir birey için modeli kalibre etmek için bir dizi parametre toplanmış ve algoritma oluşturulmuştur. Çalışma, matematiksel tahmin modelleri ile hastalardaki kan şekeri üst seviye değerlerini tahmin etmeyi amaçlamıştır.

Cheng ve diğ. (2010), yaptıkları çalışmada kandaki glikoz yoğunluğunun tahmin edilmesinde glikoz, insulin ve vücuta alınan karbonhidrat miktarlarını kullanarak kandaki şeker seviyesi ile ilgili tahmin yapmayı amaçlamışlardır. Yaptıkları ilk çalışmada hasta sayısını 4 olarak belirlemişlerdir. Aynı yıl yaptıkları diğer çalışmada hasta sayısını arttırmışlar, 58 adet hasta üzerinden çalışmalarını oluşturmuşlardır. İlk çalışmada zaman gecikmeli diferansiyel denklem modeli kurulmuş, aynı yıl yapılan sonraki diğer çalışmada modifiye edilmiş diferansiyel denklem modeli kullanılmıştır. Modeller, Matlab programı aracılığıyla çözülmüştür ve hata kareler toplamı performans ölçüm kriterleri olarak 2 çalışmada da kullanılmıştır.

Wu ve diğ. (2011), yaptıkları çalışmada 4 parametreli glikoz-insülin dinamik modelini geliştirmişlerdır. Bu model ile deri altı dokulardaki glikoz ve insulin dinamiklerinin anlaşılması amaçlanmıştır. Bununla birlikte kan şekeri seviyesinin düzenlenmesi

hedeflenmiştir. Yapılan çalışmanın yapay pankreas gelişimi için önemli olduğu sonucuna varılmıştır.

Kirchsteiger ve diğ. (2011), yaptıkları çalışmada tip 1 diyabet hastalarından elde edilmiş gerçek verileri analiz ederek sürekli dağılımlı transfer fonksiyonları kullanmışlardır. Kullanılan transfer fonksiyonu olarak çok girişli tek çıkışlı sistemler tercih edilmiştir. Kan şekeri dinamiklerini modellemek amacıyla elde edilen sonuçlar, yinelemeli tahmin hatası yöntemi aracılığıyla bir parametre tahmin prosedürü sunmuştur.

Calm ve diğ. (2011), yaptıkları çalışmada Hovorka'nın yapmış oldukları çalışmayı geliştirmişlerdir. İki farklı simülasyon yaklaşımı kullanarak yeni çözümler üretmeye çalışmışlardır. Monte Carlo ve aralıklı simülasyon modelleri kurgulanmış ve sonuçlar karşılaştırılmıştır. İnsulin duyarlılıkları, glikozun absorbe edilmesi ve gıda alımı belirsizliklerine karşın, oluşturulan glikoz düzenleyici modelin simülasyonu her 2 yöntem kullanılarak gerçekleştirilmiştir. Aralıklı simülasyon ile elde edilen sonuçlar Monte Carlo simülasyonu ile elde edilen sonuçlara göre daha üstün ve daha az hesaplama süresine sahip olmuştur.

De Pereda ve diğ. (2012), yaptıkları çalışmada tip 1 diyabet hastalarında insulin tedavisine yönelik Calm ve diğerleri gibi Hovorka'nın glikoz-insülin modelini temel alarak yeni model oluşturmuşlardır. Çalışmada tahmin araçlarının geliştirilmesi için parametrik belirsizlik altında glikoz konsantrasyon sınırlarının hesaplanması ele alınmıştır. Biyolojik süreçlerin yüksek değişkenliği sebebiyle model parametrelerinin kesin değerleri yerine aralıklarla sınırlandırma tercih edilmiştir.

Zhao ve diğ. (2012), tip1 diyabet hastalarının gelecekte kandaki glikoz yoğunluğunu tahmin etmek amacıyla yapısal eşitlik modeli tabanlı istatistiksel yöntem modeli ve zaman serileri otoregresif modeli kurmuşlardır. Kurulan yapısal eşitlik modeli tabanlı istatistiksel yöntem modelin, otoregresif modele göre daha yüksek performanslı sonuçlar vermekte olduğu görülmüştür. Çalışmada 15-30-45 ve 60 dakikalık zaman dilimlerinde kan şekeri tahminleri yapmak için 1 günlük zaman diliminde 7 adet gerçek diyabet hastasının verileri kullanılmıştır.

Duun ve diğ. (2013), yaptıkları çalışmada genişletilmiş minimal model yaklaşımını kullanmışlardır. Kullanılan model stokastik diferansiyel denklemler aracılığıyla

oluşturulmuştur. Çalışmada matlab programı aracılığıyla 300 dakikalık süre için tahminleme yapılmıştır. Kullanılan parametreler 4 adet tip 1 diyabet hastasından alınan klinik verilerden elde edilmiştir. Otokorelasyon fonksiyonlarındaki olabilirlik-oran testleri çalışmada performans kriteri olarak kullanılmıştır.

Wang ve diğ. (2014), genişletilmiş Kalman filtresi yaklaşımı kullanarak tip 1 diyabet hastalığı için glikoz dinamiğini modellemişlerdir. Kalman filtresi bir filtreleme algoritmasıdır. Bu algoritmanın sistemin ölçülemeyen durumlarını tahmin etmede oldukça güçlü ve yetenekli olduğu kabul edilmiştir. Çalışmada sürekli izlenen glikoz seviyesi, insulin dozu ve yemek aktiviteleri durumları gözlemlenmiştir. İnsülin dozunun ve öğün alımının kan şekeri dalgalanmaları üzerindeki etkisi farklı hastalar üzerinden durum uzay modeli oluşturularak incelenmiştir. Çalışmada 30 sanal ve 5 gerçek hastası verisi kullanılmıştır. Model parametreleri gerçek zamanlı olarak tahmin edildiğinden modelin kontrolünün oldukça uygun olduğu sonucuna varılmıştır.

Balakrishnan ve diğ. (2014), Bergman'ın minimal modelini geliştirmişlerdir. Matlab programı aracılığıyla gerçek 34 adet tip 1 diyabet hastasından aldıkları verileri girdi değişkeni olarak kullanmışlardır. Çalışmada 3 aşamalı metodoloji ile kandaki glikoz verileri, insulin dozajı, yemek ve spor aktiviteleri göz önüne alınarak tahminleme yapılmıştır. Algılanan efor oranı, egzersiz yoğunluğunu ölçen bir belirteç olarak modelde tanıtılmıştır. Yeni modelde 16 parametreden 6 tanesi tahmin edilebilir parametre olarak tanımlanış ve kişiselleştirilmiş parametreler olarak her denek için tahmin yapılmıştır. Hem uygunluk hem de doğrulama için performans kriteri olarak korelasyon katsayısı kullanılmıştır.

Efendic ve diğ. (2014), tip 1 diyabet hastaları için kandaki glikoz yoğunluğunu kısa periyotlar olarak 10-30 dakikalık süreçler için tahmin etmeye çalışmışlardır. 12 adet gerçek hastanın 7 günlük verileri Gaussian Karışım Modelleri kullanılarak analiz edilmiştir. Performans değerlendirme aracı olarak tahmin yüzdelerinin başarısı kullanılmıştır. Bu çalışmalar Tablo 2.3.'de özetlenmiştir.

Laguna ve diğ. (2014), Hovorka'nın çalışmasını temel alarak yeni bir çalışma yapmışlardır. Çalışma tip 1 diyabetli hastalar için bireyselleştirilmiştir. Kişiye özel yapay pankreas ve insulin tedavisi için model tabanlı stratejiler geliştirilmiştir. Bununla birlikte hastalarda meydana gelen yüksek düzeyde glisemik değişkenlik, güvenilir modellerin tanımlanmasını

engeller. Bu sorunu çözmek için çok amaçlı optimizasyon yapılmış ve hatalar minimize edilmeye çalışılmıştır. Çalışmada önerilen yöntemin uygulanabilir olduğu ve referans ölçüm sonuçlarına göre 5 saatlik bir zaman diliminde iyi bir tahmin yeteneği olduğu belirtilmiştir.

Hidalgo ve diğ. (2014), evrimsel algoritmalardan genetik algoritmanın geliştirilmiş modelini çalışmada kan şekeri tahmini amacıyla kullanmışlardır. Modeller, 5 farklı hastadan elde edilen veriler kullanılarak oluşturulmuştur. Çalışmada kulanılan gelişmiş genetik programlama modelleri tek nokta çaprazlamalı ve nokta mutasyonlu olarak düzenlenmiştir. Çalışmada 4 farklı genetik algoritması modeli oluşturulmuş ve performansı en yüksek model belirlenmiştir.

Fang ve diğ. (2015), tip 1 diyabet hastalarında kandaki glikoz seviyesini tahmin etmek için Cobelli tarafından geliştirilmiş modeli temel alarak yeni bir insulin-glikoz metabolik çözüm modeli ortaya koymuşlardır. Yetişkinler, ergenler ve çocuklar olmak üzere 3 grup sanal hastaya dayanan çözüm modeli indirgeme yöntemi kullanılarak elde edilmiştir. Bu yöntemde en iyi çözüme ulaşmak için model parametreleri indirgenmiştir ve çalışmada daha az parametre kullanılarak da doğru tahminler yapıldığı gösterilmiştir.

Zhang ve diğ. (2015), kandaki glikoz yoğunluğunu tahmin etmede veriye dayalı lineer olmayan stokastik bir model ortaya koymuşlardır. Sürekli glikoz ölçüm sistemleri verilerini kullanarak kan şekeri konsantrasyonunun gıda alımına tepkisini açıklamak için stokastik doğrusal olmayan ikinci dereceden diferansiyel denklem geliştirmişlerdir. Glikoz zaman serileri için dinamik sistem denklemleri geliştirmek ve sistem parametrelerini belirlemek için Bayes yöntemi çalışmada uygulanmıştır.

Bock ve diğ. (2015), tip 1 diyabetli hastaların tedavisi için kan şekeri konsantrasyonlarının tahmin edilmesi sorununu ele almışlardır. Sadece kan şekeri ölçümlerini kullanarak model parametrelerini belirleyebilmek, kan glikoz dinamiğini mükemmel bir şekilde modellemenin neredeyse imkansız olduğu görüşü yaygın olarak kabul edilmiştir. Çalışmada standart statik tahmin modeline dayanan basit ve tanımlanabilir bir lineer dinamik model önerilmiştir. Önerilen dinamik modelin doğrulanması, simulatöründen gelen veriler ve gerçek klinik veriler kullanılarak gerçekleştirilmiştir.

Zhou ve diğ. (2020), kan şekeri seviyesini tahmin etmede, kalp atış hızı ile ilgili daha evvel yapmış oldukları çalışmada kurmuş oldukları modeli güncellemişlerdir. Bu çalışmada optik görüntüler kullanarak veri tabanlı bir model ortaya koymuşlardır. Veriler 125 ölçüm sonucunda, 1 erkek ile 2 kadın olmak üzere 3 adet sağlıklı gönüllü kişi ve 3 erkek ile 2 kadın olmak üzere 5 adet diyabet hastalarından toplanmıştır. Çalışmada fotopletismografi aracılığıyla vücutta cilde yakın bir test bölgesinde akan kanın hacmi hakkında bilgi veren elektro-optik yönteminden yararlanarak ve Gauss fonksiyonu kullanarak tahmin modeli oluşturulmuştur.

Tablo 2.3: Matematiksel modeller kurularak yapılan çalışmaların analizi.

Yazarlar	Yıl	Tahmin Zamanı	Yazılım	Veriler (Hasta Sayısı)	Gözlem Zamanı	Girdi Değişkenleri				Performans Ölçüm Kriterleri
						Glikoz	İnsülin Dozaj	Karbon-hidrat	Diğer	
Bergman ve diğ.	1979	120 dakikaya kadar	-	5 Kırma Cins Köpek	-	✓	✓	-	-	Hate kareleri toplamı, uyumluluk (FIT) değeri
Bergman ve diğ.	1982	-	-	18 Sağlıklı Birey	-	✓	✓	-	-	Hate kareleri toplamı, FIT değeri

Tablo 2.3 (devam)

Yazarlar	Yıl	Tahmin Zamanı	Yazılım	Veriler (Hasta Sayısı)	Gözlem Zamanı	Glikoz	İnsülin Dozaj	Karbon-hidrat	Diğer	Performans Ölçüm Kriterleri
Lehmann ve Deutch	1992	24 saatlik değişim	-	486 veri	10 saat	✓	✓	✓	-	Tahmin hataları minimizasyonu
Bremer ve Gough	1999	10 dk.	-	-	40 saat	✓	-	-	-	Minimum hata oranı
Hovorka ve diğ.	2004	-	Florance	15	6 Hafta	✓	✓	✓	-	Hata kareleri ortalaması
Yamaguchi ve diğ.	2005	1 saat	Matlab	4	5 Ay	✓	✓	✓	✓	Korelasyon katsayısı
Islam ve diğ.	2007	Yemek'ten hemen sonrası	-	-	-	✓	✓	✓	-	Tahmin başarı performansı
Cheng ve diğ.	2010	-	Matlab	4	1 gün	✓	✓	✓	-	Hata kareler Toplamı
Cheng ve diğ.	2010	-	Matlab	56	1 gün	✓	✓	✓	-	Hata kareler Toplamı
Wu ve diğ.	2011	-	-	3	-	✓	✓	✓	-	Hata kareleri toplamı
Kirchsteiger ve diğ.	2011	-	-	15	3 gün	✓	✓	✓	-	-
Calm ve diğ.	2011	400 dk.	C++	1	6 saat	✓	✓	✓	-	Hata kareleri toplamı
De Pereda ve diğ.	2012	300 dk.	Matlab	1	5 saat	✓	✓	✓	-	Simülasyon sonuçları
Zhao ve diğ.	2012	15,30,45,60 dk.	Matlab	7	1 Gün	✓	-	-	-	Hata kareleri ortalaması, hata ızgara analizi ve mutlak sapma
Duun ve diğ.	2013	10 dk.	-	4	4 gün	✓	✓	✓	-	Otokorelasyon fonksiyonu (ACF) olasılık-oran testleri

Wang ve diğ.	2014	30 dk.	Matlab	30 Sanal/ 5 Gerçek	1 Gün	✓	✓	✓	-	R2 katsayısı, mutlak fark ve J index değeri
Balakrishnan ve diğ.	2014	-	Matlab	34	-	✓	✓	✓	✓	R2 katsayısı
Laguna ve diğ.	2014	300 dk.	Matlab	12	20 saat	✓	✓	✓	-	Ortalama mutlak bağıl fark
Efendic ve diğ.	2014	10-30 dk.	-	12	7 Gün	✓	✓	✓	-	Doğru ve yanlış tahmin oranları
Hidalgo ve diğ.	2014	-	Java	5	24 saat	✓	✓	✓	-	Ortalama mutlak sapma, ortalama hata kareleri toplamı kökü, ortalama hata, maximum hata
Fang ve diğ.	2015	-	-	30	1 gün	✓	✓	✓	-	Ortalama hata kareleri toplamı kökü
Zhang ve diğ.	2015	-	-	4	72 Saat	✓	-	-	-	Uyumluluk (FIT) değeri
Bock ve diğ.	2015	15- 165 dk.	Matlab/ Simulink	10 simüle (sanal), 12 gerçek	4 gün/10 gün	✓	✓	✓	-	Ortalama mutlak fark, R2
Zhou ve diğ.	2020	-	Matlab	8	214 ölçüm	✓	✓	✓	-	Ortalama hata kareleri toplamı kökü, R2 ve Clarke glikoz hata analizi

2.5.3. Zaman Serileri Modelleri Kullanılarak Yapılan Diyabet Tahmini Çalışmaları

Sparacino ve diğ. (2006), birinci mertebeden bir polinom modelinin tahmin edicisinin performansı ile otoregresif modelin performanslarını karşılaştırmıştır. Modellerin her ikisi içinde girdi değişkenleri olarak geçmişteki glikoz verileri kullanılmıştır. Uygulama 48 saatlik süre için izlenen 13 adet diyabetik gönüllüden elde edilen veriler kullanarak tamamlanmıştır. Çalışma da glikozun, sürekli glikoz izleme (CGM) tarafından sağlanan geçmiş değerlere karşı oluşturulmuş bir matematiksel modelden yararlanarak ileride öngörülebilir olduğunu göstermeye çalışmışlardır.

Reifman ve diğ. (2007), veriye dayalı otoregresif modeller (zaman serileri) kullanarak glikoz seviyesi tahmini yapmıştır. Çalışmada, 10. dereceden zaman serisi modeli kullanılmıştır. Verilerin eğitimi ve değerlendirilmesinde i-sense CGM (i-sense Şirketi-Portland) sürekli glikoz ölçüm sistemleri kullanmışlardır. Kurulan modelin sonuçları ile CGM ölçümleri birbirlerine yakın sonuçlar vermiştir.

Lu ve diğ. (2010), yaptıkları çalışmada kullandıkları veriler, sürekli glikoz izleme cihazları kullanılarak 5 günlük süre boyunca 9 hastadan toplanan dakika dakika glikoz sinyallerinden oluşmuştur. Çalışma otoregresif hareketli ortalama modeli aracılığıyla, glikoz sinyalinin

frekans bantlarının tekli ve ikili kombinasyonları kullanılarak geliştirilmiş ve tüm bantları içeren bir referans model ile karşılaştırılmıştır. 60-120 dakika ve 150-500 dakika arası periyotlar kullanılarak yapılan tekli frekans zaman serisi modelleri ile 25 dakikaya kadar tahminleme yapılmıştır. Bununla birlikte 60-120 dakikalık periyot aralığı için de ikili kombinasyon kullanarak 300 dakikaya kadar başarılı tahminleme yapıldığı belirtilmiştir.

Gani ve diğ. (2011), Matlab programı aracılığıyla 5-56 gün boyunca izlenen 34 adet hastanın glikoz verilerini kullanarak kurdukları otoregresif zaman serisi modeli ile kandaki glikoz yoğunluğunu tahmin etmeyi amaçlamışlardır. Çalışmanın performans aracı olarak ortalama hata kareleri kökü değerleri kullanılmıştır.

Boiroux ve diğ. (2012), zaman serilerinde kullanılan doğrusal olmayan durağan zaman serisi modeli (ARIMAX) tip 1 diyabetli kişilerde kan şekerinin gece boyunca kontrol altına alınması için yapmış oldukları çalışmalarda kullanmıştır. Kontrol için oluşturulan model, sürekli glikoz ölçümlerine göre hesaplamalar yapmıştır. Çalışmada, rastgele seçilen 100 adet hasta bilgileri kullanılarak parametreler düzenlenmiştir. 1 gecelik yapılan simülasyondan sonraki aşamada da gerçek klinikten elde edilen yeni hasta bilgileri ile çalışma sonuçları sunulmuştur.

Harsh ve diğ. (2013), tip 1 diyabet hastalarının kanlarındaki glikoz seviyelerini tahmin etmek için insulin dozu kullanım miktarı ve öğünlerdeki yemek durumlarına göre çözüm yöntemi geliştirmişlerdir. Oluşturulmuş deneysel model, 30 hasta ile Amerikan Gıda ve İlaç Dairesi (FDA) onaylı simulatör kullanılarak değerlendirilmiştir. Deneklerin kan şekeri seviyesi, insulin dozu ve karbonhidrat alımından oluşan bu 3 duruma göre zaman serileri ve Karman filtreleme tekniği uygulanmıştır. Önerilmiş model yapay pankreas aracılığıyla kan şekeri seviyesini tahminleme yapmış ve kan şekerini kontrol altında tutmayı amaçlamıştır.

Schmidt ve diğ. (2013), doğrusal olmayan durağan zaman serisi modeli (ARIMAX) aracılığıyla 10 saatlik bir tahmin süresi için çalışmalarını ortaya koymuşlardır. Yaptıkları çalışmada 12 adet hastadan, 2 gece boyunca toplamış oldukları verileri kullanmışlardır. Oluşturulan çözüm ile tip 1 diyabet hastalarında model tabanlı kapalı çevrim kontrol sistemi kurarak tahminleme yapılmıştır.

Türksoy ve diğ. (2013), yapay pankreas sistemleri için çok değişkenli uyarlanabilir tanımlama ve kontrol özelliklerine sahip bir model geliştirmişlerdir. Çalışmada deneklerin vücutlarına takılan kol bantları ile elde edilen fizyolojik bilgilerin kullanılması, glikoz konsantrasyonu tahminini iyileştirmiş ve fiziksel aktivitenin glikoz konsantrasyonu üzerindeki etkilerinin daha erken tanınmasını sağlamıştır. Glikoz tahmin hatası kareleri toplamı ve ortalama kare kök sapması, yanlış alarm oranı (FAR) ve duyarlılık analizi yapılarak çalışma sonuçları analiz edilmiştir.

Peng ve diğ. (2014), zaman serileri analizinde otoregresif hareketli ortalama tekniklerinden otoregresif zaman serisi modeli aracılığıyla tip 1 diyabet hastası 2 bireyin 10 günlük verisini kullanarak tahmin modeli oluşturmaya çalışmışlardır. Modelde, girdi değişkenleri olarak glikoz verileri kullanılmıştır ve performans kriteri olarak da uygunluk (fit) değerleri kullanılmıştır.

Botwey ve diğ. (2014), yaptıkları çalışmada tip 1 diyabetli hastalarda kan şekeri değerlerinin doğru tahmin edilmesini sağlayarak, hipoglisemik ve hiperglisemik olayların erken vakitte uyarılarını sağlamak ve böylece hastanın güvenliğini arttırmayı amaçlamışlardır. Bu soruna yönelik farklı çözüm modelleri geliştimişlerdir. Geliştirilen çözüm modellerinden birisi de çıktı düzeltmeli otoregresif modeller olarak görülmüştür.

Novara ve diğ. (2015), yaptıkları çalışmada yeni tasarladıkları yöntemler ile beraber 5. Dereceden doğrusal olmayan durağan zaman serisi modeli (ARIMAX) uygulamışlardır. Esas önerdikleri yöntem çalışmada yeni kara-kutu modeli olarak adlandırılmıştır. Bu yaklaşım ile hasta sistemine giren tüm bilinmeyen veya kolayca ölçümü yapılamayan girdilerin etkilerini gerçekçi olacak şekilde hesaba katarak hasta sisteminin davranışını simüle etmeyi amaçlamışlardır.

Zhao ve diğ. (2015), Tip 1 diyabetli hastalar için online cilt altı glikoz konsantrasyonu tahminini hızlı olarak yapmak için model kurmuşlardır. Kurulan model otoregresif zaman serisi modeli özelliğini taşımıştır. Önerilen yöntem, özellikle modelleme verilerinin eksikliği için tekrarlayan konu bağımlı modelleme yöntemi yerine, etkili ve ekonomik bir modelleme yöntemi olarak kabul edilmiştir.

Timar ve diğ. (2017), tip 1 diyabet hastalarının kandaki glikoz yoğunluğunu tahmin etmek için lineer olmayan kaotik özelliklerin kullanıldığı karar destek sistemi tasarlamışlardır. Burada 30 dakika ve 60 dakika sonraki zamanların tahminlerinde bulunmak için kaotik zaman serileri tercih edilmiştir. Çalışmada hızlı ve doğru olarak tahminlerin elde edilmesi için, dışarıdan gelen bilgiyi ölçme yükü olmadan kullanarak, doğrusal olmayan rejim değiştirmeli zaman serisi modelleri kullanılmıştır.

Xie ve Wang (2020) makine öğrenme algoritmaları ile beraber klasik zaman serileri modellerini tip 1 diyabet hastalarında kandaki glikoz seviyesi tahmini için kullanmışlardır. Kullanılan zaman serisi modeli otoregresif hareketli ortalamalar modeli olarak tercih edilmiştir. Hava sıcaklığı, bazal insülin hızı, kan şekeri seviyesi, bolus insülin hızı, egzersiz yoğunluğu, parmak-ölçüm kan şekeri seviyesi, karbonhidrat alımı, cilt sıcaklığı, glomerüler filtrasyon hızı, kalp hızı dahil olmak üzere her bir deneğin ayrıntılı giriş ve çıkış zaman serileri verileri kullanılmıştır. Çalışmanın sonucunda kullanılan çözüm yöntemleri ve elde edilen sonuçlar incelendiğinde, çözüm modelleri arasında önemli bir farklılık olmadığı görülmüştür. Bu çalışma sayesinde, araştırmacıların ve klinik pratisyenlerin kan şekeri tahmini için en uygun modelleri seçerek tahmin yapabileceği belirtilmiştir.

Tablo 2.4: Zaman serileri kullanılarak yapılan çalışmaların analizi.

Yazarlar	Yıl	Tahmin Zamanı	Yazılım	Veriler (Hasta Sayısı)	Gözlem Zamanı	Girdi Değişkenleri				Performans Ölçüm Kriterleri
						Glikoz	İnsülin Dozaj	Karbon-hidrat	Diğer	
Sparacino ve diğ.	2006	20 dk. Sonrası için	-	13	48 saat	✓	-	-	-	Ortalama gecikmeler
Reifman ve diğ.	2007	30 dk.	isense/ Guardion/ Dexcom	9/7/18	-	✓	-	-	-	Ortalama hata kareleri toplamı kökü ve clarke hata tablosu
Lu ve diğ.	2010	50 dakikaya kadar	Matlab	9	5 gün	✓	-	-	-	Ortalama hata kareleri toplamı kökü
Gani ve diğ.	2011	30 dk.	Matlab	34	5-56 Gün	✓	-	-	-	Ortalama hata kareleri toplamı kökü
Boiroux ve diğ.	2012	10 saat	Matlab	12	2 Gece	✓	✓	✓	-	-
Harsh ve diğ.	2013	30 dk.	Matlab	30	1 Gün	✓	✓	✓	-	R2, uyumluluk (FIT) değeri, mutlak fark
Schmidt ve diğ.	2013	10 saat	Labview	12	2 Gece	✓	✓	✓	-	-
Turksoy ve diğ.	2013	30 dk.	Matlab	14	48-60 saat	✓	-	-	-	Glikoz tahmin hatalarının karesinin toplamı, duyarlılık, yanlış alarm oranı, ortalama hata

										kareleri toplamı kökü
Peng ve diğ.	2014	30 dk.	Matlab	2	10 Gün	✓	-	-	-	Uyumluluk (FIT) değeri
Botwey ve diğ.	2014	15, 30 ve 45 dakika	-	23	-	✓	✓	-	-	Hata kareler ortalaması, gecikme süresi ve korelasyon katsayısı
Novara ve diğ.	2015	30 dk.	Matlab	5	10 Gün	✓	-	-	-	Uyumluluk (FIT) değeri
Zhao ve diğ.	2015	30 dk. ve 60 dk.	-	30	6 Gün	✓	✓	✓	-	Ortalama hata kareleri toplamı kökü, özgüllük ve duyarlılık testleri
Timar ve diğ.	2017	30 dk. ve 60 dk.	R Proje	17	-	✓	✓	✓	✓	Ortalama hata kareleri toplamı kökü ve ortalma mutlak hata
Xie ve diğ.	2020	5 dk. ve 30 dk.	-	6	8 Hafta	✓	✓	✓	✓	Ortalama hata kareleri toplamı kökü, zaman tasarrufu ve risk tahmini

2.5.4. Karar Ağacı Algoritmaları Kullanılarak Yapılan Diyabet Tahmini Çalışmaları

Eljil ve diğ. (2013), yaptıkları çalışmada diyabetli hastalarda meydana gelebilecek 2 önemli sağlık sorunundan biri olan, hipoglisemi sorununa çözüm getirmeye çalışmışlardır. Hipoglisemi diğer bir ifadeyle ani ve hızlıca kan şekerinin yükselmesi problemi, diyabet türü fark etmeksizin şeker hastalarının başına gelebilecek ciddi sonuçlar doğuran bir durumdur. Bu noktada karar ağacı temel algoritmalarıyla beraber torbalama yöntemi de kullanılmasıyla, hipoglisemi meydana gelmeden hipoglisemiyi tahmin etmek amaçlanmıştır. Her hasta için ayrı ayrı analizler yapılmış ve tahmin modelleri oluşturulmuştur.

Fong ve diğ. (2013), karar ağacı algoritmaları ve yapay sinir ağları algoritmaları kullanarak kandaki glikoz yoğunluklarını tahmin etmeye çalışmışlardır. Çalışmada 70 adet hastadan elde edilen kan şekeri ölçüm değerleri, fiziksel aktivite durumları, açlık-tokluk durumları gibi değişken durumlar analiz edilerek çeşitli yöntemlerle sonuçlar karşılaştırılmıştır. Yapay sinir ağları sınıflandırıcısının etkinlik verilerinin yüksek performans göstermediği buna karşın karar ağacı algoritmaları kullanarak elde edilen sonuçların daha yüksek doğruluğa sahip olduğu görülmüştür.

Kondiloğlu (2015) yaptığı çalışmada kullandığı veri seti içerisinde yer alan temel bilgilerden yemek durumu, fiziksel aktivite durumu ve kandaki glikoz yoğunluğu bilgilerini düzenleyerek 71 adet farklı girdi değişken parametresi belirlemiştir. Çalışmada; yapay sinir ağları, karar

ağacı algoritmaları, destek vektör makineleri ve regresyon yöntemleri sırasıyla uygulanmıştır. Performans kriteri olarak kullanılan korelasyon katsayısına göre en yüksek korelasyon ilişkisini ortaya koyan yöntem olarak karar ağacı algoritmalarının olduğu gözlemlenmiştir.

Lai ve diğ. (2017), yaşları 18-90 arasında olan Kanadalı hastalardan elde edilmiş 13.309 adet kayıt verilerini kullanarak makine öğrenimi teknikleri aracılığıyla çalışmalarını ortaya koymuşlardır. Çalışmada karar ağaçları ile beraber rastgele orman algoritmaları, gradyan arttırma algoritmaları ve lojistik regresyon yöntemleri de kullanılmıştır. Çalışma sonucunda karar ağaçları ve rastgele orman algoritmaları kullanılarak yapılan çözümlerin diğer iki yöntem kullanılarak yapılan çözümlere göre daha düşük performans gösterdiği sonucuna varılmıştır.

Alfian ve diğ. (2020), yapay zeka teknolojilerini kullanarak zamana bağlı olarak kan şekeri tahmin modelleri geliştirmişlerdir. Çalışmada karar ağaçları ile beraber destek vektör makinaları, k en yakın komşu algoritması, rastgele orman algoritması, Ada Boost yapay sinir ağları ve ekstrem gradyan arttırma (XGBoost) uygulanmıştır. Bu yöntemler ayrı ayrı veri setlerine uygulanmıştır. Çalışmada, 15-30-45-60 dakikalık süreler için tahminler yapılmaya çalışılmıştır. Önerilen tahmin modellerinin uygulanması, hastaların gelecekteki kan şekeri seviyelerininin elde edilmesine izin vermiş, böylece önleyici tedbirler sayesinde kritik hipoglisemik / hiperglisemik olaylar meydana gelmeden önce önlem alınabilmiştir. Çözüm sonuçları birbirine yakın sonuçlar üretmekle beraber, ysa ile elde edilen sonuçlar daha yüksek performans göstermiştir.

2.5.5. Mobil Uygulamalar Oluşturarak Yapılan Diyabet Tahmini Çalışmaları

Farmer ve diğ. (2005), İngiltere de, cep telefonları aracılığıyla veri aktarımına dayanmakta olan yeni bir gerçek zamanlı tele-tıp sistemi tanımlamışlardır. Tele-tıp sistemleri, diyabetli kişilerin kendi kendine yönetimini desteklemenin bir yolu olarak önerilmiştir. Bu model, cep telefonundan veri iletimi ile geri bildirimine dayanan yenilikçi bir gerçek zamanlı tele-tıp sisteminin geliştirilmesi ve uygulanması adımlarını içermiştir. Kan şekeri ölçüm cihazını cep telefonuna bağlamak için Java tabanlı programlama dilinden yardım alınmıştır. Kan şekeri verilerinin anında iletilmesine ek olarak, insülin dozu, yeme alışkanlıkları ve fiziksel egzersiz hakkında bilgiler toplanmıştır. Çalışmada, insülin dozunun kontrolü ile diyet ve egzersize yanıt olarak değiştirilme derecesi hakkında veri sağlamak için verilerin daha ayrıntılı grafiksel

gösterimleri kullanılmıştır. Sistem, diğer basamakta yürütülen bir klinik araştırmada değerlendirilmiştir ve sonrasında hasta, pratisyen hekim ve eczacı için entegre yönetim sağlamadaki etkinliğini değerlendirmek üzere tasarlanmış bir denemede kullanılmak üzere uyarlanmıştır.

Kim ve Jeong (2007), tip 2 diyabet hastalarının tedavi süreçlerine yönelik mevcut teknolojiye uyumlu bir çalışma yapmışlardır. Bu çalışmada diyabet hastalarının 6 aylık süreleri boyunca verileri takip edilmiştir. 6 aylık sürede, kandaki glikoz konsantrayonu, beslenme ve fiziksel aktivite bilgileri hergün internet ve cep telefonu aracılığıyla alınmıştır. Bu bilgiler ışığında doktor-hemşire iş birliği ile kan şekerinin istenilen seviede tutulmasını sağlayacak şeklide hastalara geri besleme mantığı ile kısa mesajlar gönderilmiştir. Sonuç bağlamında, hastaların Hemoglabin A1C seviyelerinde 6 aylık süre boyunca önemli düzeyde düşüşler sağlanmıştır. 6 aylık çalışma dönemi 2 dönem olarak ayrılmış. 1. Dönem olarak tasvir edilen ilk 3 aylık dönemde Hemoglabin A1C seviyesinde % 1,2'lik oranda düşme başarısı sağlanmıştır. İkinci 3 aylık dönemde Hemoglabin A1C seviyesinde % 1,1'lik oranda düşme başarısı sağlanmıştır.

Georga ve diğ. (2009), yaptıkları çalışmada diyabet hastaları ile ilgili topladıkları verileri kullanarak, hastaları tedavi süreci ile ilgili bilgilendirme yapacak bir geri dönüş sistemi kurmuşlardır. Kandaki glikoz yoğunluğunu ölçen araçlar, mobil iletişim cihazları, yazılımlar ve internet tabanlı programlar kullanılarak bir bilgi tabanı elde edilmiştir. Analiz işlemleri hastaya özgü olarak, belirlenmiş yaş aralığında kadın ve erkek hastalar için yapılmıştır. Çalışmada yakın zamanda kaydedilen veriler göz önüne alınarak hastaların metobolizmaları modellenmiş, hipoglisemik ve hiperglisemik olayları tahmin etmek amaçlanmış ve hastaya kısa ve uzun vadeli uyarılar sağlanmış, karar destek sistemleri oluşturulmuştur.

Chemlal ve diğ. (2011), tip 2 şeker hastaları için geliştirilmiş özel bir yazılım kullanarak hastalarda kan şekeri tahmin etmeyi amaçlamışlardır. Oluşturulmuş yazılımda parametrik olmayan regresyon yöntemi ile tahminlemeler yapılmıştır. Bu noktada hastaların kan şekeri ve fiziksel aktivite verilerini kullanarak geliştirilen yazılım, cep telefonları aracılığıyla kolayca kullanılabilmiştir. Uygulama hasta uyumunu arttırmayı amaçlayan ve kan şekeri regülasyonunu iyileştirmeye teşvik etmeyi amaçlayan bir ödül bileşenli arayüze sahip olarak düşünülmüştür. Model fiziksel aktiviteleri tanımlama ve sınıflandırma özelliklerini de kullanmıştır.

Li ve Fernando (2016) kişiselleştirmiş tahminlerde bulunmak için akıllı telefonlardan toplanan hasta verilerini kullanan diyabet tahmin mekanizması kurulmasını amaçlamışlardır. Bu model daha önceden oluşturulmuş populasyon havuzundaki bilgileri kullanarak hasta benzerliklerine göre bireyler için kan şekeri tahmini yapmayı amaçlamıştır. Önerilen model regreyon yöntemlerini kullanarak, bir mobil uygulama olarak hayata geçirilmiş ve performans değerlendirilmesi için kapsamlı deneyler gerçekleştirilmiştir. Yeni deneyler ile tahmin doğruluğunun iyileştirilebileceği ve mevcut yaklaşımlardaki az veri ile çalışma sorununun da önüne geçilebileceği sonuçları ortaya konulmuştur.

Chowdhury ve diğ. (2018), kan şekeri tahmininde akıllı telefonların kulllanıldığı video tabanlı bir yaklaşım önermiştir. Akıllı telefon kamerası kullanılarak diyabet hastası kişilerin parmak uçları çekilmiş, ardından fotopletismografi (PPG) dalga biçimine dönüştürülmüştür. Sonrasında, gauss filtresi tabanlı asimetrik en küçük kareler yöntemleri uygulanmış ve sinyal noktalarına göre tepe noktaları belirlenmiştir. Çalışmada son olarak elde edilen özellikler sayesinde temel bileşenler regresyon analizi kullanılarak glikoz seviyeleri tahmin edilmiş ve tahmin hataları belirlenmiştir.

2.5.6. Bulanık Mantık Uygulamaları Kullanılarak Yapılan Diyabet Tahmini Çalışmaları

Fong ve diğ. (2013), yaptıkları çalışmada bulanık mantık uygulamalarından bulanık kafes akıl yürütme yöntemini kullanarak kandaki glikoz seviyelerini tahmin etmek için genel bir yaklaşım önermiştir. Çalışmada uygulanan yöntem bulanık kafes modeli ile uzmanlardan bilgi edinimini ve kural tabanı oluşturarak doğruluğu yüksek oranda kan şekeri düzeyini tahmin etmeyi amaçlamıştır. Çalışmada 1 adet hastanın verileri Weka programı aracılığıyla analiz edilmiştir. Girdi değişkeni olarak glikoz ve insulin doz değerleri kullanılmıştır. Modelin doğruluğu, Cohen Kappa katsayısı aracılığıyla kontrol edilmiştir. Modelin etkinliği yeni deneyler ile kontrol edilmiştir. Diyabet veri setlerine bulanık mantık uygulayan ilk ve öncü çalışma olduğu yazarlar tarafından çalışmada belirtilmiştir.

Zarkogianni ve diğ. (2014), dalgacık aktivasyon fonksiyonları kullanarak nöro-bulanık model aracılığıyla kandaki glikoz miktarını tahmin etmeyi amaçlamışlardır. Model gerçek zamanlı glikoz seviyelerini ve fiziksel aktiviteyi kaydeden sensörlerden girdi olarak veriyi almıştır. Önerilen model bulanık mantık kural tabanlarına göre sonuçlar çıkartmıştır. Çalışmada 6 adet gerçek hastanın verileri Matlab programı kullanılarak analiz edilmiştir. Tahminler tip diyabet

hastaları için 30 dk, 45 dk., ve 60 dk.'lık süreler için yapılmıştır. Performans kriterleri olarak hata kareleri ortalaması, korelasyon katsayısı, hata analizleri ve ortalama mutlak bağıl fark değerleri kullanılmıştır.

Çanayaz (2014) tip 1 diyabet hastalarının tedavilerine yönelik kapalı döngü kontrol sistemi üzerinde çalışmıştır. Bu sistem hasta özelinde, 3 adet ardışık kan şekeri ölçüm sonuçlarını karşılaştırarak, kan şekeri seviyesini olması gereken düzeyde sağlayacak insulin miktarını hesaplamıştır. Bu model, gönüllü tip 1 diyabet hastası aracılığıyla test edilmiştir. Sonrasında hastanın kullanması gereken insulin ile sistemin önerdiği insulin miktarı karşılaştırılmıştır. Burada kullanılan çözüm modeli olarak bulanık mantık bulanık çıkarım sistemi (FIS) editörü kullanılmıştır.

Soylu ve Danışman (2017) literatürde yer alan 3 farklı yöntemi benzetim aracılığıyla karşılaştırmıştır. Analiz edilen yaklaşımlar; Bergman'ın minimal modeli, Hovorka modeli ve UVa/Padova modelleri olarak tercih edilmiştir. Karşılaştırma sırasında bulanık mantık temelli kapalı döngü kontrol sistemi kullanılmıştır. Kullanılan yöntem iki giriş ve tek çıkıştan oluşan Mamdani tipi bulanık çıkarım sistemi olarak belirlenmiştir. Giriş değişkenleri için 3'er adet Gaussian üyelik fonksiyonu ve çıkış değişkeni için 5 adet Gaussian üyelik fonksiyonu tanımlanmıştır.

Farahmand ve diğ. (2019), tip 1 diyabet hastalarının glikoz-insülin seviyelerini düzenlemek için bir kontrol sistemi kurmayı amaçlamışlardır. Çalışmada önerilen yaklaşımda, bir Takagi-Sugeno (TS) bulanık modelleme tabanlı paralel bağlanmış kompanzasyon (PDC) ve paralel olmayan dağıtılmış kompanzasyon (PDC olmayan) şemaları kullanılmıştır. Çalışmada karesel olmayan Lyapunov fonksiyonu kullanılarak sistemin kritik noktaları belirlenmiştir. Diyabet modelleri olarak, tip 1 diyabet hastalarında glikoz-insülin sürecini tanımlayan doğrusal olmayan minimal Bergman ve Tolic modellerinden yararlanılmıştır. Çalışma sonucunda tasarlanan sistem, hipoglisemi ve hiperglisemi bozukluklarından kaçınmada ve istenen bölgede kan glikoz konsantrasyonunun dengesini sağlamada başarılı sonuçlar vermiştir.

McShinsky ve Marshall (2020) şeker hastalarının uzun vadeli olarak sağlıklarını kontrol altında tutmalarına yardımcı olmak için gelecekteki kan şekeri seviyelerinin yanı sıra hem hipoglisemik olayları hem de hiperglisemik olayları tespit etmenin etkinliğini analiz etmişlerdir. Bu sebepten ötürü bir çok farklı yöntem çalışmalarında kullanmışlardır.

Kullandıkları yöntemlerden birisi de bulanık mantık ve yapay sinir ağlarını içeren uyarlamalı ağ tabanlı bulanık çıkarım (ANFIS) sistemi olarak düşünülmüştür. Çalışmanın sonucunda tüm yöntemler ile beraber bulanık çıkarım yönteminin sonuçları da karşılaştırılmıştır.

2.5.7. Yapay Sinir Ağları Kullanılarak Yapılan Diyabet Tahmini Çalışmaları

Sandham ve diğ. (1998), girdi değişkeni olarak insulin dozaj miktarlarını kullanarak kandaki glikoz seviyesinin değerini tahmin etmeye çalışmışlardır. Çalışmada ardışık bilgileri kullanmak için tekrarlayan yapay sinir ağları tercih edilmiştir. Tekrarlayan yapay sinir ağlarında diziler aynı görevi yerine getirmiş ve çıktı önceki hesaplamlara bağlı olarak geliştirilmiştir. Ayrıca egzersiz, stress, hastalık ve ameliyat gibi diğer bilgileri de dikkate almışlardır. Yapılan çalışmada 2 adet diyabet hastası kadının verilerini kullanılmıştır.

Gandia ve diğ. (2007), tarafından yapılan çalışmada yapay sinir ağları kullanılarak tahmin edici model geliştirilmiştir. Oluşturulan ağ mimarisinde ilk katman 10 nöron ve ikinci katman 5 nörondan oluşmuştur. Transfer fonksiyonları olarak sigmoid fonksiyonları tercih edilmiştir. Kullanılan ağ modeli olarak ileri beslemeli ağ kullanılmıştır. Çıktı katmanı doğrusal transfer fonksiyonlu olmuş ve ağ çıktısı glikoz tahmini olarak düzenlenmiştir. Çalışmada 15, 30 ve 45 dakikalık zaman aralıkları için glikoz tahminleri yapılmıştır. Performans göstergesi olarak hata kareleri ortalaması ve tahmin gecikmeleri kullanılmıştır. En iyi sonuçlar sırasıyla 15 dakika için, sonraki en iyi sonuç 30 dakika ve 45 dakika için olduğu gözlemlenmiştir. Çalışmada aynı zamanda otoregresif analiz yöntemiyle de çözüm yapılmış ve 2 yöntem karşılaştırılmıştır. Karşılaştırma sonucunda yapay sinir ağı kullanılarak oluşturulan modelin daha iyi performans gösterdiği sonucuna varılmıştır.

Pappada ve diğ. (2008), kan şekeri tahmini için diyabet hastalarının glikoz ölçüm değerlerini, insulin değerlerini, karbonhidrat değerlerini ve diğer verilerini kullanmışlardır. Çalışmada başarılı tahminler yapmak için veri setine, gecikmeli ileri beslemeli yapay sinir ağları makina öğrenme algoritması uygulanmıştır. Bu sinir ağları, ileri beslemeli bir şekilde birbirine bağlanan çok sayıda işlem öğesi katmanından oluşmuştur. Hasta sayılarının 11 ile 17 hasta arasında olması durumlarına göre modelin tahmin hata değerleri hesaplanmış ve birbirleriyle karşılaştırılmıştır.

Perez-Gandia ve diğ. (2010), kan şekeri tahmini için 9 adet hastanın glikoz ölçüm değerlerini değerlerini kullanmışlardır. Çalışmada başarılı tahminler yapmak için veri setine, yapay sinir

ağları makina öğrenme algoritması uygulanmıştır. Ağ mimarisinde ilk katman 10 nöron ve ikinci katman 5 nörondan oluşmuştur. Transfer fonksiyonunun her 2 katmanda da sigmoid özellikte olduğu belirtilmiştir. İleri beslemeli ağ kullanılmış ve çıktı katmanı doğrusal fonksiyon özelliğinde tercih edilmiştir. Performans kriteri olarak da ortalama hata kareler kökü değerleri hesaplanmıştır.

Robertson ve diğ. (2011), Elman tekrarlayan yapay sinir ağı modeli kullanarak kandaki glikoz yoğunluğunu tahmin etmeye çalışmışlardır. Projede geliştirilen yapay sinir ağları geri beslemesiz, ileri beslemeli yapay sinir ağlarından oluşmuştur. Çözüm modelinde 4 girdi değişkeni düşünülmüştür. Gizli katmanda 2 nöron ve çıktı katmanında 1 adet nöron çalışmada kullanılmıştır. Gizli katmanda tansig aktivasyon fonksiyonu, çıktı nöronu doğrusal aktivasyon fonksiyonuna sahip olarak modelenmiştir. Tek bir vaka senaryosundan elde edilen 28 adet veri çalışmada kullanılmıştır. Çalışmada en doğru tahminlerin 24 saatlik günlük döngünün sonunda yapıldığı görülmüştür. Performans kriteri olarak hata kareleri ortalaması kullanılmıştır.

Leon ve diğ. (2012), tip 1 diyabet hastalarının glisemik kontrolü için bu çalışmayı yapmışlardır. Çalışmada bulaşıcı hastalıkların matematiksel olarak modellemesini basitleştiren kompartman (bölmeli) model ve yapay sinir ağları ile kurulan model olmak üzere 2 farklı model kullanılmıştır. Kullanılan yapay sinir ağı modeli olarak çok katmanlı tekrarlayan yapay sinir ağı modeli tercih edilmiştir. Bu model yapay sinir ağları ile öğrenme işlemini yapmış ve sonrasında Karman filtresi tahmin gücünü kullanmıştır. Gizli katman logaritmik aktivasyon fonksiyonlu ve çıktı katmanı da doğrusal aktivasyon fonksiyonlu özellikte kurulmuştur. Çalışma verileri 2 hasta aracılığıyla elde edilmiş ve performans kriteri olarak izleme sinyalleri hataları karşılaştırılmıştır.

Fernandez ve diğ. (2012), kan şekeri tahmini yapmak amacıyla yapay sinir ağları kullanarak yaptıkları çalışmada gerçek hastalardan elde ettikleri veriler aracılığıyla Simulink yazılımını kullanmışlardır. Kurulan modelde Newton yönteminin bir varyasyonu olarak Levenberg-Marquardt yapay sinir ağı eğitim algoritması kullanılmıştır. 30 gün boyunca 20 adet diyabet hastasının gözlem değerlerini çalışmada veri seti olarak değerlendirilmiştir. Girdi değişkenleri olarak kan şekeri seviyeleri, insulin dozajları, karbonhidrat sayımı ve hastaların yapmış

oldukları fiziksel aktiviteler kullanılmıştır. Glikoz tahmini çalışmasında ysa aracılığıyla yapılan tahminlerde hedef performans kriteri olarak da hata kareleri ortalaması kullanılmıştır.

Shanti ve diğ. (2012), yaptıkları çalışmada sürekli glikoz izleme sensörü verisi ile kandaki glikoz değerlerini tahmim etmeye çalışmıştır. Çalışmada özellik tabanlı olarak adlandırılan yapay sinir ağı modeli, gradyan azalmalı yapay sinir ağı modeli ve levenberg-marquart algoritması kullanılarak oluşturulan yapay sinir ağı modelleri kullanılmıştır. Özellik tabanlı yapay sinir ağı modellerinin performansının diğer modellere göre, daha yüksek olduğu sonucuna varılmıştır. Çalışmada 30, 45 ve 60 dakikalık zamanlar için Matlab programı aracılığıyla tahminler yapılmıştır. Performans kriteri olarak hata kareler ortalaması değerleri kullanılmıştır.

Eljil ve diğ. (2013), sürekli glikoz izleme sensörleri aracılığıyla toplanan glikoz ölçümlerinin yakın geçmiş sonuçlarını kullanan bir hipoglisemi tahmin modeli önermişlerdir. Modelin, son iki glikoz ölçümü ve aralarındaki farkı kullanarak 30 dakikalık bir tahmin ufku için hipoglisemi durumlarını doğru bir şekilde tahmin edebildiğini ortaya koymuşlardır. Bu çalışmada zaman esaslı olan diğer bir ifadeyle nonlinear otoregresif modelli yapay sinir ağları kullanılmıştır. Ağ modeli 20 nöronlu 1 gizli katmandan oluşmuş ve aktivasyon fonksiyonu olarak lojistik sigmoid fonksiyonu kullanılmıştır. Çalışma veri setini oluşturan hasta grubunun hipoglisemi risklerini tahmini için genelleşmiştir.

Botwey ve diğ. (2014), 23 adet Tip 1 diyabet hastasından aldıkları verileri kullanarak hastaların kanındaki glikoz yoğunluğunu tahmin etmeyi amaçlamışlardır. Çalışmada 15, 30 ve 45 dakikalık zaman periyotları için tahminler yapılmıştır. Otoregresif modeller ile yapay sinir ağları yöntemleri kullanılmış ve performans kriterleri olarak hata kareler ortalaması, gecikme süresi ve korelasyon katsayısı değişkenlerine göre karşılaştırmalar yapılmıştır. Kurulan yapay sinir ağı çözüm modeli olarak tekrarlayan yapay sinir ağları modeli tercih edilmiştir.

Kondiloğlu (2015) diyabet hastalarında kan şekeri değerinin doğru tahmin edilmesi için öznitelik tabanlı oluşturdukları veri setlerine yönelik çeşitli yöntemler uygulamışlardır. Uygulanan yöntemlerden birisi de yapay sinir ağları yöntemleri olarak tercih edilmiştir. Kullanılan algoritma çok katmanlı ysa algoritması olarak düzenlenmiştir. Algoritma oluşturulmuş öznitelik değişkenlerinin yarısı kadar gizli katman oluşturulacak şekilde

çalıştırılmıştır. Korelasyon katsayıları çalışmada performans kriteri olarak kullanılmış ve ysa ile diğer yöntemlerin sonuçları karşılaştırılmıştır.

Zarkogianni ve diğ. (2015), yaptıkları çalışmada kan şekeri verilerini kullanarak kandaki glikoz tahmin etmek amacıyla çeşitli yöntemler kullanmışlardır. Kullandıkları yöntemler olarak; ileri beslemeli yapay sinir ağı modelleri, öz düzenleyici haritalarla öğrenme modeli, dalgacık işlevli bir sinirsel ağ modeli ve lineer regresyon modellerini tercih etmişlerdir. Çalışma Matlab programı aracılığıyla, 30-60 ve 120 dakikalık tahminler yapmak için, 10 adet Tip 1 diyabet hastasının 6 günlük gözlem sonucunda oluşturulan verileri ile analiz edilmiştir. Kullanılan 4 yöntemin performans karşılaştırılmaları yapılmıştır.

Hamdi ve diğ. (2017), kan şekeri seviyelerini doğru tahmin edebilmek amacıyla tip 1 diyabet hastası 12 kişinin sürekli glikoz ölçüm verilerini kullanarak model oluşturmuşlardır. Oluşturulan yapay sinir ağı modeli 3 katmandan oluşmuştur. Model; girdi katmanı, hiperbolik tanjant sigmoid transfer fonksiyonuna sahip gizli katman ve doğrusal aktivasyon fonksiyonlu çıktı katmanından oluşmuştur. Kullanılan optimizasyon tekniği olarak Levenberg-Marquardt algoritması tercih edilmiştir. Çalışmada her bir hasta için ayrı ayrı performans değerleri hesaplanmıştır. Bu bakımdan çalışma genelleştirici veya bütünleştirici değil, özelleştirici bir yaklaşım sunmuştur. Çalışma sonuçları önerilen yapay sinir ağı tabanlı modelin kan glikoz seviyelerini minimum hata ile tahmin etmekte olduğunu göstermiştir.

Mirshekarian ve diğ. (2017), karmaşık fizyolojik modeller geliştirerek kan şekeri seviyelerini doğru bir şekilde tahmin etmeyi amaçlamışlardır. İnsülin kullanımı, karbonhidrat alımı ve egzersiz faktörleri kan şekeri seviyesini tahmin etmede zorlaştırıcı faktörler olarak kabul edilmiştir. Çalışmada destek vektör regresyonu ve yapay sinir ağları ile tahminlemeler yapılmıştır. Kullanılan sinir ağı yinelemeli (recurrent) yapay sinir ağı modeli olarak tercih edilmiştir. Yinelemeli yapay sinir ağı modeli ile elde edilen sonuçların destek vektör regresyonuna göre daha başarılı sonuçları elde ettiği tespit edilmiştir.

Bamgbose ve diğ. (2017), hastaların tıbbi kayıtları ve gerçek zamanlı kontrol sonuçları ile ilgili verilere dayanarak tip 1 diyabet hastaları için kapalı döngü insülin uygulaması için entegre bir kontol sistemi geliştirmeyi amaçlamışlardır. Çalışmada kurulan yapay sinir ağı modeli yarı-statik ve dinamik veriler içeren 23 özellik kullanılarak dizayn edilmiştir. İki katmanlı sinir ağı modeli, 8 gizli nöron ile eğitilmiştir. Çalışmada standart geri yayılma

algoritmasına göre daha hızlı bir algoritma olarak kabul edilen Levenberg-Marquerdt algoritması kullanılmış ve veriler % 70 eğitim, % 15 doğrulama ve % 15 test işlevlerine göre kullanılmıştır.

Yadav ve diğ. (2018), diyabet alanında önemli bir yenilik olarak invazif olmayan (cerrahi işlem gerektirmeyen durumlar) ve bu sayede hastada meydana gelebilecek ciddi komplikasyonların azaltılmasını amaçlayan kan şekeri ölçüm cihazı geliştirilmesi için çalışmışlardır. Çeşitli katmanlardan oluşması düşünülen sistemde glikoz konsantrasyonunu elde etmek için Levenberg-Marquardt tabanlı yapay sinir ağı kullanılmıştır. 2 veri setinden elde edilen toplam 84 veriden; 66 adeti eğitim için kullanılmış, kalan 18 adeti sinir ağının test edilmesi için kullanılmıştır. Giriş katmanı 4 nörondan, 2 gizli katmanın her biri 5 nörondan ve çıktı katmanı tek nörondan oluşan 4-5-5-1 sistemine göre model kurulmuş ve Matlab aracılığıyla çözümler yapılmıştır.

Özkaya (2018) kan şekeri tahmininde yapay sinir ağları, destek vektör makinaları, karar ağacı, Gauss regresyonu, k-en yakın komşu algoritması ve rastgele orman algoritmaları kullanmışlardır. En iyi sonuçlara yapay sinir ağları ile yapılan tahminlerde ulaşılmıştır. Kullanılan yapay sinir ağları modellerinden ileri beslemeli yapay sinir ağı çözümünde gradyan iniş özellikli model oluşturulmuştur. Diğer kullanılan yöntemler tekrarlayan yapay sinir ağları ve bulanık yapay sinir ağları olmak çalışmada tercih edilmiştir.

Kim ve diğ. (2020), tip-2 diyabet hastası olan ve hastanede yatan hastalar için derin öğrenmeyi kullanan kişiselleştirilmiş bir glikoz tahmin modeli önermişlerdir. Çalışmada 3 adet farklı çözüm algoritması uygulanmıştır. Bu algoritmalar; tekrarlayan yapay sinir ağları, geçitli tekrarlayan birim yapay sinir ağları ve uzun-kısa süreli bellek birimli yapay sinir ağları olarak tercih edilmiştir. Çalışmada glikoz ölçüm cihazı ile 20 hastadan 1 hafta boyunca veri toplanmıştır. Çalışma sonucunda en iyi çözümün tekrarlayan yapay sinir ağları ile elde edildiği ortaya çıkmıştır.

Tablo 2.5: Yapay sinir ağları kullanılarak yapılan çalışmaların analizi.

Yazarlar	Yıl	Tahmin Zamanı	Yazılım	Veriler (Hasta Sayısı)	Gözlem Zamanı	Girdi Değişkenleri				Performans Ölçüm Kriterleri
						Glikoz	İnsülin Dozaj	Karbon-hidrat	Diğer	
Sandham ve diğ.	1998	30 dakika	Matlab	2	10 Gün	✓	✓	-	✓	Ortalama hata kareleri toplamı kökü

Gandia ve diğ.	2007	15, 30 ve 45 dakika	-	9-6	72 Saat	✓	✓	✓	-	Hata kareler ortalaması
Pappada ve diğ.	2008	50-180 dakika	Neuro Solution	18	9 Gün	✓	✓	✓	-	Tahmin performansı, hata kareler ortalaması
Perez-Gandia ve diğ.	2010	15, 30 ve 45 dakika	Matlab/ Weka	9-6	Haftalık 72 saat / 4 Hafta	✓	✓	✓	-	Hata kareler ortalaması, hata ızgara analizi, tahmin sapmaları, ortalamadan sapma, korelasyon katsayısı
Robertson ve diğ.	2011	15, 30, 45 ve 60 dk.	Matlab	28	24 saat	✓	✓	✓	-	Ortalama hata kareleri toplamı kökü
Leon ve arkaadaşları	2012	-	Matlab	2	4/5 gün	✓	✓	✓	-	Sinyal izleme performansı, sinyal izleme hatası
Fernandez ve diğ.	2012	-	Simulink	20	30 gün	✓	✓	✓	✓	Hata kareler toplamı
Shanti ve diğ.	2012	30, 45 ve 60 dakikalık	Matlab	-	-	✓	-	-	-	Hata kareler ortalaması
Eljil ve diğ.	2013	30 dk.	Weka	10	-	✓	-	-	-	Duyarlılık, Youden indeks, doğruluk değerleri
Botwey ve diğ.	2014	15, 30 ve 45 dakika	-	23	-	✓	✓	-	-	Hata kareler ortalaması, gecikme süresi ve korelasyon katsayısı
Kondiloğlu	2015	-	Weka	70	Birkaç ay	✓	✓	✓	✓	Korelasyon katsayısı

Tablo 2.5 (devam)

Yazarlar	Yıl	Tahmin Zamanı	Yazılım	Veriler (Hasta Sayısı)	Gözlem Zamanı	Glikoz	İnsülin Dozaj	Karbon-hidrat	Diğer	Performans Ölçüm Kriterleri

Zarkogianni ve diğ.	2015	30,60,120 dk.	Matlab	10	10.70 ± 4.69 gün	✓	-	-	✓	Glikoz hata analizi, ortalama hata kareleri toplamı kökü, korelasyon katsayısı, ortalama mutlak bağıl fark
Hamdi ve diğ.	2017	15 dk.	-	12	14 Gün	✓	✓	-	-	Minimum hata oranı
Mirshekarian ve diğ.	2017	30 ve 60 dk.	-	5	400 Gün	✓	✓	✓	✓	Hata kareleri ortalaması
Bamgbose ve diğ.	2017	-	Matlab	300 Sanal Hasta	9 Saat	✓	✓	✓	-	Ortalama hata kareleri toplamı kökü
Yadav ve diğ.	2018	-	Matlab	84 Veri	-	✓	-	-	-	Ortalama mutlak yüzde hatası, korelasyon katsayısı
Özkaya	2018	15,30 ve 60 dk.	Phyton/ Matlab	-	30 Gün	✓	✓	✓	✓	Ortalama hata karesi, Clarke glikoz hata analizi, korelasyon katsayısı,tahmin süresi,bağıl fark
Kim ve diğ.	2020	30 dk.	-	20	7 Gün	✓	✓	-	-	Ortalama hata kareleri toplamı kökü, Clarke glikoz hata analizi ve ortalama mutlak yüzde hata

2.5.8. Destek Vektör Makineleri Kullanılarak Yapılan Diyabet Tahmini Çalışmaları

Makine öğrenme tekniklerinden destek vektör makinaları kullanılarak bir çok makalede kanda bulunan glikoz düzeyi tahmin edilmeye çalışılmıştır.

Georga ve diğ. (2010), tip1 diyabetik hastalarda glikoz tahminini içeren bir modelleme metadolojisi sunmuştur. Yöntem olarak glikoz konsantrasyonunun tahminlerini sağlamak için destek vektör regresyon makineleri kullanılmıştır. Elde edilen modelin tahmin yetenekleri, 3 tip diyabetik hastadan alınan veriler kullanılarak değerlendirilmiştir. Önerilen tahmin yönteminin klinik faydasını değerlendirmek için Clarke analiz değerleri kullanılmıştır. 1 sene sonra yaptıkları diğer çalışmada Georga ve diğ. (2011), kan şekeri tahmini için farklı tipte ve farklı dozlarda insulin kullanan 7 adet hastanın glikoz ölçüm değerlerini, insulin değerlerini ve karbonhidrat değerlerini kullanmışlardır. Çalışmada başarılı tahminler yapmak için veri

setine, destek vektör makinaları regresyonu uygulanmıştır. Uygulanan yöntem olarak lineer kernel tabanlı destek vektör makinaları tercih edilmiştir. Model her hasta için ayrı ayrı eğitilmiş ve aşırı uyum sorununun çözülmesi için çapraz doğrulama algoritması kullanılmıştır. Performans kriteri olarak da ortalama hata kareler kökü değerleri hesaplanmış ve sonuçlar yorumlanmıştır.

Yıldırım (2011) yapmış olduğu çalışmada avuç içi nem parametrelerini kullanarak destek vektör makineleri ile kandaki glikoz ve HbA1c değerlerini sınıflandırmıştır. Çalışmada destek vektör makinaları oluşturmada yaş, tansiyon, vücut kütle endeksi ve avuç içi nem öznitelikleri girdi değişkenleri olarak belirlenmiştir. Avuç içi nem parametreleri her 2 saniyede 1 sefer olmak üzere, 80 saniye süre boyunca kaydedilmiştir. Zaman ile avuç içi nem ölçüm sonuçları grafikleri çıkarılmış ve 2 saniyelik süreler için eğim değerleri hesaplanmışır. Çalışmada hem lineer tabanlı sınıflandırma ve hem de radyal tabanlı çekirdek fonksiyonu kullanılarak yapılan sınıflandırma, farklı öznitelik senaryoları altında yapılmıştır. Çalışmada açlık kan glikoz tahmininde % 82'lik ve HbA1c tahmininde % 84'lük sınıflandırma başarısı elde edimiştir.

Bunescu ve diğ. (2013), tarafından yapılan çalışmada hastaya özgü veriler üzerinde eğitilmiş bir destek vektör regresyon modeli için bilgilendirici özellikler oluşturmak üzere genel kan şekeri dinamiği fizyolojik modelini kullanan bir çözüm elde edilmiştir. Veri kümesindeki belirli bir test noktası için destek vektör regresyon sistemleri ile Gauss çekirdeği kullanılarak veriler eğitilmiştir. Fizyolojik bir model tarafından bilgilendirilen ve hastaya özel veriler üzerinde eğitilen bir destek vektör regresyon modeli, kan şekeri seviyelerini tahmin etmede diyabet uzmanlarından daha iyi performans göstermiştir ve 30 dakika önceden hipoglisemik olayların % 23'ünü tahmin edebilmiştir.

Georga ve diğ. (2013), bu çalışmada glikoz tahmini yapmak için çok değişkenli bir regresyon problemi olarak destek vektörü regresyonunu kullanmışlardır. Önerilen yöntem; glikoz profili, plazma insülin yoğunluğu, sistem çevriminde yemek kaynaklı glikoz ve fiziksel aktiviteler sırasında enerji harcama değişikenlerini kullanmıştır. Model 27 adet hastadan oluşan bir veri kümesi kullanılarak değerlendirilmiştir. Destek vektör regresyonu ile çöüzmde hem optimize ve hem de test etmek için her bir veri kümesine ayrı ayrı 10 kat çapraz doğrulama işlemi uygulanmıştır. Kurulan destek vektör regresyonu olarak non-lineer model

tercih edilmiştir. Elde edilen sonuçlar, kısa ve uzun vadeli doğru tahminlerin yapılabileceğini ortaya koymuştur.

Kondiloğlu (2015) yapay sinir ağı modelleri ile birlikte destek vektör makinaları aracılığıyla da diyabet hastaları için kan şekeri tahminleme modelleri kurmuşlardır. Çalışmada destek vektör makinası olarak adaptif sıralı minimizasyon optimizasyon modeli SMOReg (Sequential Minimal Optimization for Regression) olarak literatürde yer alan yapay öğrenme algoritması kullanılmıştır.

Özkaya (2018) yapmış olduğu çalışmada kan şekeri tahmini yapmak için uyguladığı destek vektör makinasında radyal tabanlı çekirdek fonksiyonunu tercih etmiştir. Modelde Lagrange C katsayısının alabileceği üst sınır değeri olarak ceza parametresi için 1000 değeri kullanılmıştır. Gamma katsayısı 0,1 ve epsilon hata değeri olarak 0,2 alınmıştır. Her bir tahmin ufku için, belirli bir süreden sonraki 15, 30 veya 60 dakika sonraki değerler çıktı olarak seçilmiştir. Çalışmada önceki 2 saat için tüm özellikler girdi vektörleri olarak kullanılmıştır ve bu şartlar altında çıktı değerleri tahmin edilmeye çalışılmıştır.

Mayo ve diğ. (2019), tip 1 diyabet hastalarında kısa vadeli kan şekeri seviyesinin tahminini yapmak için makine öğrenimi tekniklerini kullanmışlardır. Çalışmanın amacı farklı makine öğrenimi tekniklerini değerlendirmek ve performansı en iyi tekniği bulmak olarak düzenlenmiştir. Çalışmada en iyi performansı gösteren yöntemler yapay sinir ağları ve destek vektör makinaları olarak görülmüştür. Polinom özelliklerle eğitilmiş doğrusal destek vektör regresyon tabanlı modelin, normal ve hiperglisemik aralıklardaki kan şekeri seviyesi tahmininde en iyi model olduğu tespit edilmiştir. Hipoglisemik aralıklardaki kan şekeri seviyesi tahmininde en yüksek performanslı modelin çok katmanlı yapay sinir ağı modeli olduğu görülmüştür.

Tablo 2.6: Destek vektör makinaları kullanılarak yapılan çalışmaların analizi.

Yazarlar	Yıl	Tahmin Zamanı	Yazılım	Veriler (Hasta Sayısı)	Gözlem Zamanı	Girdi Değişkenleri				Performans Ölçüm Kriterleri
						Glikoz	İnsülin Dozaj	Karbon -hidrat	Diğer	
Georga ve diğ.	2010	15, 30, 60 ve 120 dk.	-	3	5/11/13 Gün	✓	✓	✓	✓ (Isı Akısı, Cilt Sıcaklığı, Egzersiz)	Korelasyon katsayısı, Clarke glikoz hata analizi, ortalama hata kareleri toplamı kökü

Tablo 2.6 (devam)

Yazarlar	Yıl	Tahmin Zamanı	Yazılım	Veriler (Hasta Sayısı)	Gözlem Zamanı	Glikoz	İnsülin Dozaj	Karbon -hidrat	Diğer	Performans Ölçüm Kriterleri
Georga ve diğ.	2011	15, 30, 60 ve 120 dk.	-	7	5-14 Gün	✓	✓	✓	-	Korelasyon katsayısı, Clarke glikoz hata analizi, ortalama hata kareleri toplamı kökü
Yıldırım	2011	80 sn	Matlab	199		-	-	-	✓	Sınıflandırma başarısı
Bunescu ve diğ.	2013	30 dk. ve 60 dk	Java	5	3 Ay	✓.	✓	✓	-	Ortalama hata kareleri toplamı kökü
Georga ve diğ.	2013	15, 30, 60 ve 120 dk.	-	27	5-22 Gün	✓	✓	✓	✓ (Ani enerji tüketimi)	Korelasyon katsayısı, Clarke glikoz hata analizi, ortalama hata kareleri toplamı kökü, hiperglisemik ve hipoglisemik tahminlerin başarısı
Kondiloğlu	2015	-	Weka	70	Birkaç ay	✓	✓	✓	✓	Korelasyon katsayısı
Özkaya	2018	15,30 ve 60 dk.	Phyton	-	30 Gün	✓	✓	✓	✓	Ortalama hata karesi, Clarke glikoz hata analizi, korelasyon katsayısı,tahmin süresi,bağıl fark
Mayo ve diğ.	2019	120 dk.	-	6	8 Hafta	✓	✓	✓	-	Ortalama mutlak bağıl fark, Clarke glikoz hata analizi

2.5.9. Rastgele Orman Algoritmaları Kullanılarak Yapılan Diyabet Tahmini Çalışmaları

Georga ve diğ. (2012), yaptıkları çalışmada tip 1 diyabet hastaları özelinde rastgele orman algoritmaları kullanılarak regresyon yöntemi ile glikoz konsantrasyon tahmini yapmayı amaçlamışlardır. Çalışmada 27 adet tip 1 hastası kayıtları kullanılmıştır. Önerilen yöntemin performansı 10 kat çapraz doğrulama ile değerlendirilmiştir. Çalışmada kullanılan yöntemlerin başarı ölçüm kriterleri olarak koreleasyon katsayısı ve ortalama hata kareleri toplamı kullanılarak, karşılaştırmalar yapılmıştır.

Xia ve diğ. (2015), tip 2 diyabet hastalarına yönelik Naive Bayes algoritması, ID3 algoritması, rastgele orman algoritması ve diğer sınıflandırma algoritmaları kullanmışlardır. Virginia Üniversitesi tıp fakültesi tarafından elde edilen veriler; 4 farklı gruba ayrılmıştır. 4

farklı grup verileri, 4 farklı veri madenciliği yöntemleri aracılığıyla hastalık risklerini tahmin etmek için kullanılmıştır. En iyi performansa sahip tahmin algoritmasının rastgele orman algoritması olduğu tespit edilmiştir.

Butwall ve Kumar (2015) yaptıkları çalışmada rastgele orman algoritmasının sınıflandırıcı özelliğinden yararlanmışlardır. Bu çalışma, veri madenciliği yaklaşımını planlamak, yürütmek ve özellikle yaşlı diyabet hastalarında fiziksel aktivite ve duygusal durumlar da dahil olmak üzere belirli yaşam tarzı parametreleriyle uyumlu olarak diyabet davranışını öngörmek için ilişkilendirme standartları tasarlamak için yapılmıştır. Çalışmada rastgele orman algoritması için farklı ağaç sayıları, farklı derinlik özellikleri gibi birbirinden farklı parametre değerleri kulllanılmıştır. En iyi yöntem sayesinde % 99,7 doğruluk oranında başarılı tahminler yapılmıştır.

Hidalgo ve diğ. (2017), tarafından yapılan çalışmada kandaki glikoz konsantrasyonlarını tahmin etmede yapılan gözlemler aracılığıyla 30, 60, 90 ve 120 dakikalık süreler kullanmışlardır. Bu çalışmada, glikoz konsantrasyonlarını modellemek ve tahmin etmek için rastgele orman algoritması ile beraber çeşitli makine öğrenme teknikleri kullanılmıştır. Çalışmada ki veriler İspanya'daki bir devlet hastanesinden elde edilen 10 adet hasta aracılığıyla elde edilmiştir. Java yazılımı kullanılarak, Clarke glikoz hata analizi ile performans ölçümleri yapılmıştır.

Özkaya (2018) kan şekeri değerleri tahmininde makine öğrenme tekniklerini kullanmıştır. Bu çalışmada 15, 30, 45 ve 60 dakika sonrası zamanlar için tahminler yapılmıştır. Kullanılan rastgele orman algoritması için, randomForestRegressor sci-kit öğrenme kitaplığı çalışmada kullanılmıştır. Bu kitaplığın tercih edilmesi sayesinde kompakt, kolay ve kullanışlı sonuçlar elde edilmiştir. Algoritmadaki girdi ölçeklendirmesi 0 ile 1 arasında düzenlenmiş, her ağacın büyümesi için maksimum derinlik değeri 10 olarak belirlenmiştir. Çalışmada yapay sinir ağları, rastgele orman algoritmaları ve destek vektör makinaları yöntemlerinin, Farklı tahmin zaman durumları için farklı yöntemlerin daha başarılı çözümler sunduğu tespit edilmiştir.

İslam ve diğ. (2021), kandaki glikoz seviyelerinin kantitatif tahmini için akıllı telefon videolarına dayalı bir teknik önermişlerdir. Videolar, akıllı telefon kameraları kullanılarak deneğin parmağının ucundan sabit bir şekilde toplanmış ve ardından bir fotopletismografi (PPG) sinyaline dönüştürülmüştür. Glikoz seviyesinin tahmini için temel bileşen analizine

dayanan temel bileşen regresyonu, kısmi en küçük kareler regresyonu, destek vektörü regresyonu ve rastgele orman algoritma modelleri çözümde kullanılmıştır. Rasgele orman algoritması çözümü için düğümler veya diğer ifadesiyle ağaçlar oluşturulmuştur. En düşük hata ile tahmin için gerekli olan optimum ağaç sayısı 70 olarak belirlenmiştir. Çalışmada kullanılan veri setine, dört istatistiksel öğrenme tekniği uygulandığında en iyi sonuç kısmi en küçük kareler regresyonu modeli ile elde edilmiştir.

Tablo 2.7: Rastgele orman algoritmaları kullanılarak 2010 yılı sonrasında yapılan bazı çalışmaların analizi.

Yazarlar	Yıl	Tahmin Zamanı	Yazılım	Veriler (Hasta Sayısı)	Gözlem Zamanı	Girdi Değişkenleri				Performans Ölçüm Kriterleri
						Glikoz	İnsülin Dozaj	Karbon-hidrat	Diğer	
Georga ve diğ.	2012	15, 30, 60 ve 120 dk.	-	27	5-22 Gün	✓	✓	✓	✓	Korelasyon katsayısı, Clarke glikoz hata analizi, ortalama hata kareleri toplamı kökü
Xia ve diğ.	2015	-	-	Hasta sayısı belirsiz - 7109 veri	4 yıl	✓	✓	-	-	Ortalama hata kareleri toplamı kökü, eğrinin altında kalan alan (AUC) ve ortalama mutlak hata
Butwall ve Kumar	2015	-	Matlab	768	-	✓	✓	✓	✓	-
Hidalgo ve diğ.	2017	30, 60, 90 ve 120 dk.	Java	10	10 Gün	✓	✓	✓	-	Clarke glikoz hata analizi
Özkaya	2018	15,30 ve 60 dk.	Phyton	-	30 Gün	✓	✓	✓	✓	Ortalama hata karesi, Clarke glikoz hata analizi, korelasyon katsayısı,tahmin süresi,bağıl fark
İslam ve diğ.	2021	60 saniye	Matlab	52	-	-	-	-	✓	Standart tahmin hatası

2.5.10. Hibrit Modeller Oluşturarak Yapılan Diyabet Tahmini Çalışmaları

Volker Tresp ve diğ. (1999), yaptıkları çalışmada oluşturdukları hibrit model aracılığıyla tekrarlayan sinir ağı modeli ile lineer hata modeli sistemi oluşturulmuş ve bölüm modelleri yaklaşımı uygulanmıştır. Bölüm modelleri yaklaşımı ile 1927 yılında Kermack ve McKendrick tarafında bulunan bulaşıcı hastalıkların matematiksel olarak modellenmesini sağlayan deterministik çözümlerini ortaya koymuştur. Çalışmada kullanılan veriler için 1 hasta 15 dakikalık aralıklarla 63 gün boyunca gözlemlenmiştir. CM, karbonhidrat alımının

emilimini, fiziksel egzersizi ve uygulanan insülin dinamiğindeki gecikmeyi simüle etmiştir. Sinir ağları (NNM) ise glikoz tahmini yapmak için CM tahminleri ile kan glikozunun mevcut ve önceki tahminlerini kullanmıştır.

Mougiakakou ve diğ. (2005), sinir ağı modeli ve bölüm modelleri yaklaşımlarından oluşan bir hibrit model oluşturarak kan glikoz seviyesi tahmini yapmaya çalışmışlardır. Oluşturulmuş hibrit modelde kullanılan yapay sinir ağları olarak gerçek zamanlı tekrarlı yapay sinir ağları tercih edilmiştir. Çalışmada Tip 1 Diyabet hastasından 69 gün boyunca elde edilen veriler kullanılmıştır.

Castillo ve diğ. (2010), yaptıkları çalışmada 45 dakikalık süre sonrası için, kandaki glikoz yoğunluğunun tahmin edilmesinde; Dalla Man'ın glikoz emilim modelini, Verdolink'in insüliln oranları ile ilgili yapmış olduğu çalışmayı ve otoregresif özellikte zaman serilerini kullanarak yeni çözüm yöntemleri geliştirmişlerdir. Teorik olarak elde edilen sonuçların aynı zamanda sürekli glikoz izleme araçları kullanılarak da performansları değerlendirilmiştir.

Zecchin ve diğ. (2011), fizyolojik bir model aracılığıyla kantitatif olarak tanımlanan karbonhidrat alımları hakkındaki bilgileri de kullanan bir sinir ağına dayalı bir kısa vadeli glikoz tahmin algoritması geliştirmişlerdir. Çalışmada lineer tahmin algoritmasıyla beraber Dalla Man glikoz emilim modeli ve yapay sinir ağlarından oluşan hibrit bir model önerilmiştir. 1 adet gerçek hasta ve 5 adet oluşturulmuş sanal hastadan alınan veriler kullanılmıştır. Yemek sonrası karbonhidrat değişimi ile kan şekeri yoğunluğu arasındaki ilişkiye yönelik model oluşturulmuştur.

Zarkogianni ve diğ. (2011), yaptıkları çalışmada 7 günlük gözlem sonucunda 10 adet diyabet hastasından alınan verileri kullanarak nonlineer model aracılığıyla kandaki glikoz miktarının tahmin edilmesini amaçlamışlardır. Nonlineer model ile bölüm modeli ve tekrarlayan sinir ağı modellerini kullanarak tahminleme yapmaya çalışmışlardır. Çalışmada yapılan tahminler 30 dakikalık zaman için düşünülmüştür. Girdi değişkenleri olarak kandaki glikoz değerleri, insulin dozajı ve karbonhidrat (CHO) değerleri kullanılmıştır.

Georga ve diğ. (2012), gözlem altında olan 27 adet diyabet 1 tip hastanın 5-22 gün arasındaki verilerini kullanarak rastgele orman algoritması tekniğine dayalı kandaki glikoz yoğunluğunu tahmin etmeyi amaçlamışlardır. Çalışmada çok değişkenli bir veri kümesi kullanılmıştır.

Kullanılan veriler olarak; glikoz yoğunluğu, plazma insülin konsantrasyonu, yemek kaynaklı glikozun bağırsak emilimi ve günlük enerji harcamaları tercih edilmiştir. Tarin'in plazma insülin konsantrasyon modeli, Lehmann'ın glikoz oranı görünüş modeli ve rastgele orman algoritması kullanılarak hibrit model oluşturulmuştur.

Balakrishnan ve diğ. (2012), oluşturdukları model de tüm hastalardan ziyade her bir hasta için ayrı ayrı kan şekeri tahmini modeli kurulması gerektiğini düşünmüşlerdir. Bu amaçla yaptıkları çalışmada inceledikleri 12 hasta için bütünleşik bir model yerine hastalara uygun model stratejisi izlemişlerdir. 9 hasta için kişiselleştirilmiş modeller olarak doğrusal modeller kullanılmıştır. Yöntemler olarak 1 hasta için Box-Jenkins zaman serileri, 1 hasta için durum uzay fonksiyonu ve kalan 7 hasta için transfer fonksiyonu kullanılmıştır. Kalan 3 hasta için de doğrusal olmayan Hammerstein-Wiener modelinin daha uygun olduğunu çalışmada göstermişlerdir. Bu nedenle, her hasta verileri farklı bir kişiselleştirilmiş modeli izlediği için, hastalar arası değişkenliğin önüne geçmek için hastaların uygun modeller tarafından yakalanması gerektiğini belirtmişlerdir.

Naumova ve diğ. (2012), meta-öğrenme makine öğrenimi yaklaşımı sunarak kan şekeri tahmin modeli geliştirmişlerdir. Geliştirilen meta öğrenme algoritmasında doğru kernelin ve parametrelerin seçilmesinin önemli olduğu belirtilmiştir. Algoritma 3 aşamada uygulanmıştır. İlk aşama optimizasyon olarak adlandırılmıştır. Bu aşamada girdi parametre değerleri ve kernel seçimi yapılmıştır. İkinci aşamada eğitim setlerinde hesaplamalar yapılmıştır. Son aşamada öğrenme sürecinin tamamlanması ve sonuçların elde edilmesi gerçekleştirilmiştir. Bu yaklaşım, diyabetik hastalar için çeşitli klinik denemelerde test edilmiştir ve literatürde bilinen algoritmalara özgü olmayan özellikler gösteren yeni bir kan şekeri öngörücüsü tasarımının geliştirilmesine olanak sağlamıştır.

Zecchin ve diğ. (2012), yaptıkları çalışmada Dalla Man tarafından geliştirilmiş yemekten sonra glikoz değişimini gösteren matematiksel modeli temel almışlardır. Bu model mantığı çerçevesinde lineer tahmin algoritması ve yapay sinir ağlarını uygulayarak paralel olarak sonuç üreten fizyolojik bir çözüm ortaya koymuşlardır. Kullanılan yapay sinir ağı, birbiri ile bağlantılı ve ileri beslemeli olarak seçilmiştir. Sinir ağında gizli katman 8 nörondan oluşmuş ve tanjant sigmoid fonksiyonunu kullanmıştır. Çıktı katmanı ise 1 nörondan oluşmuş ve lineer transfer fonksiyonunu kullanmıştır.

Balakrishnan ve diğ. (2013), Hovorka modelini, Berger modelini ve yapay sinir ağlarını kullanarak hibrit bir çözüm modeli geliştirmişlerdir. 34 adet hasta, 2 gün boyunca gözlemlenerek Matlab programı aracılığıyla çalışma geliştirilmiştir. Bu çalışmada yemek sonrası karbonhidrat kinetiğini hesaplamak için Hovorka matematiksel eşitlik modeli, insülin takibi ve değişimi için Berger'in insulin deneysel modeli ve kan şekeri tahmini için yapay sinir ağları kullanılmıştır. Kişiselleştirilmiş hibrit modellerin uygunluk ve çapraz doğrulama sonuçları, sadece kan şekeri dinamiklerinin doğru tahminini göstermekle kalmamış, aynı zamanda hibrit modellerin hastalar içinde de değişkenliği yakalama potansiyelini ortaya çıkarmıştır.

Wang ve diğ. (2013), 3 farklı tahmin yönteminin kullanarak kandaki glikoz tahmini yapmak için hibrit bir çözüm modeli oluşturmuşlardır. Kullanılan yöntemler olarak otoregresif model, öğrenme algoritmaları ve destek vektör regresyon yöntemleri kullanılmıştır. Kullanılan veriler 10 adet tip 1 diyabet hastasından elde edilmiş ve 860 adet gözlem değeri üzerinde çalışılmıştır. Performans kriterleri olarak hata kareler ortalaması, bağıl hata, Clarke analiz değerleri ve J indeksi kullanılmıştır.

Zarkogianni ve diğ. (2013), diyabet hastalarında kandaki glikoz miktarı tahmini yapmak amacıyla 12 adet diyabet hastasından aldıkları 10 günlük veriyi kullanmışlardır. Tahmin amacıyla oluşturulan model bölüm modeli ve öz düzenleyici haritalar modellerinin birleştirilmesi ile meydana gelmiştir. Model, 30 dakikalık süre ve 60 dakikalık süre sonraları için tahminleme yapmışlardır.

Zecchin ve diğ. (2014), yaptıkları çalışmada Dalla Man glikoz emilim modeli ve yapay sinir ağlarını bir arada kullanarak hibrit bir çözüm modeli ortaya koymuştur. Kullanılan modelde, 10 farklı denekten elde edilen veriler kullanılmıştır. Tahminler 30 dakika süre sonrası için yapılmıştır. Çalışmada yemek, geçmiş glikoz verileri, insulin ve fiziksel aktivite bilgileri kullanılmış ve zamana bağlı glikoz tahmini yapılmaya çalışılmıştır. Kullanılan yapay sinir ağı modeli olarak daha çok sismik problemlerin çözümünde kullanılan atlama sinir ağı modelleri seçilmiştir. Yazarın 2012 yılında yapmış olduğu referans çalışma ile karşılaştırılmış ve bu yöntemin daha iyi sonuç verdiği tespit edilmiştir.

Cescon ve diğ. (2015), bu makalede, kan şekerini gelecekte tahmin etmek amacıyla, çok adımlı veri odaklı bir model önermişlerdir. Model hem fizyolojik yöntemleri ve hem de

altuzay öğrenim modelini içermiştir. Literatürde yer alan Dalla Man fizyolojik modelleri karbonhidrat ve insülin alımları tahminleri için kullanılmıştır. Bununla birlikte boyutsallığı azaltan yaklaşım olarak tercih edilen çoklu doğrusal alt uzay öğrenimi yaklaşımı problemin çözümünde uygulanmıştır. Veriler, tip 1 diyabet hastalarının 3 gün boyunca hastanede izlenmesi sonucunda elde edilmiştir. Verilerin yarısı modeli geliştirmek için, kalan yarısı da modelin doğrulanması için kullanılmıştır. Doğrulama sonucunda 30 dakika, 60 dakika, 90 dakika ve 120 dakika ilerisi için ortalama popülasyon tahmin hatası, standart sapma değerleri ve tahmin değerleri bulunmuştur.

Gyuk ve diğ. (2019), ayakta tedavi edilen diyabet hastalarında kan şekeri değerlerinin hesaplanması için bir tahmin algoritması önermişlerdir. Önerilen modeller; genetik algoritma, Nelder-Mead yöntemi ve biyoritm teorisinin kullanıldığı hibrit modeller olarak tasarlanmıştır. Çalışmada besin tüketimi ve insulin kullanımı, girdi değişkeni olarak kullanılmıştır. Nelder-Mead simplex yöntemi sayesinde, algoritma optimize edilmeye çalışılmıştır. Çalışma da 2 farklı temellendirme yapılarak hibrit modeller oluşturulmuştur. İlki genetik algoritma ve biyoritm teorisi kullanılarak oluşturulan modeller ve ikincisi ise genetik algoritma ve Nelder-Mead yöntemi kullanılarak oluşturulan modeller olarak düzenlenmiştir. İkinci modeller ile daha yüksek doğrulukla sonuçlar elde edilmiştir.

Isfahani ve diğ. (2020), girdi değişkenlerinin sezgisel olarak seçimine dayalı hibrit sinir ağı modelleri kullanmışlardır. Önerilen modeller, hibrit dinamik dalgacık sinir ağları, hibrit dinamik bulanık dalgacık sinir ağları ve atlamalı yapay sinir ağı modeli olarak tercih edilmiştir. Çalışmada kullanılan dalgacık tabanlı ağlar, genetik algoritma-ortogonal en küçük kareler yöntemi ile seçilen baskın dalgacıklara göre tasarlanmıştır. Hibrit dinamik bulanık dalgacık sinir ağları model yapısı, bulanık dalgacık modellemesinde önemli bir yenilik olan bulanık kuralları kullanılarak geliştirilmiştir. Çalışma, 12 tip 1 diyabet hastasından alınan gerçek veriler üzerinde test edilmiştir. Çalışmada en iyi modelin hibrit dinamik bulanık dalgacık sinir ağları modeli olduğu sonucu ortaya çıkmıştır.

Tablo 2.8: Hibrit modeller oluşturularak yapılan çalışmaların analizi.

Yazarlar	Yıl	Tahmin Zamanı	Yazılım	Veriler (Hasta Sayısı)	Gözlem Zamanı	Girdiler				Performans Ölçüm Kriterleri	Model Tipleri
						Glikoz	İnsülin Dozaj	Karbonhidrat	Diğer		
Volker Tresp ve diğ.	1999	15 dk.	-	1	63 Gün	✓	✓	✓	✓	Hata kareler ortalaması	Tekrarlayan sinir ağı modeli + Lineer hata modeli
Mougiakakou ve diğ.	2005	15 dk.	-	-	69 Gün	✓	✓	✓	-	Ortalama hata kareleri toplamı kökü	Gerçek zamanlı tekrarlamalı öğrenme + Bölüm modelleri
Castillo ve diğ.	2010	45 dk.	-	15	3 gün	✓	✓	✓	-	Uyumluluk (FIT) değeri ve Clarke analiz değerleri	Dalla Man glikoz emilim modeli + verdonk insülin oranı modeli + Zaman serisi otoregresif exogen (ARX) modeli
Zecchin ve diğ.	2011	30 dk.	Matlab	5 sanal/1 gerçek	-	✓	-	✓	-	Ortalama hata kareleri toplamı kökü, J indeksi, zaman kazancı	Lineer tahmin algoritması + Yapay sinir ağı modeli + Dalla Man glikoz emilim modeli
Zarkogianni ve diğ.	2011	30 dk.	Matlab	10	7 Gün	✓	✓	✓	-	Ortalama hata kareleri toplamı kökü, korelasyon katsayısı, sürekli hata tablosu analizi	Dalla Man glikoz emilim modeli + Dalla Man insülin emilim modeli + Tekrarlayan yapay sinir ağı modeli
Georga ve diğ.	2012	15,30,60 ve 120 dk.	-	27	5-22 Gün	✓	✓	✓	✓	Ortalama hata kareleri toplamı kökü, korelasyon katsayısı ve Clarke glikoz hata analizi	Tarin'in plazma insülin konsantrasyon modeli + Lehmann'ın glikoz oranı görünüm modeli + rastgele orman algoritması
Balakrishnan ve diğ.	2012	-	Matlab	12	14 saat	✓	✓	✓	✓	Uyumluluk (FIT) değeri	Hovorka'nın yemek emilim modeli + İnsülin absorpsiyon kinetiği modeli + Zaman serisi modelleri
Naumova ve diğ.	2012	20, 40, 60 ve 75 dk.	Matlab	90	3-10 Gün Arası	✓	-	-	-	Clarke glikoz hata analizi, tahmin hatası analizi	Regularize edilmiş (yeniden düzenlenmiş) öğrenme yaklaşımı
Zecchin ve diğ.	2012	30 dk.	Matlab	20 sanal / 15 gerçek hasta	-	✓	-	✓	-	Ortalama hata kareleri toplamı kökü, J indeksi, zaman kazancı	Lineer tahmin algoritması + Yapay sinir ağı modeli (yemek dahil) + Dalla Man glikoz emilim modeli

Tablo 2.8 (devam)

Yazarlar	Yıl	Tahmin Zamanı	Yazılım	Veriler (Hasta Sayısı)	Gözlem Zamanı	Glikoz	İnsülin Dozaj	Karbonhidrat	Diğer	Performans Ölçüm Kriterleri	Model Tipleri
Balakrishnan ve diğ.	2013	-	Matlab	34	2 Gün	✓	✓	✓	✓	Hata Kareler Ortalaması, R2	Berger'in insülin kinetiği modeli + Hovorka'nın yemek emilim modeli + Yapay sinir ağları modeli
Wang ve diğ.	2013	15, 30 ve 45 dk.	Matlab	10	-	✓	-	-	-	Hata kareler ortalaması, bağıl hata, Clarke analiz değerleri ve J indeksi	Zaman serisi (AR) modeli + Öğrenme algoritması + Destek vektör regresyon
Zarkogianni ve diğ.	2013	30, 60 dk.	-	12	10 Gün	✓	✓	✓	-	Ortalama hata kareleri toplamı kökü, korelasyon katsayısı, sürekli hata tablosu analizi	Dalla Man glikoz emilim modeli + Dalla Man insülin emilim modeli + Öz düzenleyici haritalar
Zecchin ve diğ.	2014	30 dk.	Matlab	10	2-3 Gün	✓	✓	✓	✓	Ortalama hata kareleri toplamı kökü, zaman kazancı	Dalla Man Glikoz emilim Modeli+ Atlamalı yapay sinir ağı modeli
Cescon ve diğ.	2015	30, 60, 90, 120 dk.	Matlab	14	3 Gün	✓	✓	✓	-	Tahmin hatasının standart sapması, Clarke hatası analizi, mutlak fark, bağıl fark	Dalla Man glikoz emilim modeli + Dalla Man insülin absorpsiyon modeli + çoklu doğrusal alt uzay öğrenimi
Gyuk ve diğ.	2019	30-60 dk	Lavinia Lifestyle Mirror	26	142 gün	✓	✓	✓	-	Ortalama hata kareleri toplamı kökü ve Clarke analiz değerleri	Genetik Algoritma + Nelder Mead yöntemi, Genetik algoritma+Biyoritim teorisi
Isfahani ve diğ.	2020	30 dk.	Dexcom	12 gerçek hasta / 33 sanal hasta	36 saat	✓	✓	✓	-	Ortalama hata kareleri toplamı kökü, R2	Hibrit dinamik dalgacık sinir ağları, Hibrit dinamik bulanık dalgacık sinir ağları ve Atlamalı yapay sinir ağı modeli

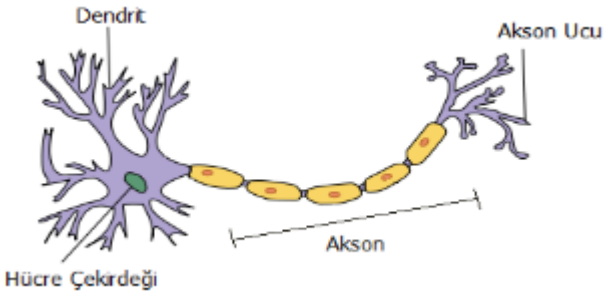
3. MALZEME VE YÖNTEM

3.1. YAPAY SİNİR AĞLARI

Bu bölümde, hibrit çözüm modellerini oluşturmada kullanılan yöntemlerden yapay sinir ağları hakkında bilgi verilmesi amaçlanmıştır.

3.1.1. Yapay Sinir Ağları Yöntemi Ortaya Çıkışı ve Zaman İçerisindeki Gelişimi

Modern bilimin başarılı olduğu en önemli husus doğayı ve canlıları gözlem yapması ve bunu icatlarda, keşiflerde ve yöntem geliştirmede kullanabilmesidir. İnsan, günümüz dünyasında diğer canlılar arasında mantıklı düşünebilme yetisinin en yüksek düzeyde olduğu canlıdır. İnsanoğlu, yeryüzündeki tüm canlı-cansız varlıkları kendi lehine kulanmakta veya kullanmaya çalışmaktadır. Doğadaki en mükemmel canlı olan insanın diğer canlılardan üstün olmasının sebeplerinden en önemlileri düşünebilme yetisi, değişen durumlara göre farklı kararlar alabilmesi, yorumlayabilmesi ve sonuçlar çıkartabilmesi özellikleridir. Bu özelliklerin kaynağı olarak insan beyni ve sinir sistemi yapısı bilim insanları tarafından incelenmiş ve incelenmeye devam etmektedir (Zhang ve diğ, 1998).



Şekil 3.1: Biyolojik sinir hücresi (Maltarollo ve diğ., 2013).

İnsanda sinir sistemi merkezi sinir sistemi ve çevresel sinir sistemi olmak üzere 2 ayrı sınıflamaya tabi tutulmuştur. Merkezi sinir sistemi; beyin ve omurilikten oluşmaktadır. Çevresel sinir sistemi ise beyin ve omurilik haricinde gangliyon ve sinirlerden oluşmaktadır. Merkezi sinir sistemindeki bu 2 önemli organ iç veya dış ortamlardan gelen uyarıları almakta ve değerlendirmektedir. Beyinde, yaklaşık 100 milyar nöron adıyla tabir edilen sinir hücreleri ve çok fazla sayıda destek hücreleri vardır. Sinir sistemi günümüzde dahi tam anlamıyla bilim insanları tarafından çözümlenebilmiş ve anlaşılabilmiş değildir. Klasik yapay sinir

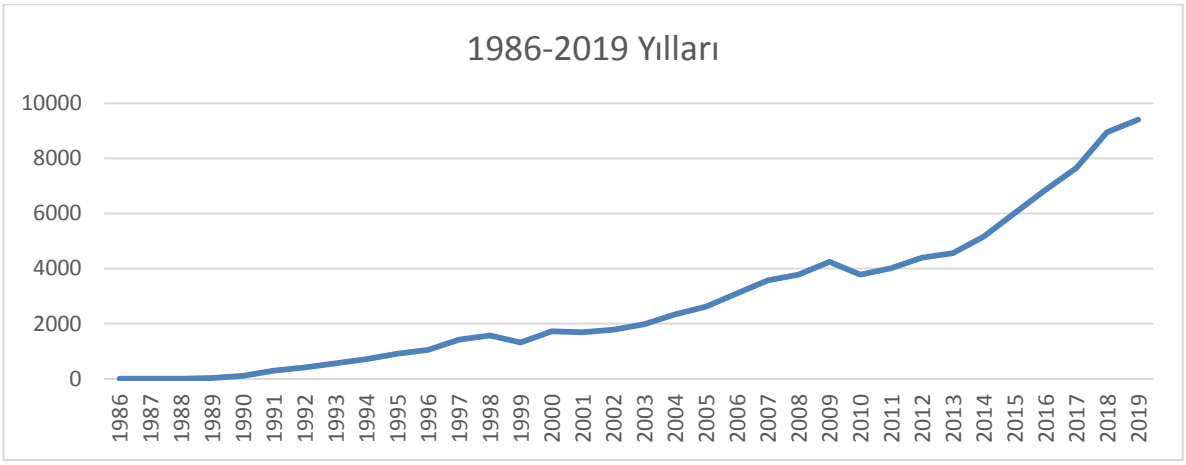
ağları mantığı da beyin-sinir sistemi yapısı örnek alınarak ortaya konmuştur. Şöyle ki; beyin nasıl ki yaşamış olduğu olayları ve bu olayların sonuçlarını, gelecekte yaşayacağı farklı durumları ve bu durumların sonuçlarını yorumlayabilme ve tahmin edebilme kapasitesine sahipse, yapay sinir ağları bunu başarmaya çalışmaktadır (Yazıcı ve diğ., 2007).

Geçmişten günümüze kadar çok farklı alanlarda kullanılan yapay sinir ağlarında kırılım yılları 1980'li yıllardan sonra olmuştur. 1980 öncesine bakıldığında, 1890-1940 arasında mühendislik anlamında değerli olmayan ama insan beyninin gözlemlendiği ve insan beyni özellikleri ile çalışma mantığının tespiti anlamında araştırmalar yapılmaya çalışılmıştır. 1943 yılında ilk yapay sinir ağları Warren S. McCulloch ve Walter H. Pitts tarafından ortaya çıkarılmıştır. Bu 2 araştırmacı yapay sinir ağlarının temellerini atmış olarak kabul edilmektedir (Piccini, 2004).

Tablo 3.1: 1975 yılı öncesi yapay sinir ağları gelişimi (Elmas, 2007).

Tarih	Yazar / Yazarlar	Çalışma/Makale/Kitap
1943	Warren S. McCulloch ve Walter H. Pitts	İlk yapay sinir ağları
1949	Donald Hebb	Hebb Kuralı
1954	Farley ve Clark	Rassal Ağlar ile Adaptif Tepki Üretme
1957	Frank Rosentblatt	Perceptron kavramı keşfi (1956 yapay zeka kavramı ortaya çıkışı)
1958	Frank Rosentblatt	Algılayıcı modeler (Adaline modeli benzeri bir model)
1959	Bernard Widrow ve Marcian Hoff	Adaline ve Madaline ysa modelleri
1961	Caianiello	Caianiello nöron denklemi geliştirme
1963	Bernard Widrow	Ters sarkaç denetleyicisi
1965	Nilss. J. Nilson	Öğrenen Makinalar
1969	Marvin Minsky ve Seymour Papert	Algılayıcılar: hesaplama geometrisine Giriş
1972	John R. Anderson	İlişkisel-Çağrışımlı bellek
1972	Teuvo Kohonen	İlişkisel-Çağrışımlı bellek

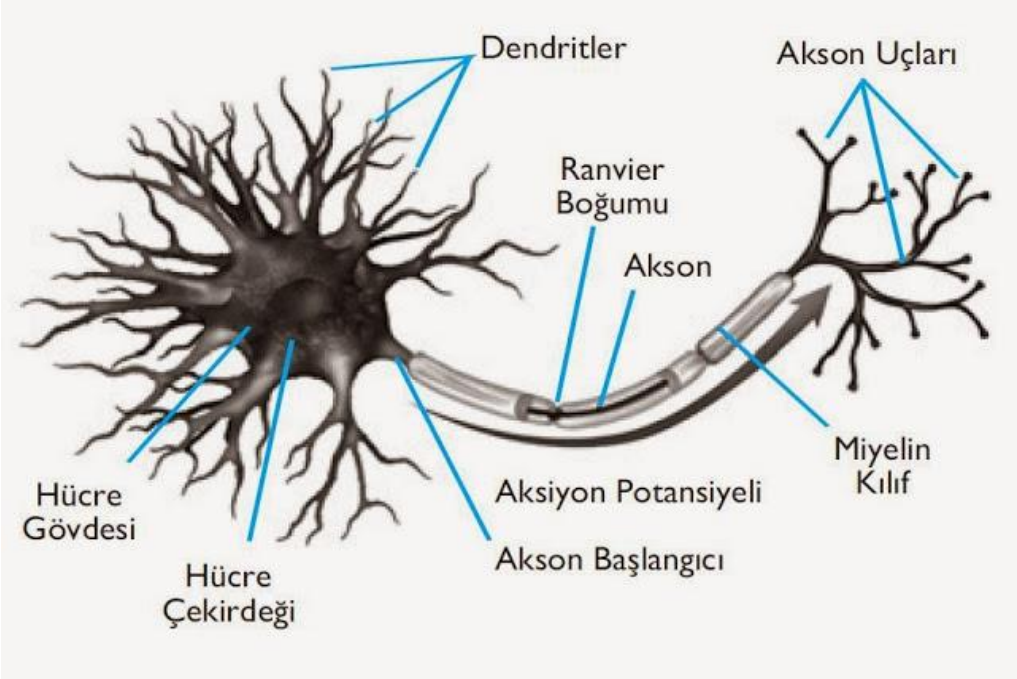
Yapay sinir ağlarının mühendislik çalışmalarında kullanılmaya başlandığı tarihler 1960'lı yıllardan sonraki tarihlerdir. Bu noktada bilgisayar teknolojileri gelişimi ile beraber yapay zeka kavramınının ortaya atılması ve geliştirilmesi etkili olmuştur.



Şekil 3.2: Yıl bazlı yapay sinir ağlarından bahsedilen çalışmaların sayısal değişimi (www.webofknowledge.com, 2019).

1986 yılından sonra yapay sinir ağları ile ilgili çalışma konuları farklılaşmış ve bu alanda meydana gelen çalışma sayıları artmıştır.

3.1.2. Yapay Sinir Ağları Modeli Oluşturma Araçları ve Temel Elemanları



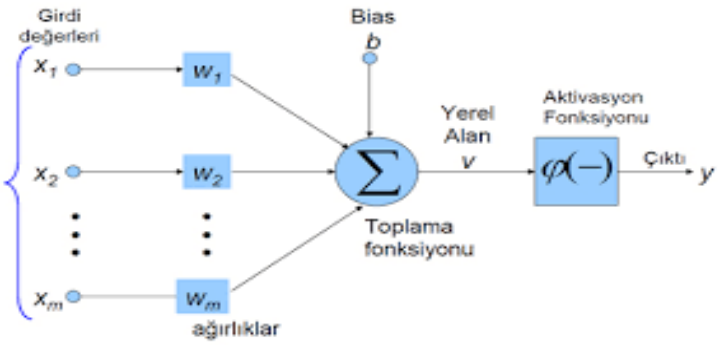
Şekil 3.3: Biyolojik sinir ağ model yapısı (Maltarollo ve diğ., 2013).

Şekil 3.3.'de görüldüğü gibi biyolojik sinir ağları; hücre gövdesi, hücre çekirdeği, dendrit, ranvier boğumu, akson ve miyelin kılıftan oluşmaktadır.

Tablo 3.2: Biyolojik sinir ağ yapısı ile yapay sinir ağları yapısı karşılaştırılması (Sarıoğlu, 2013).

Biyolojik Sinir Ağı	Yapay Sinir Ağı
Sinaps	Ağırlık
Dendrit	Toplama Fonksiyonu
Hücre Gövdesi	Aktivasyon Fonksiyonu
Akson	Sinir Hücresi Çıkışı
Sinir Hücresi	Proses Elemanı
Sinir Sistemi	Hesaplama Sistemi

Sağıroğlu ve diğ. (2013), yaptıkları çalışmada biyolojik sinir ağları ve yapay sinir ağları karşılaştırılması tabloda gözükmektedir.



Şekil 3.4: YSA elemanları (Öztemel, 2006).

Notasyon

Xi: Girdi değişkenleri

wi: Ağırlıklar

yi: Çıktı değişkenleri

Girdi ve çıktı kavramları anlaşıldıktan sonra işin yöntemin matematik kısmının anlaşılabilmesi için ağırlık ve fonksiyonlar devreye girer.

3.1.2.1. Girdiler

Yapay sinir ağlarında yer alan temel elemanlardan girdiler, bilgilerin dışarıdan alınması ve bu bilgilerin işlenmesi durumudur. Girdiler 2 şekilde temin edilebilir. Birincisi doğrudan dış dünyadan alınması, ikincisi ise kendinden önceki katman yada katmanlardan elde edilen işlem elemanın çıkış değeri de olabilir. Bilgiler, doğrudan dış ortamdan gelebildiği gibi aynı zamanda bazı girdiler diğer bilgilerin çıktıları olarak da düşünülebilir. Yani farklı girdilerden

oluşan çıktılar, başka bir yapıda girdi olarak kullanılıp yeni çıktılar elde edilmek amaçlanabilir (Öztemel, 2012).

3.1.2.2. Ağırlıklar

Yapay sinir ağ model yapılarında ağırlıklar girdilerin ağ model yapısında bağlantıların önem derecesini verir. Ağırlıklar, girdilerin çıktı üzerindeki değerini belirlemek için düşünülen katsayılardır. Bir başka ifadeyle, her girdi kendisine ait bir ağırlığa sahip olmakla beraber girdilerin önemini belirleyen katsayılar aracılığıyla çıktı tahmin edilir. Yapay sinir ağlarında ağırlıklar ile bağlantıların önem sıraları tespit edilebilir. Ağırlıklar'ın negatif (-) olması, pozitif (+) olması veya sıfır (0) olması durumları olabilir. Ağırlık değerinin eksi değer olması, önemsiz demek değildir. Ağırlıkların (+) ve (-) olması durumları, bu ağırlıkların etkilerinin pozitif veya negatif olduğu anlamına gelir (Öztürk ve Şahin, 2018).

3.1.2.3. Birleştirme Fonksiyonu

Birleştirme fonksiyonları; toplam fonksiyonu, çarpım fonksiyonu, maksimum fonksiyonu, minimum fonksiyonu, çoğunluk fonksiyonu ve kümülatif fonksiyonu olarak sınıflandırılmıştır. Birleştirme fonksiyonları hücrelerin net girdisini hesaplayan fonksiyonlardır. Birleştirme fonksiyonu seçimi deneme-yanılma yöntemi ile seçilebileceği gibi tasarımcı isteğine bağlı olarak da seçilebilir (Çayıroğlu, 2015).

1.Toplam Fonksiyonu

Toplam fonksiyonunda, girdilerin ağırlık değerleri ile çarpılması ve çarpılan değerlerin birbirleri ile toplanması ile girdi hesaplanır.

Notasyon

Xi: Girdi değişkenleri

wi: Ağırlıklar

yi: Çıktı değişkenleri

$$\text{Net Girdi} = \sum_{i=1}^{N} \text{Xi} * \text{wi} \quad (3.1)$$

2.Çarpım Fonksiyonu

Çarpım fonksiyonunda, girdilerin ağırlık değerleri ile çarpılması ve çarpılan değerlerin birbirleri ile çarpılması ile girdi hesaplanır.

$$\text{Net Girdi} = \prod_{i=1}^{N}(Xi * wi)$$

(3.2)

3.Maksimum Fonksiyonu

Maksimum fonksiyonunda, ağırlıklar girdiler ile çarpılarak, en büyük çarpım sonucu girdi olarak kullanılır.

$$\text{Net Girdi} = \text{Max}(Xi * wi)$$

(3.3)

4.Minimum Fonksiyonu

Minimum fonksiyonunda, ağırlıklar girdiler ile çarpılarak, en küçük çarpım sonucu girdi olarak kullanılır.

$$\text{Net Girdi} = \text{Min}(Xi * wi)$$

(3.4)

5.Çoğunluk Fonksiyonu

Çoğunluk fonksiyonunda, ağırlıklar girdiler ile çarpılır. Çarpılan değerler, pozitif ve negatif olarak 2 sınıfa ayrılır. Ayrılan sınıflardaki büyük olan sayı girdi olarak kullanılır.

$$\text{Net Girdi} = \sum_{i=1}^{N} \text{sgn}(Xi * wi)$$

(3.5)

6. Kümülatif Fonksiyonu

Kümülatif fonksiyonunda, hücrelere gelen bilgilerin ağırlıkları toplanır. Daha önceki girdi bilgileri, hesaplanan yeni girdi değerleri ile toplanarak net girdi hesap edilir.

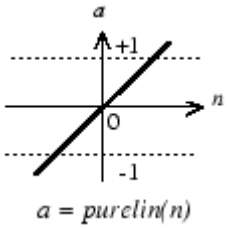
$$\text{Net Girdi} = \text{Net(eski)} + \sum_{i=1}^{N} Xi * wi$$

(3.6)

3.1.2.4. Aktivasyon Fonksiyonu

Aktivasyon fonksiyonları sinir hücrelerinin davranışlarını belirleyen önemli faktörlerdendir. Biyolojik hücrelerde hücre gövdesinin yaptığı görevi sinir sisteminde aktivasyon fonksiyonu yerine getirir.

1. Doğrusal (Lineer) Transfer Fonksiyonu

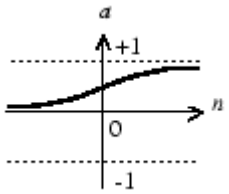


Şekil 3.5: Lineer transfer fonksiyonu (Çayıroğlu, 2015).

$$f(x) = n * a \quad (3.7)$$

Purelin lineer fonksiyonu değişim aralığı [-1 +1] aralığıdır. Bu aralıkta fonksiyonda lineer olmayan değişimler olur.

2. Logaritmik Sigmoid Transfer Fonksiyonu

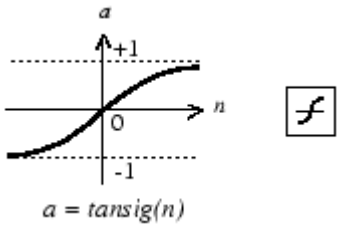


Şekil 3.6: Logaritmik sigmoid fonksiyonu (Çayıroğlu, 2015).

$$a = \frac{1}{1+e^{(-n)}} \quad (3.8)$$

Logaritmik sigmoid fonksiyonları değişim aralığı [0 +1] aralığıdır. Bu aralıkta fonksiyonda lineer olmayan değişimler olur.

3. Hiperbolik Tanjant Sigmoid Tipi Transfer Fonksiyonu



Şekil 3.7: Tanjant-sigmoid fonksiyonu (Çayıroğlu, 2015).

S: Net giriş (sütun) vektörlerinin S-Q matrisi

FP: Fonksiyon parametrelerinin yapısı

A = tansig (N, FP)

$$a = \frac{2}{1+e^{(-2n)}} - 1$$

(3.9)

Tanjant-sigmoid fonksiyonu değişim aralığı [-1 +1] aralığıdır. Bu aralıkta fonksiyonda lineer olmayan değişimler olur.

4. Adım Aktivasyon Tipi Transfer Fonksiyonu

Gelen girdiler eşik değerlerine göre tasniflenir. Eşik değeri durumuna göre çıktı değerleri 0 veya 1 değerlerini almaktadır.

$$f(n) = \begin{cases} 1, & n > \text{Eşik Değer} \\ 0, & n \leq \text{Eşik Değer} \end{cases}$$

(3.10)

5. Eşik Değer Tipi Transfer Fonksiyonu

Gelen girdiler eşik değerlerine göre tasniflenir. Eşik değeri 0 ve altında ise 0 çıktısı, eşik değeri 1'den büyükse 1, 0-1 aralığında olursa n çıktıları üretilir (Çayıroğlu, 2015).

$$f(n) = \begin{cases} 0, & n \leq 0 \\ n, & 0 < n < 0 \\ 1, & n \geq 1 \end{cases}$$

(3.11)

3.1.2.5. Gizli Katman

Gizli katman, uçakların kara kutusuna benzetilebilir. Gizli katmanlar, girdi katmanı ile çıktı katmanı arasındaki köprü vazifesi görmektedir. Gizli katman bir veya daha fazla katmandan oluşmaktadır. Gizli katman sayısı ile gizli katmanlardaki hücre sayıları tasarımcılar tarafından deneme-yanılma yöntemi veya farklı değerlendirme ölçütleri aracılığıyla belirlenir (Bilen, 2014).

3.1.2.6. Çıktı Katmanı

Çıktılar elde edilmek istenen sonucun değeri veya değerleridir (Öztemel, 2006). Amaç optimum ve optimum yakın çıktılar elde etmektedir. Elde edilen çıktıların başarısı, yöntemin başarısı ve aynı zamanda modelin başarısı anlamına gelir.

3.1.3. Yapay Sinir Ağları Sınıflandırılması

3.1.3.1. Tek Katmanlı Algılayıcılar

Tek katmanlı algılayıcı yapay sinir ağları yapısı sadece girdi ve çıktı katmanlarından oluşmaktadır. Basit ağlar olarak da kabul edilen algıyacılarda, çıktı fonksiyonu özellik olarak lineer fonksiyondur. Çıktı değeri 0 veya 1 değerlerinden birisini alır. Yapıda eşik değeri vardır, elde edilen eşik değerine göre çıktı uygun değeri alır (Öztemel, 2006).

$\Theta = Eşik\ Değeri$

$$f(x) = \begin{cases} 0, & \text{Net Girdi} < \Theta \\ 1, & \text{Net Girdi} > \Theta \end{cases}$$

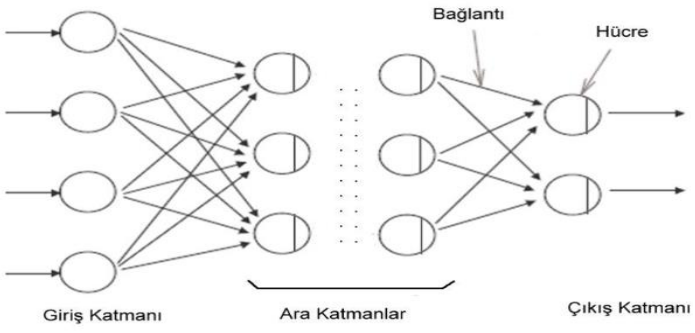
(3.12)

3.1.3.2. Adaline ve Madaline Ağlar

1959'da Stanford Üniversitesinden Bernard Widrow ve Marcian Hoff, adaline ve madaline ysa modellerini ortaya koymuşlardır. Adaline, adaptif lineer nöron olarak da adlandırılan basit sinir ağı yapılarındandır. Madaline ise adaline yapının 2 tabakalı olan halidir. İlk yapay sinir ağları modelleri içerisinde yer almaktadır. Transfer fonksiyonu olarak lineer transfer fonksiyonu kullanır (Kıraç, 2011).

3.1.3.3. İleri Beslemeli Ağlar

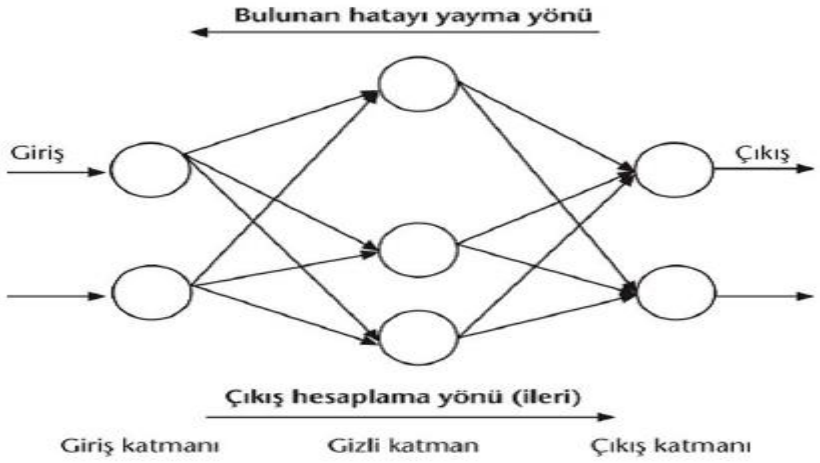
Algılayıcılar ile adaline-madaline yapılar bazı problemlerin çözümüne yönelik yeterli çözümler üretememesinden dolayı çok katmanlı algılayıcılar geliştirilmiştir. Çok katmanlı algılayıcılardan ileri beslemeli ağlar tek yönlü olarak sinyal akışını sağlamaktadır. İleri beslemeli ağlar giriş katmanı, gizli katman ve çıkış katmanlarından oluşmaktadır. Wilamowski 2003'de yaptığı tespitte aktivasyon fonksiyonları gradyan tabanlı eğitime olanak sağlar.



Şekil 3.8: İleri beslemeli sinir ağı modeli (Öztemel, 2006).

3.1.3.4. Geri Beslemeli Ağlar

Çok katmanlı algılayıcılardan geri beslemeli ağlarda katmanlar arasındaki bağlantılar tek yönlü değil, çift yönlüdür. Bu ağ yapılarında minimum 1 tane geri besleme akışı bulunmaktadır. Bu tip ağ yapıları güçlü ve oldukça karmaşık yapıdadır (Kabalcı, 2015).



Şekil 3.9: Geri beslemeli sinir ağı modeli (Kabalcı, 2015).

3.1.4. Yapay Sinir Ağlarının Eğitim Yöntemleri

Yapay sinir ağlarını eğitmek için literatürde birçok farklı eğitim yöntemi bulunmaktadır.

3.1.4.1. Eğim Azaltımı Yöntemleri

Temel eğim azaltımı (gradient descent) yöntemine göre nöronların ağırlıklarını performans fonksiyonunun gradyan değerinin azaldığı yönde güncelleyen en popüler eğitim yöntemleridir. Gradient descent eğitim yöntemleri geriye yayılımlı, momentumlu ve esnek geriye yayılımlı olmak üzere üç yöntemden oluşmaktadır (Sharma ve Venugopalan, 2014).

3.1.4.2. Birleşik (Eşlenik) Eğim Yöntemleri

Temel eğim azaltımı yöntemleri (gradient descent yöntemi) ağırlıkları, performans fonksiyonunun gradyan değerinin en hızlı şekilde azaldığı yöne doğru ayarlamaktadır. Bu en hızlı şekilde yakınsama sağlayamamaktadır. Birleşik eğim (conjugate gradient) yöntemleri, en çok azalma yönlerini bulmak yerine eşlenik yönler arayarak genellikle daha hızlı yakınsama sağlayan yöntemlerdir. Conjugate gradient yöntemleri diğer yöntemlere göre biraz daha fazla yer tutmaktadır. Bu nedenle, bu yöntemler çok sayıda ağırlığa sahip ağlar için iyidir. Conjugate gradient eğitim yöntemleri, ölçekli, Fletcher-Reeves yenilemeli geri yayılımlı ve

Polak-Riebre yenilemeli geri yayılımlı olmak üzere üç yöntemden oluşmaktadır (Hager ve Zhang, 2006).

3.1.4.3. Ölçekli Birleşik Eğim Yöntemleri

Bu yöntem 1993 yılında Moller tarafından geliştirilen ölçekli conjugate gradient yöntemi, sayısal analiz problemlerinde çok kullanılmaktadır. Diğer birçok standart geriye yayılım yöntemlerinden farklı olarak, bu yöntemde kullanıcıdan herhangi bir parametre girişi beklenmemektedir ve hesaplama gücü diğer yöntemlere göre daha yüksektir.

Ölçekli conjugate gradient yönteminde, conjugate gradient yöntemlerinde olduğu gibi satır araması gerektirmemektedir. Her bir öğrenme iterasyonunda, satır araması yapmasını önleyen bir mekanizma kullanarak diğer yöntemlerden daha hızlı çalışmaktadır. Diğer conjugate gradient yöntemlerine göre daha fazla iterasyona ihtiyaç duymaktadır, fakat işlem sayısı her iterasyonda satır taraması yapılmadığı için önemli derecede azalmıştır. Satır taraması, bir hedef fonksiyonun yerel minimum değerini bulmaya yarayan iterative bir yaklaşımdır (Moller, 1993).

3.1.4.4. Quasi- Newton Yöntemleri

Newton yöntemleri conjugate gradient yöntemlerine göre daha iyi ve hızlı optimizasyon sağlamaktadır. Newton yöntemlerinin temel basamağını, başarı ölçütündeki güncel ağırlık ve eğilim değerlerine ait Hessian matrisi (ikinci türevler) oluşturmaktadır. Newton yöntemleri conjugate gradient yöntemlerine göre daha hızlı yakınsamaktadır, fakat bu yöntemler daha karmaşıktır ve ileri beslemeli yapay sinir ağlarında Hessian matrisini hesaplamak daha fazla zaman gerektirmektedir. Newton yöntemlerinden ikinci türevin hesaplanmasını gerektirmeyenler Quasi – Newton ya da Secant yöntemleri olarak adlandırılmaktadır. Bu yöntemler algoritmalarının her bir iterasyonunda Hessian matrisini güncellemektedir. Quasi – Newton yöntemleri Broyde-Fletcher-Goldfarb- Shanno ve Levenberg-Marquardt geri yayılım yöntemlerinden oluşmaktadır.

Quasi – Newton yöntemleri ikinci türevleri kullanılamayan veya hesaplanması çok uzun işlemler gerektiren düzgün doğrusal olmayan fonksiyonların minimizasyonunda en etkili yöntemlerdir. Bu yöntemler, bir dizi arama yönü boyunca eğriliği tahmin ederek ikinci türev bilgilerini oluşturmaktadır. Bu yöntemlerden en başarılı yenileme yöntemi Broyde-Fletcher-Goldfarb-Shanno (BFGS) yöntemidir. Bu yöntemlerin çok çeşitli problemlerde başarısına rağmen, klasik Quasi – Newton yöntemleri çok büyük sayılarda iterasyon ve bazı

problemlerde fonksiyon değerlendirmesi gerektirebileceği iyi bilinmektedir (More ve diğ., 1978).

3.1.4.5. Levenberg- Marquardt Geri Yayılım Yöntemi

Levenberg– Marquardt geri yayılım yönteminde algoritmanın her iterasyonunda performans fonksiyon değerini daima azaltarak çalışan iteratif bir yöntemdir. Bu özelliği, Levenberg–Marquardt geri yayılım yöntemini orta büyüklükteki ağlarda en hızlı eğitim algoritması yapmaktadır.

Levenber – Marquardt geri yayılım yöntemi de Broyde-Fletcher-Goldfarb-Shanno yöntemi gibi fazla bellek ihtiyacı ve eğilimin ve Hessian matrisinin hesaplanmasının güçlüğü konusunda dezavantaja sahiptir (Çavuşlu ve diğ., 2012).

3.1.5. Yapay Sinir Ağları Model Yöntemlerinin Diğer Yöntemlerle Karşılaştırılması ve Değerlendirilmesi

3.1.5.1. Diğer Yöntemlere Göre Üstün Yönleri

Yapay Sinir Ağları, yapıları itibarıyla bilgi eksikliğinde dahi çalışabilmekte ve sonuçlar üretebilmektedir. Girdi bilgilerinin ağ üzerinde saklanabilmesi sayesinde münferit bilgi kaybı durumlarında yöntemin çalışması engellenememektedir. Eksik bilginin ve tam bilginin durumuna bağlı olarak performans sonuçları etkilenebilmektedir (Aras, 2008).

Yapay sinir ağları hücre yapılarında bir veya daha fazla sayıda hücrelerde meydana gelen bozulmalar çıktı üretiminde sorun yaratmaz (Saraç, 2004).

Yapay sinir ağları modelleri belirli hata payına sahiptirler. Bu özellik bu yönteme esneklik katmaktadır (Yurtoğlu, 2005).

Yapay sinir ağları geleneksel yöntemlerden farklı olarak aynı anda birden fazla işi yapabilecek çoklu ve eş zamanlı çalışabilme özelliğine sahiptir.

Yapay sinir ağları olayları öğrenebilme yeteneğine sahiptir. Bu özellik sayesinde gelecekte karşısına çıkabilecek olaylara karşı mantıklı çıkarım yapabilme ve yorumlama yeteneğini ağa verir (Bayır, 2006).

3.1.5.2. Diğer Yöntemlere Göre Zayıf Yönleri

Yapay sinir ağları ile kurulan modellerde elle çözüm oldukça zordur. Yöntemin sağlıklı sonuç vermesi donanım ve yazılıma bağlıdır (Yurtoğlu, 2005).

Yapay sinir ağlarında en önemli problemlerden birisi olarak kabul edilen durum ağın davranış durumunun açıklanamamasıdır. Hangi ağ yapısının modeli en iyi şekilde temsil ettiğinin bulunması da oldukça zordur. Bu noktada en iyi ağ yapılarının deneme-yanılma yöntemi ile bulunmaya çalışılması bu yöntemin zayıflıklarındandır.

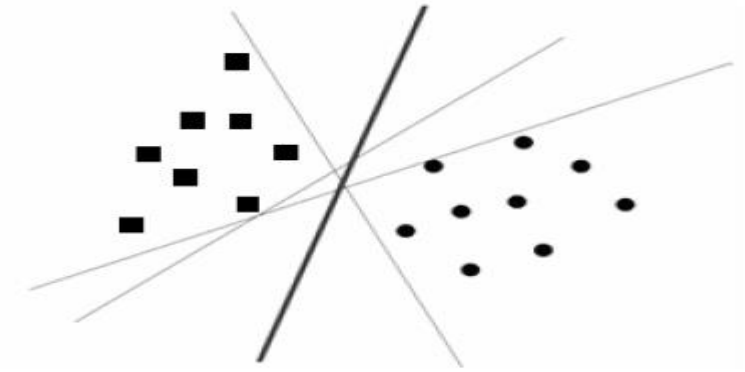
Doğru ağ yapısı kurulması için problem tüm temsil kabiliyeti ile ağa tanıtılmalıdır. Bu noktada kullanıcının yeteneğine bağlı olarak temsil gücü performansı değişir. Ağ gösterim mekanizmasının istenildiği gibi doğru kurulamaması ağın performans sonuçlarını olumsuz etkileyecektir (Öztemel, 2003).

Modeli oluşturan ağ yapısında, ağın eğitim süresinin bilinememeside ağ yapısı için de bir soru işaretidir.

3.2. DESTEK VEKTÖR MAKİNALARI

3.2.1. Destek Vektör Makinaları Yöntemi Ortaya Çıkışı ve Zaman İçerisindeki Gelişimi

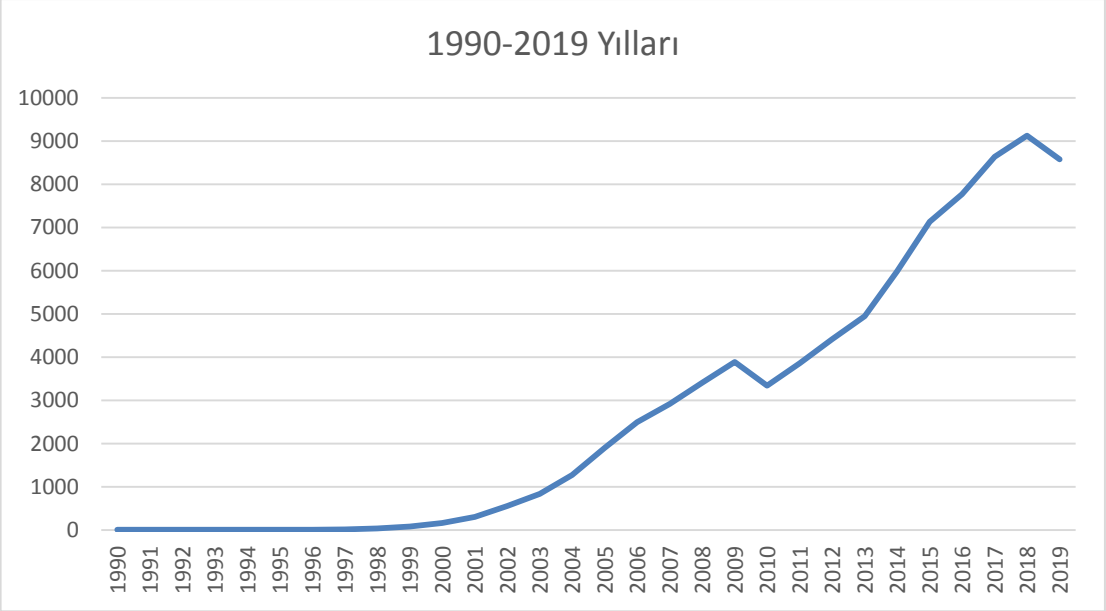
DVM ilk olarak 1979 yılında, Vapnik tarafından bulunmuştur. Destek vektör makinaları temelde 2 amaçla kullanılmaktadır. Bu amaçlar sınıflandırma ve tahminlemedir.



Şekil 3.10: Optimum ayırıcı düzlemler (Cherkassky, 1998).

Optimum düzlemler aracılığıyla düzlemler farklı tip ve çeşitlerde 2 farklı sınıfa ayrılabilmektedir.

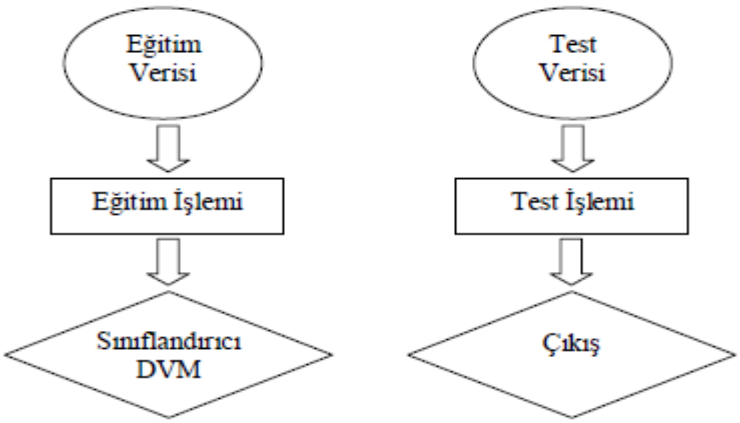
Makine öğrenmesi teknikleri içerisinde en çok veri madenciliği alanında kullanılmaktadır. Yıllar itibarıyla yöntemin uygulandığı alanlar ve çalışma sayıları artmakta, bu da yöntemin kullanım sahasının arttığı anlamına gelmektedir (www.webofknowledge.com, 2019).



Şekil 3.11: Yıl bazlı destek vektör makinalarından bahsedilen çalışmaların sayısal değişimi (www.webofknowledge.com, 2019).

3.2.2. Destek Vektör Makinaları Mimarisi ve Temel Elemanları

DVM mimarisi şekilde görüldüğü üzere tüm makine öğrenmesi sınıflandırıcıları için geçerli olan bir mimaridir. Sınıflandırma işlemlerinde, sınıflandırma işlemine tabi tutulacak verilerin bir kısmı eğitim amacıyla ayrılırken diğer bir kısmı da test amacıyla ayrılmaktadır. Eğitim verilerinin test amacıyla kullanılmaması, sınıflandırma performansının gerçek uygulamalara göre daha yüksek çıkmasını önlemektedir. DVM yönteminde, şekilde de görüldüğü gibi, öncelikle eğitim verileri kullanılarak sistem eğitilmekte ve sınıflandırıcı modeli oluşturulmaktadır. Daha sonra çıkış değerlerini önceden bildiğimiz test verileri kullanılarak sistemin çıkış değerleri belirlenir. Eldeki çıkış değerleri ile sistemin üretmiş olduğu çıkış değerleri karşılaştırılarak sistemin performansı belirlenmiş olur (www.semanticscholar.org, 2018).



Şekil 3.12: Destek vektör makinaları temel örnek mimarisi (Güraksın, 2015).

3.2.3. Destek Vektör Makinaları'nın Sınıflandırılması

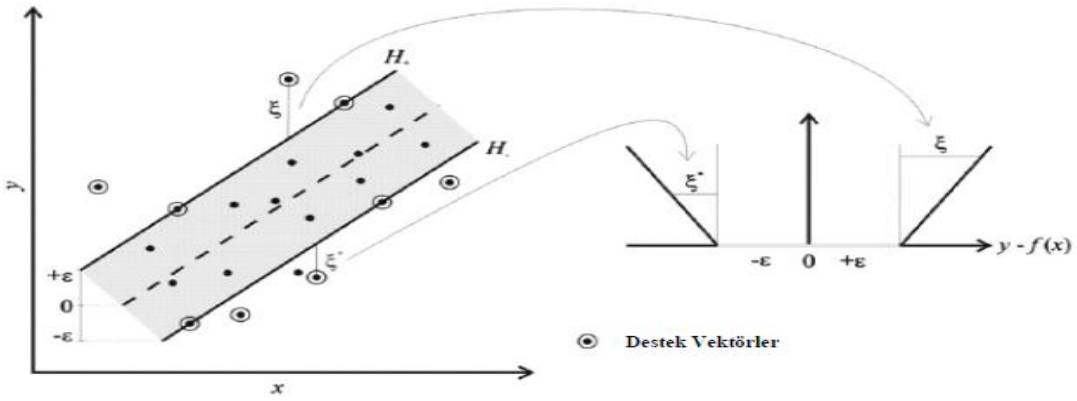
DVM için belirlenen öğrenme probleminde; yüksek boyutlu bir giriş vektörü (x) ve skaler çıkış vektörü (y) arasında bazı bilinmeyenler ve doğrusal/doğrusal olmayan bağımlılıklar vardır. Bu bağımlılıklarda y=f(x) fonksiyonu ile ifade edilir.

Destek vektör makinelerinin performansını etkileyen üç faktör mevcuttur: Hata terimi, düzenleme faktörü ve kernel parametresi ile denklemi. Model parametreleri ve kernel tipinin seçimi destek vektör makinelerinin optimum performansını etkiler (Ayhan ve Erdoğmuş, 2014).

Notasyon

Є=Hata Terimi

C=Düzenleme Faktörü



Şekil 3.13: Hata dilim değeri kayıp fonksiyonu (Moura ve diğ., 2011).

Destek vektör makinelerindeki hata terimi değişkeni olan ϵ, çözüm modelinin etkinliğinin arttırılması için gürültü faktörünün kontrol altına alınması ile görevlidir. Hata terimi değişkeni değerinin değişmesi, destek vektör sayılarının ve makina hesaplama sürelerinin değişmesine sebebiyet vermektedir. Örneğin; hata teriminin küçük seçilmesi makina tahmin hesaplama süresini ve destek vektör sayısının artışına neden olur. Tam tersi durumda da hata teriminin büyük seçilmesi, makina tahmin hesaplama süresini ve destek vektör sayısının azalmasını sağlar.

Düzenleme faktörü, eğitim hatasının minimuma indirgenmesini ve marjın maximuma çıkarılmasını amaçlamaktadır. Düzenleme faktörünün değeri modelin çözümünde ve hatanın tespitinde önemlidir. Düzenleme faktörünün yüksek olması veya başka bir ifadeyle sonsuza doğru gitmesi çözüm modelinin daha karmaşık hale gelmesine sebep olacaktır. Düzenleme faktörünün düşük olmasıyla çözüm modelinin büyük miktarda hataya göz yummasına sebep olacaktır. Aynı zamanda düşük düzenleme faktörü modelin daha az karmaşık olmasını sağlamaktadır.

Diğer önemli husus çözüm fonksiyonunun tipi ve denklemidir. Bazı kaynaklar çözüm fonksiyonunu, kernel fonksiyonu olarak da ifade etmektedir. Model çözümünde oluşturulacak destek vektörlerin hesaplama süresini ve performansını kernel fonksiyonların tipleri etkilemektedir.

Lineer destek vektör makinaları da kendi içerisinde 2 sınıfa ayrılacaktır. Bu sınıflar lineer olarak ayrılabilenler ve lineer olarak belirli bir hata payı altında ayrılabilenlerdir (Özkan, 2008).

3.2.3.1. Doğrusal Olarak Ayrılabilir Veriler

Destek vektör makinelerinin iki amacı vardır. Birinci amacı eğitim verilerinden geleceği tahmin eden F(x) fonksiyonunu bulmaktır. İkincil olarak optimizasyon problemi formülünde destek vektör regresyonu yaklaşımı oluşturmak için kullanılır (Gürsoy ve Subast, 2008).

Notasyon

xi: $Veri$ kümesi $elemanları$

yi: $Vektörlerin\ ait\ oldukları\ sınıf$

w: $Ağırlık\ vektör$ü

$||w||$ = Ağırlık vektörü olarak adlandırılan w normal düzleminin normu

$b = Eğilim\ değeri$

$ai = Lagrange\ çarpanı$

$f(x): Doğrusal\ fonksiyon$

$(x_1,y_1),(x_2,y_2),\ldots.(x_n,y_n),\quad x \in R^N,\quad y \in \{-1,+1\}$

(3.13)

$f(x)= (w.x)+b,\quad w \in R^N,\quad b \in R$

(3.14)

Sınıflandırmada y durumuna bakılarak karar alınır. Şöyle ki;

$w^T x_i+b \geq 1,\quad y_i=1$

(3.15)

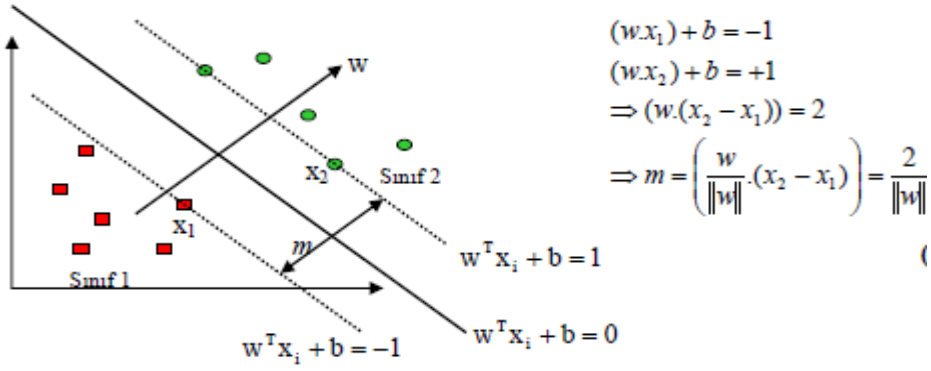
$w^T x_i+b \leq 1,\quad y_i=-1$

(3.16)

Başka bir ifadeyle

$\text{sign}((w \times xi) + b = y,\quad i = 1,..,n$

(3.17)



Şekil 3.14: Ayırıcı düzlem ve sınıflar (Kabaoğlu, 2010).

$\text{minimum}\ \frac{1}{2} \times ||w||^2$

(3.18)

Öyle ki; $y_i(w^T x_i + b) \geq 1$

(3.19)

Lagrange çarpanları ile ele alınır;

$$L(w, b, \alpha) = \frac{1}{2}\left||w|\right|^2 - \sum_i a_i(y_i(w^T x_i + b) - 1)\, a_i \geq 0\ \forall i$$

(3.20)

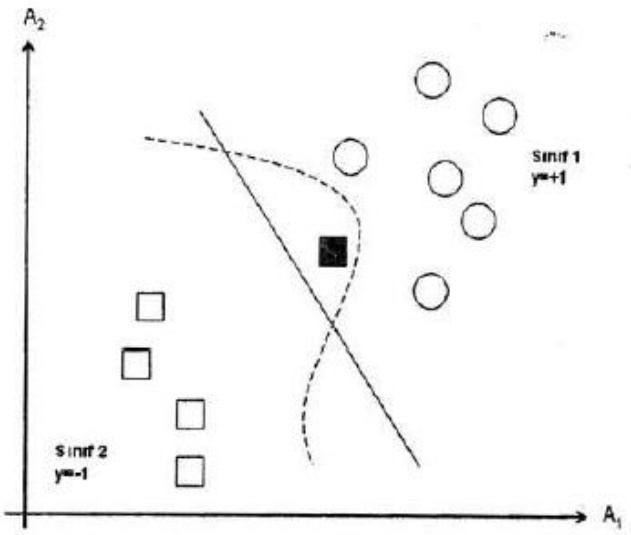
$$\max_\alpha \min_{w,b} L(w, b, a)$$

(3.21)

Vektörel w ile b minimize ve α ise maximize edilmeye çalışılır (Kabaoğlu, 2010).

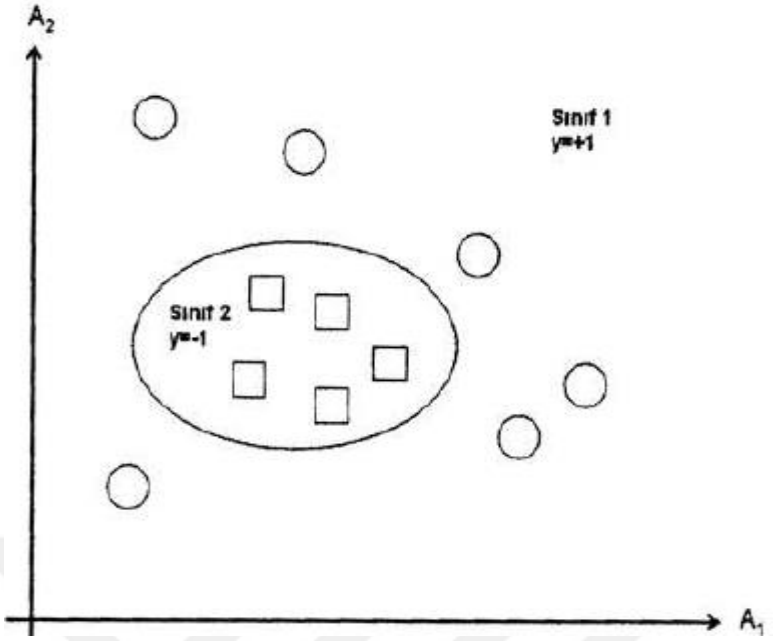
3.2.3.2. Doğrusal Ayrımı Yapılamayan Veriler

Verilerin teorik olarak doğrusal düzlemlere ayrılmadığı durumlarda lineer ayrıma engel olan veriler belirli hata oranlarında kabul edilebilir ve lineer ayrım yapılır.



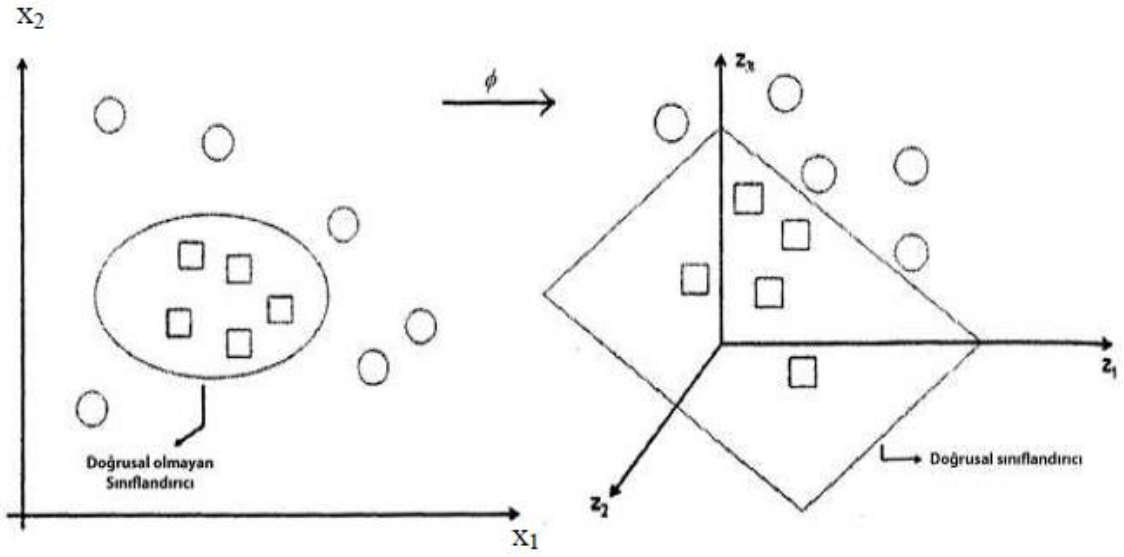
Şekil 3.15: Verilerin doğrusal olarak ayrılmama durumu (Özkan, 2008).

Verilerin lineer olarak ayrılamadığı zamanlarda ve hatta kabul edilebilir hata oranlarının ötesinde verilere sahip yapılarda, doğrusal özellikte sınıflandırıcılar kullanılamaz. Bu durumda doğrusal olmayan özelliklerdeki sınıflandırıcı yapıları kullanılmaktadır. Şekil'de lineer sınıflandırılamayacak veri kümesi örneği görülmektedir. Bunun için örnek yapılar lineer olarak ayrılabilecekleri yüksek boyutlu özelliğe sahip uzaya taşınarak yeni çözüm yeni oluşturulacak uzayda aranır (Özkan, 2008).



Şekil 3.16: Doğrusal olmayan sınıflandırma (Özkan, 2008).

Şekil 3.16'da görüldüğü üzere lineer olmayan veriler çok boyutlu uzaya taşınmış ve bir hiper düzlem ile sınıflandırılmıştır.



Şekil 3.17: Lineer olmayan verilerin üç boyutlu uzaya taşınarak sınıflandırılması (Ayhan ve Erdoğmuş, 2014).

Doğrusal olmayan verilerin ayrılmama durumlarının üstesinden gelmek için gevşek değişkenler çalışmada modellere eklenir.

ε_i: *Gevşek Değişken*

$w^T x_i + b \geq 1 - \varepsilon_i$ eğer $y_i = 1$

(3.22)

$w^T x_i + b \leq 1 + \varepsilon_i$ eğer $y_i = -1$

(3.23)

Bu noktada daha evvel bahsedilen C parametresi devreye girer. C parametresi, amaç fonksiyonuna eklenmelidir.

$$\text{minimum } \frac{1}{2} \times ||w||^2 + C\sum_{i=1}^{n} \varepsilon_i$$

(3.24)

2009'da Tso ve Mather tarafından yaptıkları çalışmada yeni Lagrange Formülünü notasyon anlamında ilkel Lagrange tabiri ile kullanmışlardır.

$$L_{ilkel} = \frac{1}{2}\left|\left|w\right|\right|^2 - \sum_{i=1}^{n} a_i(y_i(w^T x_i + b) - 1 + \varepsilon_i) + C\sum_{i=1}^{n} \varepsilon_i - \sum_{i=1}^{n} \mu_i \varepsilon_i$$

(3.25)

Verilerin uzaya taşınması için hesaplama yükünün azaltılması gerekmektedir. Bunun için kernel veya çekirdek fonksiyonları gerekmektedir.

$$K(x_i, x_j) = \varphi(_{x_i})^T \varphi(x_j)$$

(3.26)

Elde edilen yeni karar kuralları;

$$f(x) = \text{sign}(\sum_{i=1}^{n} a_i y_i K(x_i, x_j) + b)$$

(3.27)

$$f(x) = \text{sign}(\sum_{i=1}^{n} a_i y_i \varphi(_{x_i})^T \varphi(x_j) + b)$$

(3.28)

şeklinde elde edilmektedir.

Çalışma literatüründe çok farklı yapıda kernel fonksiyonları mevcuttur. Tso ve Mather 2009 yılında yaptıkları çalışmada kernel fonksyonlarını polinom, radyal ve sigmoid fonksiyonları şeklinde sıralamışlardır.

Homojen Yapıda Polinom Tip Kernel Fonksiyon:

$$K(x_i, x_j) = \beta(x_i, x_j)^d$$

(3.29)

Homojen Yapıda Polinom Tip Kernel Fonksiyon:

$$K(x_i, x_j) = (\beta(x_i, x_j) + \gamma)^d, \beta > 0 \text{ ve } \gamma > 0$$

(3.30)

Radyal Tip Kernel Fonksiyon:

$$K(x_i, x_j, \sigma) = \exp\left(-\beta \left|\left|x_i - x_j\right|\right|^2\right), \qquad \beta > 0$$

(3.31)

Gauss Radyal Tip Kernel Fonksiyon:

$$K(x_i, x_j, \sigma) = \exp\left(\frac{-\beta\left|\left|x_i - x_j\right|\right|^2}{2\sigma^2}\right), \qquad \beta > 0$$

(3.32)

(Tso ve Mather, 2009).

3.2.4. Destek Vektör Makinaları Yöntemlerinin Diğer Yöntemlerle Karşılaştırılması ve Değerlendirilmesi

3.2.4.1. Diğer Yöntemlere Göre Üstün Yönleri

Destek vektör makinaları sadece 2 boyutlu problemli çözümlerin değil, yüksek boyutlu problemlerin çözümünde de kullanılabilmektedir. Sınıflama ve kümeleme özellikleri dışında spesifik tahminleme, noktasal tahminleme özelliği de bulunabilmektedir. Karmaşık karar sınırlarına sahip problemleri modelleyebilme yeteneğine sahiptir. Yöntem, karmaşık bir

yapıya sahip sınıflama ve tahminleme problemlerinde yüksek başarı yüzdesi ile sonuç üretebilmektedir (Osowski ve diğ., 2004).

Uygulaması kolay ve aynı zamanda yüksek doğruluk oranına sahiptir. Karar verici fonksiyonlarındaki eğitim özelliği sayesinde, hafızasını verimli olarak kulanılabilmektedir (Kabaoğlu, 2010).

Veri tiplerine göre sınıflandırma işlemine tabi tutulan yöntem; doğrusal verilere uygulanabildiği gibi aynı zamanda doğrusal olmayan verilere de uygulanabilmektedir.

Çekirdek kernel fonksiyon çeşitleri sayesinde problemlerin çözümüne esneklik katarak, farklı çözüm alternatifleri meydana gelebilmektedir (Vapnik ve Chervonenkis, 1979).

3.2.4.2. Diğer Yöntemlere Göre Zayıf Yönleri

Geleneksel istatistikte, doğru modelin fonksiyonu ve formu bilinerek kabuller yapılarak çözüm sağlanır. İleri istatiksel öğrenme veya yapay zeka yöntemlerinden olan destek vektör makinalarında çözüm için gerekli en uygun çekirdek kernel fonksiyonu karar verici tarafından seçilmektedir. Bu noktada seçimde subjektiflik, modelin doğru olarak çözülmesine engel olabilir.

Çalışmada destek vektör makinalarının tahminleme özelliğinden faydalanılmaktadır. Fakat sınıflandırma problemleri için yöntem değerlendirilecek olunursa olasılıksal olarak tahminler üretememesi yöntemin zayıf yönlerindendir. Üretiği sonuçlar var veya yok yada K sınıfı veya M sınıfı gibi olabileceğinden, bazı durumlarda ara değerlere veya ara derecelere ihtiyaç duyulanabilir.

Veri setinin dağılımı hakkında varsayım veya ön bilgilerin olmaması modelin doğru çözülmesine engel olabilir (Platt, 1998).

3.3. RASTGELE ORMAN ALGORİTMASI

Rastgele orman algoritması başlıca 2 amacı gerçekleştirmek için çalışmalarda kullanılmaktadır. Bu amaçlar; ilgili verilerin doğru sınıflara ayrılmasını sağlayan sınıflandırma durumu ve ilgili verilerden doğru sonuç elde etmek için yapılan doğru tahminlerin yapılması durumlarıdır (Kormaz ve diğ., 2018). Çalışmada net kandaki glikoz

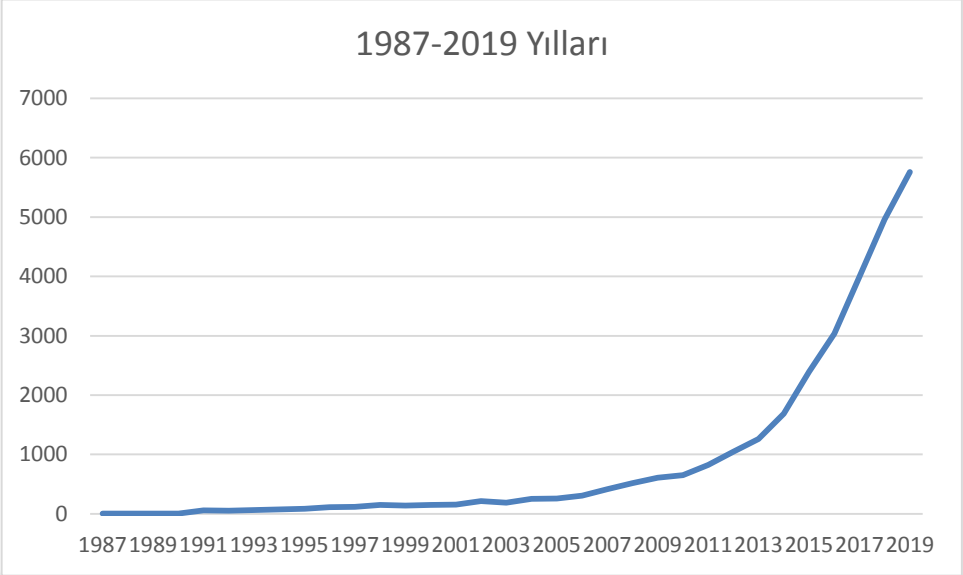
yoğunluğu değerlerini tahmin etmek amacıyla, rastgele orman algoritmasının tahmin özelliğinden yararlanılmaktadır.

Rastgele Orman Algoritması, denetimli makine öğrenme yöntemleri arasında yer alan ve hem sayısal (regresyon) hem de sayısal olmayan (sınıflandırma) bağımlı değişkenli sistemlerde kullanılabilen ağaç tabanlı bir yöntemdir. Ağaç tabanlı yöntemler, temelde, tahmin değişken uzayını ayrıştırarak bölgelere ayırırlar. Ağaç tabanlı yöntemler, veri kümesini "eğer-öyleyse (if-else)" önermeleriyle bölerek, sayısal bağımlı değişken durumunda tahmin hata kareleri toplamını; sınıflandırma durumunda hatalı sınıflandırma oranını en küçüklemeye çalışmaktadır (Kuzey, 2012).

3.3.1. Rastgele Orman Algoritmaları Yöntemi Ortaya Çıkışı ve Zaman İçerisindeki Gelişimi

Rastgele orman algoritmaları, Leo Breiman ile Adele Cutler aracılığıyla geliştirilmiştir. Genel yapısı analiz edildiğinde rastgele orman algoritması yönteminin, karar ağaçlarının geliştirilmiş versiyonu olduğu söylenebilir. Karar ağaçları mantığında, bölünme-sınıflara ayrılmada, bir değişken kullanılarak, veri tabanı 2 dala ayrılmaktadır. En iyi dallara ayırma kriteri kullanılarak hangi düğümün kök yada düğüm olarak kullanılacağına karar verir. Tespit edilen düğümün, hangi noktasında 2'ye ayrılması gerektiğini de bulmaktadır.

Makine öğrenmesi tekniklerinden olan rastgele orman algoritmalarının yıllar itibarıyla uygulandığı alanlar ve çalışma sayıları yıllar itibarıyla artmakta, 2000'li yıllardan sonra yöntemin kullanım sahasının arttığı görülmüştür (www.webofknowledge.com, 2019).



Şekil 3.18: Yıl bazlı rastgele orman algoritmalarından bahsedilen çalışmaların sayısal değişimi (www.webofknowledge.com, 2019).

3.3.2. Rastgele Orman Algoritmaları Temel Yapısı ve Uygulama Aşamaları

Rastgele orman algoritması, tahmin edici ve değişken bulunduktan sonra karar ağacı geliştirerek ilerlemektedir. Doğru olarak kurulamayan ağaçlardan oluşan yapı sonucu elde edilen başarıda önemli bir performans göstermeyecektir. İlaveten, bu ağacın standart karar ağaçlarından farklı olduğu bazı önemli noktalar vardır. Bu hususlar, rastgele orman algoritması ağacı büyütmek için mevcut tüm eğitim verilerini kullanmamaktadır. Bu yöntem yerine, torbalama yöntemini kullanarak örnekleme yapmaktadır. Diğer husus, ağaç yapımında bir düğümü bölmek için kullanılan değişken seçimini kısmen rastgele veya tamamen rastgele olarak gerçekleştirmektedir. Örnek olarak; her düğüm bölücünün tamamen rastgele seçildiği bir ağaç düşünülebilir. Rastgele orman algoritması uygulamalarında ağaçlar tamamen rastgele değildir. Fakat rastgelelik bölücü seçimini etkilemektedir. Karar ağacı son aşamada, mümkün olan en büyük boyuta getirilmektedir. Daha sonra budanmadan bırakılmakta ve bölünme/alt sınıflara ayırma işlemi başka bölünme mümkün olmayana denk tekrarlanmaktadır (Breiman, 2001).

Aşırı büyük bir karar ağacını rastgele büyütmek ve sonrasında budamadan bırakmak, akıllı bir strateji veya iyi bir modelleme uygulaması olarak görülmemektedir. Breiman, budamayı her ağaçtaki aşırı uyumu ortadan kaldırmak için topluluklarda denemiş ve yeni verilere uygulama yaptığında daha kötü sonuçlar vermiştir. Aşırı büyük ağaçlar, dinamik olarak inşa edilmiş en yakın komşu sınıflandırıcılarının başka bir varyasyonu olarak ortaya çıkmaktadır ve düşük

yanlılık - yüksek varyans özellikli tahmin edicilere sahip olmaktadırlar. Birçok tahmin edicinin ortalaması alınarak düşük yanlılık korunmakta ve yüksek varyans ortadan kaldırılmaktadır (ortalamanın her zaman yaptığı şey budur). Ağaç budama yanlılığı artıracaktır ve bu ortalama alınarak tamir edilemeyecektir. Bu nedenle budanmamış ağaçlar tercih edilmektedir (Mingers, 1989).

Rastgele orman algoritması yöntemi için InBag (in bag) ve OOB (out of bag) kavramları çok önemlidir. ROA süreci bir öğrenme veri (orijinal veri) kümesiyle başlamaktadır. Öğrenme verisinden her seferinde bir ağaç için Torbalama yöntemiyle örneklem çekilmektedir. Örneklemlerin her biri şansa bağlı olarak orijinal verinin yaklaşık % 63'lük kısmını ihtiva etmektedir ve InBag olarak adlandırılmaktadır. Başka bir ifadeyle, modelin başarısını test etmek amacıyla veri setinin yaklaşık 2/3'ü eğitim veri seti olarak algoritmanın öğrenmesi için kullanılır. Bu nedenle her ağacın örneklemi ya da InBag verisi farklı olmaktadır. Rastgele Orman Algoritması, her bir örneklem için öğrenme verisinden çekilen % 63'lük kısımdan geriye kalan % 37'lik kısmı, model performansının güvenilir tahminlerini elde etmek için test bölümünde OOB verisi olarak kullanmaktadır. Bir ROA modelinin performansı OOB verileri kullanılarak değerlendirilirken, öğrenme verilerindeki her kayıtta, aslında yetiştirilen ağaçların farklı bir alt kümesinden tahminler elde edilmektedir. Bu durum, yalnızca OOB ölçümü için geçerlidir ve ROA modelini farklı kılan bir başka özellik olarak da gösterilmektedir (Akar ve Güngör, 2012).

Rastgele orman algoritması sınıflandırmanın yanında regresyon tahmini analizinde de kullanılmaktadır. Regresyon tahminlemede amaç, kategorik veriler elde etmeden farklı olarak sürekli değere sahip bir veri elde etmektir. Oluşturulan ağaçlar devamlı yeni tahminlerde bulunur. Genel manada modelin tahminini bulmak için, her bir ağaçtan elde edilen tahminlerin ortalaması alınmalıdır (Albayrak ve Yılmaz, 2009).

Rastgele Orman Algoritması yöntemini sırasıyla aşamalara ayırarak algoritmayı uygulamak sonuç elde etme sürecini kolaylaştıracaktır. Yöntemin uygulama adımları literatür incelendiğinde bazı kaynaklarda 5, bazı kaynaklarda 6 ve bazı kaynaklarda 7 aşama/adım olarak tespit edilmiştir. Bu noktada çalışmada uygulama adımlarını 5 aşamada özetlenmesi amaçlanmıştır.

Aşamalar

1.Aşama:

İlk aşamada torbalama yöntemi ile seçim yapılacaktır. Torbalama yöntemi, bir öğrenme algoritmasının genelleme yeteneğini önemli ölçüde geliştirebilen etkili bir eğitim yöntemidir. Yöntem, makine öğrenme algoritmalarıyla birlikte kullanıldığında zaman kazandırmaktadır. Ayrıca; hata, standart hata ve parametrelerin güven aralıklarının tahmin edilmesinde popüler bir yöntemdir (Breiman, 1996).

İlk aşamada torbalama yöntemiyle N adet örneklem oluşturulmaktadır. Breiman 1996' da eğitim verisini düzenleyerek sınıflandırıcıları üretmek ve temel sınıflandırıcıların doğruluğunu geliştirmek amacıyla torbalama yöntemini geliştirmiştir. Bu metotta her bir sınıflandırıcının eğitim verisi N tane örneğin yer değiştirmesiyle rastgele olarak üretilir. N, orijinal eğitim verisinin boyutunu temsil eder. Bu durumda çoğu orijinal örnek, oluşturulan eğitim verisinde tekrarlayan olarak yer alabilirken diğerleri dışarıda kalabilir. Öğrenme sistemi örneklerden sınıfları üretir. Sınıflandırıcıyı oluşturmak için farklı denemelerde üretilen tüm sınıflandırıcıları toplar. Sınıflandırma için her bir sınıflandırıcı o nesnenin ait olduğu sınıf için bir oy kaydeder. Ve en çok oyu alan sınıfın üyesi olarak etiketlenir (Pal, 2003).

Torbalama yöntemi adımları

1.Ham (Orijinal) eğitim verisi, rastgele Ĺ eğitim verisine ve Ŧ test verisine bölünüyor. Ŧ' deki veriler orijinal veri setinde olup Ĺ de olmayan verileri içerir.

Notasyonlar

Ń=Orjinal eğitim verisinin boyutu,

n=Orjinal veri setindeki gözlem sayısı,

m=Veri setinden seçilen örneklem,

Ŧ=Test verisi (Önyükleme örnekleri),

Ĺ*=Eğitim verisi (Önyükleme örnekleri),*

Ĺ *(k) = Alt veri setleri (Alt eğitim setleri),*

βi=Önyükleme örnekleri, (β1, β2, β3 v.b.*)*

Ći= Önyükleme örnekleri için sınıflandırıcılar,

Ɵ=Önyükleme tekrarları,

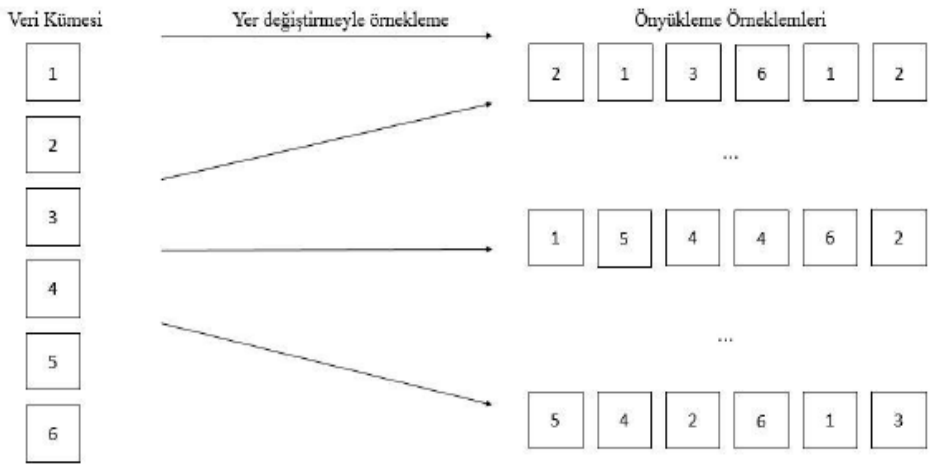
Ć=Final sınıflandırıcı,

y=Sınıf etiketi,

x=Girdi verisi.

Torbalama algoritmasında temel mantık farklı önyükleme örnekleri kullanarak üretilen sınıflandırıcıları oylamaktır.

Önyükleme yöntemi ilk kez 1979 yılında Bradley Efron tarafından geliştirilmiştir. Mevcut durumdaki veri setinden her seferde farklı örnekler seçilerek yeni veri kümeleri oluşturulur. Bu şekilde yeni veri kümeleri oluşturmaya önyükleme metodu denilmektedir. Önyükleme metodu ile örnek seçimi Şekil ’de gösterilmektedir. RO algoritmasında da eğitim veri setinden, kullanıcı tarafından belirlenen oluşturulacak ağaç sayısı kadar önyükleme örnekleme yapılmaktadır (Efron, 1979).



Şekil 3.19: Önyükleme örnek yapısı (Ayas, 2014).

Önyükleme örnekleri, yer değiştirmeli olarak eğitim veri setinden m tane örneğin birbirinden bağımsız olarak farklı şekillerde örneklenmeleri ile elde edilir. Ŧ test verisi önyükleme örnekleri (ß1, ß2, ß3 v.b.) üretildikten sonra her ßi önyükleme örnekleri için Ći sınıflandırıcıları oluşturulur. Final Sınıflandırıcı (Ć), Ći sınıflandırıcılardan elde edilir. Burada çıktının en çok tahmin edildiği sınıf dikkate alınarak işlemler yapılır. Önyükleme tekniği, veri tabanlı bir simülasyon yöntemidir. İstatiksel sonuçlar elde etmek amacıyla orjinal veriden alt veri setleri üretilir. Bu veriler önyükleme örnekleri olarak adlandırılır. Örnekleme işlemleri çok fazla sayıda tekrarlanır. (25 defa, 100 defa, 200 defa v.b.) Elde edilen tekrarlar, veriden çıkarım yapmak için kullanılabilecek bilgileri içerir.

2. Ön yükleme yöntemi kullanılarak Ĺ verisinden rastgele seçim ile yer değiştirmeli olarak birbirinden bağımsız üretilen N tane Ĺ (B) alt veri setleri oluşturulur. Bir önyükleme örneği, yer değiştirmeli olarak eğitim veri setinden m tane örneğin birbirinden bağımsız olarak farklı şekilde örneklenmesiyle elde edilir. T önyükleme örnekleri ($B1$, $B2$ … Bn) üretildikten sonra her bir Bi örnek için Ci sınıflandırıcısı oluşturulur. C* final sınıflandırıcı, oluşturulan sınıflandırıcılardan ($C1$, $C2$ … Cn) çıktının en çok tahmin edildiği sınıf baz alınarak elde edilen sınıflandırıcıdır (Bauer ve Kohavi, 1999).

3. Her bir Ĺ (B) için ağaç geliştirilir.

4. Geliştirilen ağaçlar oylanır ve bu oylara göre tahminler birleştirilir.

5. Maksimum oya göre sınıflara atama işlemi yapılır.

Akar 2013'de yaptığı çalışmada torbalama yönteminde ana fikir modellerde, ortalamaların alınması ve varyansın azaltılması olduğunu belirtmiştir. Yöntem, karar ağaçları eğer yeterli oranda büyümüşlerse ve nispeten düşük hataya sahiplerse veri setinde yer alan karmaşık ilikileri yakalama şansına sahiptir (Akar, 2013).

2.Aşama:

Bu aşamada InBag ve OOB olarak adlandırılan verilerin tasnifi yapılmaktadır. Orijinal veri setindeki gözlemlerin: 2/3'ü InBag verisi olarak örneklemde bulunurken, kalan kısmı da modelde ki hata oranınını test edilmesinde OOB verisi olarak örneklem dışında kalmaktadır.

3.Aşama:

InBag verisi ile en büyük genişlikte karar ağacı oluşturulmaktadır. Bu ağaç oluşturulurken her düğümde mevcut tüm tahmin değişkenleri içerisinden en iyi değişkenleri seçmek yerine, her düğüm bölünmesinde toplam *m* tane bağımsız değişkenden *p* tanesi rastgele seçilmektedir (*p* < *m*). Bu hususa dikkat edilmesinin sebebi ağacın çok büyük olmasının istenmemesidir. Aynı zamanda uyum anlamında yüksek başarı oranına sahip olması da istenmemektedir. *p* değişken değerlerine göre bilgi kazanç değeri maximum olan değişken ile ağacın dallara ayrılması sağlanmaktadır. Bu noktada devreye Gini indeksi veya Gini katsayısı girmektedir. Hesaplama sonucunda veri set yapısı 2 ayrı dala ayrılmaktadır. Bu süreç yeni dal kalmayana kadar, yeni dallar oluşturularak devam eder.

SVRA veya başka bir ifadeyle CART, karar ağacı oluşturmada kullanılan algoritmalardan birisidir. İkili ağaç yapısı vardır. Algoritma 3 temel aşamadan oluşur. Maximum sayıda ağaç

oluşturmak, ağaç derinliğini tespit etmek ve test verilerini ağaç üzerine uygulamak. Birinci aşama olan ağaç oluşturmak aşamasında temel amaç en saf ve en iyi bölme durumundaki maximum ağaç sayısını oluşturmaktır. Bu noktada 2 algoritma Twong ve Gini algoritmaları karşımıza çıkmaktadır. Çalışmada Gini algoritması kullanılarak değerlendirme yapılmaktadır. Gini indeksi İtalyan ekonomist Corrado Gini tarafından ilk başlarda gelir eşitsizliğini ölçmek için kullanılmış olup, her bir girdi verinin homojenliğinin ölçüldüğü bir yöntemdir. Düğümlerdeki özniteliklerin tek bir sınıfa girmesi durumunda Gini indeksi minimum ulaşır yani sıfır olur. Karar ağaçlarının bölünme kriteri, en düşük Gini indeksine göre girdi verilerini ayırmaktadır. Gini indeksinin azalması istenen bir durumdur çünkü saflığın arttığına işarettir ve bu indeksin nihai olarak sıfıra eşit olması demek maksimum saflık demektir (Tso ve Mather, 2009).

$T: Veri\ seti$

$p_j: Veri\ setinde\ yer\ alan\ sırayla\ tüm\ verilerin\ kendisinden\ küçük\ olan\ eleman\ sayılarının\ ve\ kendisinden\ büyük\ olan\ eleman\ sayılarına\ bölümü$

$n: Seçilen\ veri$

$$\text{Gini (T)} = 1 - \sum_{j=1}^{n} pj^2$$

(3.33)

4.Aşama:

Ormanı oluşturacak olan N ağaç sayısı elde edinceye kadar önceki adımlar tekrar edilmektedir. Ardından N tane ağacın ayrı ayrı yapmış olduğu sınıf tahminleri bir araya getirilerek yeni bir tahminde bulunulmaktadır. Analiz edilen gözlemlerin kategorisel anlamda sınıflandırılma sayıları hesaplanır. Yapılan gözlemlere, oy çokluklarına göre sınıf atamaları yapılır. Örnek olarak 2 kategoriden oluşan bir model yapısında, takip edilen gözlem, ağaçların en az % 50,1 oy çoğunluğunu aldığı taktirde sınıf etiketine sahip olur ve bu sınıf tahmin edilen sınıf değeri olmaktadır.

5.Aşama:

Bireysel ağaçlarda kullanılmayan OOB gözlemleri ile yapılan tahminler ise ormanın iç hata oranı kestirimini yapmak için kullanılmaktadır. Ormanı oluşturan her karar ağacının OOB hata oranı hesaplanmaktadır. Yanlış sınıflandırmanın yüzdesi rastgele orman algoritması'nın tahmin veya sınıflandırma hatası oranı olarak belirlenmektedir (Ayas, 2014).

3.3.3. Rastgele Orman Algoritmaları Yöntemlerinin Diğer Yöntemlerle Karşılaştırılması ve Değerlendirilmesi

3.3.3.1. Diğer Yöntemlere Göre Üstün Yönleri

Rastgele orman algoritması esnek veri yapısı sayesinde hızlı öğrenme gerçekleştirmektedir. Çok şıklı kategorik değişkenler de dâhil olmak üzere, sürekli ve kategorik bağımlı değişkenlerle çalışabilmektedir. Küçük veri setleri ve çok sayıda parametre ile üstün performans göstermektedir. Kayıp veri değerleri otomatik olarak ele alınmaktadır (Cutler ve diğ., 2007).

Rastgele orman algoritmasını belirli sayıdaki tahminleyici arasından uygun tahminleyicileri seçebilmektedir. Analizcinin değişken seçimi yapması veya veri azaltması gerekmemektedir. Ön işleme gerek kalmadan verileri işleyebilme özelliği bulunmaktadır. Verileri yeniden ölçeklendirme, dönüştürme, değiştirme gerektirmemektedir. Herhangi bir sütundaki aykırı değerlere karşı dayanıklılık göstermektedir (Breiman ve Cutler, 2003).

RF modelleri genellikle tek bir ağacın verdiği sonuçlardan daha doğru sonuçlar vermektedir ve çoğu zaman en iyi alternatif yöntemlerle rekabet etmektedir (Suchetana ve diğ., 2017).

Uyum konusunda, aşırı uyuma karşı dayanıklılık göstermektedir. Yöntem uygulanırken fazla sayıda rastgele orman ağacı yetiştirmek uyum riskini arttırabilir. Her ağaç tamamen bağımsız bir rastgele deneme ürünü olarak yetişmektedir.

Çözüm yapısında yer alan OOB verilerini kullanarak kendi kendini test edebilmektedir. Bu sayede rastgele orman algoritması modelleri güvenilirlik değerlendirilmesinde oldukça başarılıdır.

Rastgele orman algoritmaları hızlı çözülebilmesi ve kendilerine özgün sonuca sahip grafik görüntüleri vermesi dolayısıyla tercih edilmektedir. Bilindiği üzere yöntem uygulamasında ağaçlar hızı bir şekilde büyümektedir. Değişken sayısı az olsa bile yöntem kullanılabilmektedir (Guo ve diğ., 2017).

Rastgele orman algoritmaları sadece 2 boyutlu uzayda değil, metrik uzayda da kullanılarak ağaç tabanlı kümelenmeler oluşturmaktadır. Kümeleri oluşturan değişkenler otomatik olarak seçilmektedir ve kümeler arasında değişebilmektedir (Watts ve diğ., 2011).

3.3.3.2. Diğer Yöntemlere Göre Zayıf Yönleri

SVRA karar ağaçları modelleri sonucu ortaya çıkan tek bir karar ağacı görseli, rastgele orman algoritması modelinde elde edilememektedir. Rastgele orman algoritmaları modellerinde oldukça fazla karar ağaçları vardır.

İstatistiki modellerde elde edilen güven aralıkları, rastgele orman algoritması modellerinde elde edilmemektedir.

Rastgele orman algoritmaları yöntemlerini uygulamak için kullanılacak donanım ve yazılım önemlidir. Süreç içerisinde ve sonrasında, uygulama sonuçları elde edilirken her ağaç ve sonuçlar belleklerde saklanmaktadır. Bunun için yüksek bellek özelliğine sahip işlemciler gerekmektedir (Guo ve diğ., 2011).

3.4. PERFORMANS ÖLÇÜTLERİ

Literatür incelendiğinde performans ölçütleri olarak; tahmin başarı oranı, korelasyon katsayısı, ortalama hata kareleri toplamı kökü, Clarke glikoz hata tablosu, ortalama mutlak yüzde hatası, ortalama mutlak bağıl fark ve hata kareler ortalaması gibi performans hata ölçütlerinin çalışmalarda kullanıldığı görülmüştür.

Çalışmada 3 temel yöntemin birbirleriyle karşılaştırılmasında ve hibrit yöntemlerin birbirleriyle karşılaştırılmasında kullanılan performans kriterleri mutlak sapma, mutlak yüzde hata ve tahmin başarı oranı kriterleri kullanılmıştır.

Bir gözlem için değerlendirilecek olunursa mutlak sapma, gerçek değer ile tahmin değeri arasındaki mutlak fark olarak bulunmuştur. Gözlem sayısı artması durumunda, n gözlem sayısı için, gerçek değerler ile tahmin değerleri arasındaki mutlak farkların n gözlem sayısına bölünmesi ile elde edilen değer ortalama mutlak sapma performans ölçütü değerini vermektedir.

Mutlak yüzde hata da bir gözlem için hesaplanmak istendiğinde gerçek değer ile tahmin değeri arasındaki farkın, gerçek değere bölünmesinin 100 ile çarpılması yani yüzdesi olarak ifade edilir. Gözlem sayısı artması durumunda, n gözlem sayısı için, gerçek değerler ile tahmin değerleri arasındaki mutlak farkların, gerçek değerlere bölünmesi sonucu değerlerin

100 ile çarpılıp toplanıp, n gözlem sayısına bölünmesi ile elde edilen değer ortalama mutlak yüzde hata performans ölçütü değerini vermektedir. Elde edilen değerler modellerin performans değerlendirmelerinde kullanılır. Mutlak yüzde hata değerlerinin düşük çıkması istenmektedir.

Notasyon:
OMS: Ortalama Mutlak Sapma
OMYH: Ortalama Mutlak Yüzde Hata
n: Gözlem Sayısı
D:Gerçek Değer
F:Tahmin Değeri

$$\text{OMS} = \frac{\sum_{i=0}^{n} |D-F|}{n} \quad (3.34)$$

$$\text{OMYH} = \frac{\sum_{i=0}^{n} 100\times|D-F|/D}{n} \quad (3.35)$$

$$\text{TBO} = 100 - \frac{\sum_{i=0}^{n} 100\times|D-F|/D}{n} \quad (3.36)$$

(Kartal, 2006)

Örnek:

Gözlem no	1	2	3	4	5	6	7	8	9	10
Kan şeker ölçüm sonuçları (mg/dl)	128	195	85	77	165	146	120	88	190	144

Kan şekeri ölçüm sonuçları ile ilgili gözlemlenen hastanın kan şekeri sonuçları ve tahmin değerleri tabloda gösterilmiştir. Kan şekeri ölçüm sonuçlarının tahmin sonuçlarına göre ortalama mutlak sapma ve ortalama mutlak yüzde hata performans değerlerinin hesaplanması şu şekilde olmuştur;

Tablo 3.3: Örnek model yapısı çözüm sonuçları.

Gözlem	Ölçüm	Tahmin	Gerçek–	|Gerçek–	100*|Gerçek–	100*|Gerçek–Tahmin|/Gerçek

No	Sonucu	Sonucu	Tahmin	Tahmin|	Tahmin|	
1	128	150	-22	22	2200	17,19
2	195	130	65	65	6500	33,33
3	85	95	-10	10	1000	11,76
4	77	120	-43	43	4300	55,84
5	165	88	77	77	7700	46,67
6	146	125	21	21	2100	14,38
7	120	195	-75	75	7500	62,50
8	88	175	-87	87	8700	98,86
9	190	125	65	65	6500	34,21
10	144	100	44	44	4400	30,56

Her bir gözlem için ölçüm sonucu ve tahmin sonuçları değerleri kullanılarak OMS ve OMYH değerleri aşağıdaki eşitlikler aracılığıyla bulunmuştur.

$$\mathrm{OMS} = \frac{\sum_{i=0}^{n} |D - F|}{n}$$

$$OMS = \frac{22 + 65 + 10 + 43 + 77 + 21 + 75 + 87 + 65 + 44}{10} = \frac{509}{10} = 50{,}9$$

$$\mathrm{OMYH} = \frac{\sum_{i=0}^{n} 100 \times |D - F|/D}{n}$$

$$= \frac{17{,}19 + 33{,}33 + 11{,}76 + 55{,}84 + 46{,}67 + 14{,}38 + 62{,}5 + 98{,}86 + 34{,}21 + 30{,}56}{10}$$

$$= \frac{405{,}31}{10} = \%\ 40{,}53$$

$$\mathrm{TBO} = 1 - \frac{\sum_{i=0}^{n} 100 \times \frac{|D - F|}{D}}{n} = 100 - \frac{405{,}31}{10} = \%\ 59{,}47$$

3.5. R STUDIO

Çalışmada analizlerde kullanılan yöntemlerin çözümünde R Studio programı aracılığıyla sonuçlar elde edilmiştir. R studio, R programlama dilinin geliştirilmesi ile ortaya çıkarılmıştır. R studio, 2010 yılında geliştirilmeye başlanmış ve 2011 yılının Şubat ayında piyasaya girmiştir. R Programalama dili, açık kaynak kodlu bir yazılım ve programlama dillerinden birisi omasından dolayı çalışmada kullanımının uygun olacağı düşünülmüştür. R Programlama dili, ücretsiz olarak kullanılabilinmiştir. İstatistiksel hesap yapmak, veri analizi

yapmak, vektör uygulamaları ve matris fonksiyonları oluşturma v.b. amaçlarla kullanılmaktadır. Ücretsiz olarak https://rstudio.com/products/rstudio/download adresi aracılığıyla bilgisayara indirilip, kullanılabilir. R studio, R programlama dilinin daha etkin ve verimli olarak kullanılmasını sağlayan entegre bir yazılım olarak da değerlendirilebilir.

Çalışmada yapay sinir ağları, destek vektör makinaları, rastgele orman algoritmaları ve bu algoritmalardan elde edilen hibrit algoritmaların kullanımında R Studio programından yararlanılmıştır.

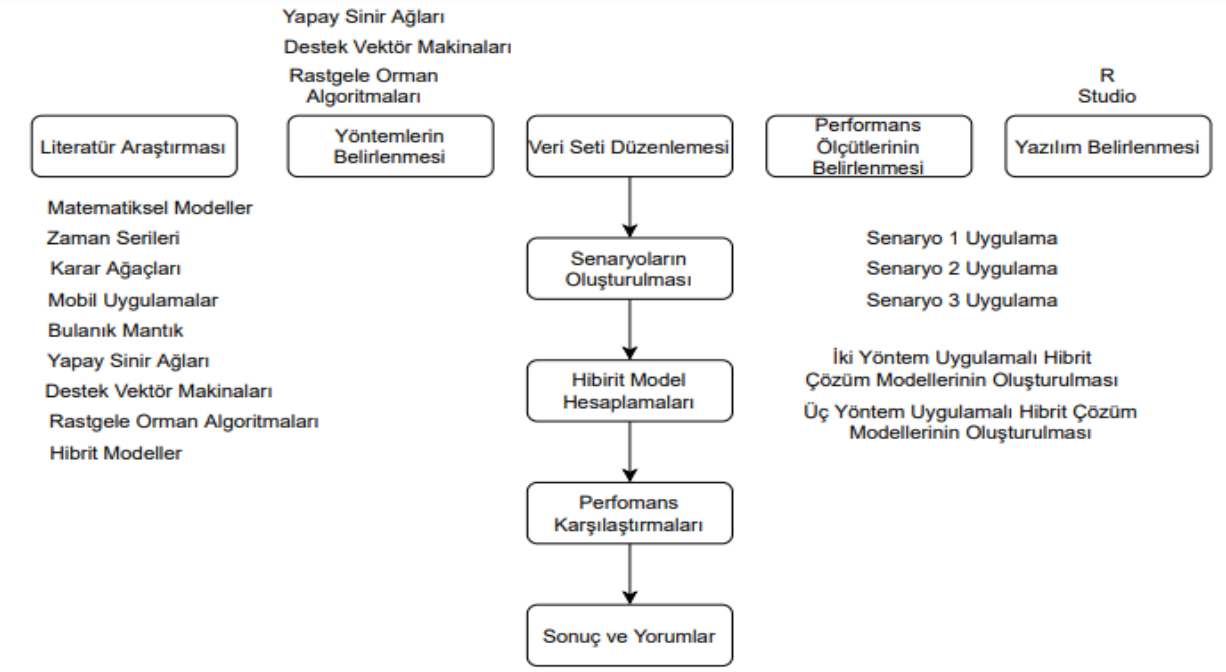
R Studio Programında, "AMORE" adı verilen kütüphane aracılığıyla yapay sinir ağlarında eğitim ve simülasyon işlemleri yapılmıştır. Rastgele orman algoritmaları kullanılarak tahminleme yapmada "e1071" isimli kütüphane kullanılmıştır. Destek vektör makinaları algortimaları uygulamada da "randomForest" kütüphanesi kullanılmış ve bu yöntemler aracılığıyla çalışma analizleri yapılmıştır (Göztepe, 2018).

4. BULGULAR

Çalışmada diyabet hastalarından elde edilmiş ve düzenlenmiş veri setlerine, veri madenciliği yöntemleri uygulanmıştır. Yapay zeka tabanlı karar destek sistemleri oluşturmada; yapay sinir ağı, rastgele orman algoritması ve destek vektör makina algoritmalarına dayalı yöntemlerinin kullanılarak yeni hibrit modellerin oluşturulması amaçlanmıştır.

Bu çalışmayı diğer çalışmalardan ayıran ve çalışmaya özgünlük katan özellik, 3 yöntemin birarada kullanılarak 27 adet farklı hibrit çözüm yöntemi elde edilmesi durumudur. Bu çözüm algoritmaları içerisinden, en iyi performansa sahip hibrit modellerin seçilmesinin uygun olacağı düşünülmüştür. Bu 3 yöntemin bir arada hibrit olarak çözüme girdiği bu durum literatürde görülmeyen bir durum olduğu gözlemlenmiş ve buna göre çözüm şekillenmiştir. Her bir veri seti için değişen senaryolar altında ayrı ayrı modeller ile parametreler oluşturulmuştur. Yapılan çözümler ile performans sonuçları birbiri ile karşılaştırılmış ve nihai sonuçlara ulaşılmıştır.

Çalışmanın bütünsel akış diyagramı şekil 4.1.'de gösterilmiştir. Literatür araştırması ile başlayan akış metadolojisi süreci yöntemlerin belirlenmesi, performans ölçütlerinin belirlenmesi, kullanılacak yazılım programlarının belirlenmesi ve çalışmada kullanılacak veri setinin düzenlenmesi ile eş zamanlı analiz edilmiştir. Bu çizilen çerçeve sonrasında alternatif senaryolar oluşturulmuş, hibrit modeller analiz edilmiş ve çalışma düzenlenmiştir.



Şekil 4.1: Çalışmanın bütünsel akış diyagramı.

4.1. VERİ SETİ DÜZENLEMESİ

Çalışmaların başarılı olması için uygulanacak yöntem ile beraber doğru ve güvenilir verilerin seçilmesi çok önemlidir. Bu noktada, çalışmada kullanılan girdi değişkenlerin oluşturulmasını sağlayan veriler Makine Öğrenimi ve Akıllı Sistemler Merkezinde yer alan diyabet hastalarının çeşitli süreler boyunca gözlemlendiği ve kayıtlar altına alındığı veri setleri havuzundan elde edilmiştir. Elde edilen veriler tekrardan hesaplamalara tutularak yeni veri setleri oluşturulmuştur.

Çalışmada kullanılan veri setindeki verilerin tiplerini ve kodlarını analiz ederek girdi değişkenleri ve çıktı değişkenleri belirlenmiştir. Gözlemci veri setini oluştururken farklı zamanlarda, farklı gözlem süreleri boyunca, farklı diyabet hastalarını incelemiş ve gerekli gördüğü ölçüm sonuçlarını kayıt altına almıştır (Kahn, 1994).

Gözlemci, diyabet hastalarını gözlemleyerek veritabanı oluşturmak için kayıt yaparken geleneksel kayıt yapma yöntemlerinden olan manuel olarak aldığı notları ve teknolojik altyapılı elektronik araçlar kullanarak yaptığı kayıtları toplamış, birleştirmiş ve düzenlemiştir. Bu kaynaklar sırasıyla otomatik elektronik kayıt cihazı ve kağıt kayıtları şeklinde

tasnifenebilir. Veri setini oluşturan dosyaların isimleri ve yapısı ile ilgili bilgiler şu şekilde özetlenmiştir;

Tablo 4.1: İnsülin ile ilgili kodlar ve açıklamaları.

İlgili Kod Numarası	Kod Numarası Açıklaması
33	*Regular Tip İnsülin*
34	*NPH Tip İnsülin*
35	*Ultralente Tip İnsülin*

Tablo 4.1'de insülin ile ilgili kodlar ve açıklamalar italik yazı tipi ile belirtilmiştir. İnsülin tipleri regular tip insulin, NPH tip insulin ve ultralente tip insulin olmak üzere 3 temel tasniflemeye tabi tutulmuş. Hastalar, durumlarına göre farklı dozajlarda farklı insulin tiplerindeki ilaçları kullanmışlardır. Farklı tiplerdeki insulinlerin etki süreleri de farklılık göstermektedir. Çalışmada bu husus dikkate alınmış ve etki sürelerine göre hesaplamalar yapılmıştır.

Veri setinde yer alan diğer bilgilerden olan kandaki glikoz yoğunluğu, kan şekeri ölçüm cihazları aracılığıyla elde edilmiştir (Kahn, 1994). Kan şekeri ölçüm cihazları, hastalara iğne batırılarak çıkarılan kanın tahlil edilmesini sağlamaktadır.

Tablo 4.2: Kandaki glikoz yoğunluğu ölçüm sonuçları ile ilgili kodlar ve açıklamaları.

İlgili Kod No	Kod Numarası Açıklaması
48	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)
57	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)
59	Kahvaltı öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)
61	Öğle yemek öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)
63	Akşam yemek öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)
64	Atıştırmalık öğün yemek öğünü öncesi kandaki glikoz yoğunluğu (mg/dl)
66	Tipik yemek öğün alımı sonrası kandaki glikoz yoğunluğu (mg/dl)
67	Normalden daha fazla yemek yenilmesi durumunda kandaki glikoz yoğunluğu (mg/dl)
68	Normalden daha az yemek yenilmesi durumunda glikoz yoğunluğu (mg/dl)

Gözlemlenen şeker hastalarında, yemek sonrası ve yemek öncesi durumlar önemli kritik durumlar olarak değerlendirilmiştir. Bu sebepten ötürü bu önemli durumlar dikkate alınarak hastalarda teşhis ve tedavi süreci yürütülmüştür. Yemek durumları öğün bazlı da farklılıklar oluşturabileceğinden hastalar; sabah kahvaltısı öncesi, sabah kahvaltısı sonrası, öğle yemeği

öncesi, öğle yemeği sonrası, akşam yemeği öncesi, akşam yemeği sonrası, atıştırmalık öğün öncesi, tipik yemek öğünü alımı sonrası, normalden daha fazla veya daha az yemek yenilmesi durumları gibi tasniflemeye tabii tutulmuştur.

Kod numaraları ve açıklamalar incelendiğinde 65 kodunun tabloda yer almadığı görülmüştür. Veri setinde yer alan 65 kodu ile ifade edilen durum hipoglisemik semptomplar durumunda kan şekeri ölçüm sonucunu göstermiştir.

Diyabet hastalarında diğer önemli husus olarak hastaların yapmış olduğu fiziksel aktiviteler olarak düşünülmüştür. Yapılan araştırmalar şeker hastalarının sadece % 13,5'lik kısmının normal insan kilosunda olduğunu, kalan % 86,5'lik kısmının ise kilolu olduğu sonucunu ortaya koymuştur. Bu da diyabet hastalarının düzensiz beslendiklerini ve fiziksel aktivite yapmadıklarını sonucunu ortaya koymuştur.

Tablo 4.3: Egzersiz / Etkinlik ile ilgili kodlar ve açıklamaları.

İlgili Kod Numarası	Kod Numarası Açıklaması
69	Normal egzersiz aktivitesi etkinliği
70	Normalden daha fazla egzersiz etkinliği
71	Normalden daha az egzersiz etkinliği
72	Belirtilmemiş tip özel etkinlik

Veri setinde yer alan fiziksel aktivite verileri 4 sınıfa ayrılmış ve yapılan gözlemler bu hususlar dikkate alınarak tamamlanmıştır. Egzersizler; normal egzersiz, normalden daha fazla egzersiz ve normalden daha az egzersiz olarak 3 sınıfta toplanmıştır. Diğer fiziksel aktivitesi de belirtilmemiş tip özel etkinlik olarak adlandırılmıştır.

Diyabet dosyaları kayıt başına dört bilgi alanından oluşmuştur. Bu bilgi alanları içerisinde tarih bilgisi, saat bilgisi, durum bilgisi ve kan şekeri / ilaç dozaj bilgisi yer almaktadır. Bu şekilde hasta bazlı gözlemler yapılmış ve kayıtlar tutulmuştur.

Diyabet dosyaları yapısı:
Tarih: Ay-Gün-Yıl biçiminde tarih
Saat: AA:BB biçiminde saat ve dakika
Durum: Kod
Kan Şekeri: Ölçüm değeri
İlaç Dozaj: Alınan ilaç

Tablo 4.4: Örnek veri seti yapısı.

Tarih	Saat	Durum	Kan Şekeri / *İlaç Dozaj*
21.04.1991	9:09	58	100
21.04.1991	*9:09*	*33*	*9*
21.04.1991	*9:09*	*34*	*13*
21.04.1991	17:08	62	119
21.04.1991	*17:08*	*33*	*7*
21.04.1991	22:51	48	123
22.04.1991	7:35	58	216
22.04.1991	*7:35*	*33*	*10*
22.04.1991	*7:35*	*34*	*13*
22.04.1991	*13:40*	*33*	*2*
22.04.1991	16:56	62	211
22.04.1991	*16:56*	*33*	*7*
23.04.1991	7:25	58	257
23.04.1991	*7:25*	*33*	*11*
23.04.1991	*7:25*	*34*	*13*
23.04.1991	17:25	62	129
23.04.1991	*17:25*	*33*	*7*
24.04.1991	7:52	58	239
24.04.1991	*7:52*	*33*	*10*

Örnek veri yapısı tablosunda normal yazı karakteri ile gösterilen 1. Satır incelenecek olunursa; tarih verisi gözlemin zamanını göstermiştir. Gözlem 21 Nisan 1991 yılında yapılmıştır. Saat veri setinde 24 saatlik dilimlerinde gözlem zamanlarının saat ve dakika cinsinden bilgileri düzenlenmiştir. 1. Satırda yapılan gözlem saat 09:09'da yapılmıştır. Durum bilgisinde ise ilgili kodlar yer almıştır. 1. satırda 58 kod durumu, yapılan gözlemin kahvaltı öğünü öncesi zamanda yapıldığını göstermiştir. Satırlarda yer alan son veri olarak kan şekeri verisi ve ilaç dozaj bilgileri, duruma göre değişmektedir. Bu 2 veri setinden sadece birisi satırlarda yer almıştır. 1. Satırda yer alan kan şekeri verisinin 100 olması kandaki glikoz konsantrasyonu ölçüm sonucunun 100 olduğu anlamına gelmiştir (Kahn, 1994).

Tablo 4.4.'de 2. ve 3. satırlarda italik yazı karakteri ile gösterilmiş 9 ve 13 değerleri ilaç dozaj miktarlarıdır. 2. satırda yer alan 33 durumu, ultralente tip insülin ilaç miktarının 9 doz olarak hastaya uygulandığını göstermiştir. 3. satırda yer alan 34 durumu, NPH tip insülin ilaç miktarının 13 doz olarak hastaya uygulandığını göstermiştir.

4.2. ALTERNATİF ÇÖZÜM YÖNTEMLERİNİN UYGULANMASI

4.2.1. Senaryo 1 Uygulama

Senaryo 1 olarak adlandırılan durum veri setinde yer alan 6 farklı hasta için kandaki glikoz yoğunluğunun tahmin edilmesi durumu olarak düşünülmüştür. Bu noktada 3 farklı girdi değişken kullanılmıştır. Girdi değişkenleri kullanılarak kan şekeri ölçüm sonucu tahmin edilmeye çalışılmıştır. Çalışmada kullanılmış girdi değişkenleri; ölçüm tarihleri arasındaki fark, kan şekeri ölçüm durumunun özelliği ve ilaç dozaj miktarıdır.

Tablo 4.5: Senaryo 1'de yer alan girdi ile çıktı değişkenleri ve açıklamaları.

Girdi/Çıktı	Değişken Açıklaması
Girdi 1	Ölçüm Tarihleri Arasındaki Fark
Girdi 2	Kan Şekeri Ölçüm Durumunun Özelliği
Girdi 3	İlaç Dozaj Miktarı
Çıktı	Kandaki Glikoz Yoğunluğu

Kan şekeri durum özelliği değişkeni; spesifik olarak, kan şekeri ölçümü yapıldığı vakit hastanın hangi durumda olduğuyla ilgili olarak düşünülmüştür. Örnek; 58 = Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)

Diğer girdi değişkeni olan ilaç dozaj miktarı değişkeni insulin tipleri ve miktarı verileri kullanılarak hesaplanmıştır. Çalışmada gözlemlenen hastalar genel olarak 3 tip ilaç kullanmıştır. Bunlar regular tip insülin, nph tip insülin ve ultralente tip insülin'dir. Çalışmada, çıktı değişkenini doğru tahmin etmek amaçlanmıştır. Çıktı değişkeni olarak kan şekeri ölçüm sonucu (kandaki glikoz konsantrasyon yoğunluğu) değeri elde edilmeye çalışılmıştır.

Tablo 4.5.'de italik yazı karakteri ile gösterilen ilaç tipleri ve miktarları görülmektedir. Bunlar regular, nph ve ultaralente ilaç miktarlarını göstermektedir. Diyabet hastalarının bu ilaçları kullanım amaçları; vücudun yemek sonrası üretmesi gereken bolus insülin ve yemek sonrası haricinde her vakit olarak vücudun üretmesi gereken bazal insülini üretememelerinden kaynaklı, kan şekeri seviyelerini dengelemektir. Kısa ve orta vadeli etki süresine göre düzenlenen bu ilaçların kullanımı ve ölçüm miktarları çalışmada ilaç dozaj miktarı olarak belirtilen girdi 3 değişkenini göstermektedir.

Tablo 4.6: Senaryo 1'de yer alan Veri Seti 1 çalışma yapısı kesiti görünümü.

Tarih	Tarihler Arası Fark (saat)	*Regular Miktar*	*NPH Miktar*	*Ultralente Miktar*	Ölçüm Tarihleri Arası Fark (dk.) (Girdi 1)	Kan Şekeri Durum Özelliği (Girdi 2)	İlaç Dozaj Miktarı (Girdi 3)	Ölçüm Sonucu (Çıktı)
21.04.1991 09:09	0	*0*	*0*	*0*	0	58	0	100
21.04.1991 17:08	07:59	*0*	*13*	*0*	479	62	13	119
21.04.1991 22:51	05:43	*7*	*13*	*0*	343	48	20	123
22.04.1991 07:35	08:44	*0*	*0*	*0*	524	58	0	216
22.04.1991 16:56	09:21	*2*	*13*	*0*	561	62	15	211
23.04.1991 07:25	14:29	*0*	*0*	*0*	869	58	0	257
23.04.1991 17:25	10:00	*0*	*13*	*0*	600	62	13	129
24.04.1991 07:52	14:27	*0*	*0*	*0*	867	58	0	239
24.04.1991 17:10	09:18	*4*	*14*	*0*	558	62	18	129
24.04.1991 22:09	04:59	*0*	*0*	*0*	299	48	0	340
25.04.1991 07:29	09:20	*0*	*0*	*0*	560	58	0	67
25.04.1991 17:24	09:55	*4*	*14*	*0*	595	62	18	206
25.04,1991 21:54	04:30	*7*	*0*	*0*	270	48	7	288
26.04.1991 05:52	07:58	*0*	*0*	*0*	478	58	0	77

Senaryo 1'de 6 hasta ve dolayısıyla 6 veri üzerinde çalışılmıştır.

Yapay sinir ağları çözüm algoritmaları ayrı ayrı her bir veri setine uygulanmış ve performans sonuçları değerlendirilmiştir. Kullanılan parametre değerleri şu şekilde özetlenmiştir;

Girdi Değişkenleri: 3 Girdi

Çıktı Değişkenleri:1 Çıktı

Ağ Modeli: Çok Katmanlı İleri Beslemeli Sinir Ağı (newff)

Momentum katsayısı: 0,5

Öğrenme hızı: 0,01

Gizli katman nöron sayısı: 2

Gizli katman aktivasyon fonksiyonu: Tanjant sigmoid transfer fonksiyonu (tansig)

Çıktı katmanı transfer fonksiyonu: Lineer transfer fonksiyonu (purelin)

Eğitim Optimizasyon Yöntemi: İniş Tabanlı Optimizayon Algoritması - Momenum Tabanlı Gradyan İniş Eğitim Yöntemi ("ADAPTgdwm": Adaptative gradient descend with momentum.)
İterasyon sayısı: 100
Yapay Sinir Ağları Eğitim Kriteri: En Küçük Ortalamalar Kareler Hatası (LMS)

Destek vektör makinaları çözüm algoritmaları ayrı ayrı her bir veri setine uygulanmış ve performans sonuçları değerlendirilmiştir. Kullanılan parametre değerleri şu şekilde özetlenmiştir;
Çekirdek (Kernel) Fonksiyonu: Radyal Tabanlı Çekirdek Fonksiyonu (radial basis function)
Radyal tabanlı fonksiyon derece değeri: 3
Gama parametresi: 0,333
Coef0 parametresi :0
Sonlandırma tolerans kriteri: 0,001
Kayıp fonksiyon: Epsilon değeri: 0.1
İşlemci ön bellek: 40 Mb

Rastgele orman algoritmaları ayrı ayrı her bir veri setine uygulanmış ve performans sonuçları değerlendirilmiştir. Kullanılan parametre değerleri şu şekilde özetlenmiştir;
Geliştirilecek ağaç sayısı: 100
Değişken sayısı: 3
Her bir düğümde kullanılmış m değişkenlerin sayısı: 2
Ölçüt kriteri: Gini katsayısı
Eğitici veri seti için ağaç sayısı: 67
Test veri seti için ağaç sayısı: 33
Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız
Minimum örnek düğüm sayısı: 2
Minimum yaprak sayısı: 1
Maximum özellik sayısı: Rastgele (random) özellik seçimi

Veri Seti 1

Veri Seti 1, 135 gün boyunca yapılan gözlemler sonucundan oluşmuştur. Veri seti 1'de başlıca 33 ve 34 kodlu olmak üzere 2 tip insulin tipi verisi ile 48, 58, 60 ve 62 kodlu olmak üzere 4 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır.

Tablo 4.7: Senaryo 1'de yer alan Veri Seti 1 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri Sayısı
33	Regular Tip İnsülin	384
34	NPH Tip İnsülin	139
48	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	78
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	134
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	54
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	102
Toplam veri sayısı		891

1 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin olmadığı görülmüştür. Destek vektör makineleri yöntemi ile ortalama mutlak sapma değeri en düşük sonucu sağlamış, rastgele orman algoritması ile de ortalama mutlak yüzde hata değeri en iyi sonucu vermiştir.

Tablo 4.8: Senaryo 1-Veri Seti 1 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	57,52873	0,164
Destek Vektör Makineleri	54,17913	0,193
Yapay Sinir Ağları	57,53458	0,245

Veri Seti 2

Veri seti 2, 95 gün boyunca yapılan gözlemler sonucundan oluşmuştur. Veri seti 2'de başlıca 33 ve 34 kodlu olmak üzere 2 tip insulin tipi verisi ile 48, 58, 60 ve 62 kodlu olmak üzere 4 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır.

Tablo 4.9: Senaryo 1'de yer alan Veri Seti 2 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri Sayısı
33	Regular Tip İnsülin	378

34	NPH Tip İnsülin	3
48	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	94
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	95
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	96
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	95
Toplam veri sayısı		761

2 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin olmadığı görülmüştür. Rastgele orman algoritması yöntemi ile ortalama mutlak sapma değeri en düşük sonucu sağlamış, yapay sinir ağları yöntemi ile ortalama mutlak yüzde hata değeri en iyi sonucu vermiştir.

Tablo 4.10: Senaryo 1-Veri Seti 2 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	57,53458	0,413
Destek Vektör Makineleri	60,17913	0,433
Yapay Sinir Ağları	58,79682	0,296

Veri Seti 3

Veri Seti 3, 38 gün boyunca yapılan gözlemler sonucundan oluşmuştur. Veri Seti 3'de başlıca 33, 34 ve 35 kodlu olmak üzere 3 tip insulin tipi verisi; 48, 57, 58, 60, 61, 62, 63, 64 ve 67 kodlu olmak üzere 9 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi ile 72 kodlu etkinlik verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır.

Tablo 4.11: Senaryo 1'de yer alan Veri Seti 3 özellikleri.

İlgili Kod	Kod Numarası Açıklaması	Veri Sayısı

Numarası		
33	Regular Tip İnsülin	60
34	NPH Tip İnsülin	1
35	Ultralente Tip İnsülin	73
48	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	94
57	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	2
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	39
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	25
61	Öğle yemek öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	5
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	30
63	Akşam yemek öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	21
64	Atıştırmalık öğün yemek öğünü öncesi kandaki glikoz yoğunluğu (mg/dl)	16
67	Normalden daha fazla yemek yenilmesi durumunda kandaki glikoz yoğunluğu (mg/dl)	10
72	Belirtilmemiş tip özel etkinlik	4
Toplam veri sayısı		380

3 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin rastgele orman algoritması yöntemi olduğu görülmüştür.

Tablo 4.12: Senaryo 1-Veri Seti 3 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	29,52774	0,296
Destek Vektör Makineleri	56,44051	0,405
Yapay Sinir Ağları	44,33432	0,343

Veri Seti 4

Veri Seti 4, 21 gün boyunca yapılan gözlemler sonucundan oluşmuştur. Veri Seti 4'de başlıca 33, 34 ve 35 kodlu olmak üzere 3 tip insulin tipi verisi; 57, 58, 60, 61, 62, 63, 64 ve 67 kodlu olmak üzere 8 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi ile 72 kodlu etkinlik verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır.

Tablo 4.13: Senaryo 1'de yer alan Veri Seti 4 özellikleri.

İlgili Kod	Kod Numarası Açıklaması	Veri

Numarası		sayısı
33	Regular Tip İnsülin	69
34	NPH Tip İnsülin	58
35	Ultralente Tip İnsülin	18
57	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	4
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	35
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	26
61	Öğle yemek öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	2
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	31
63	Akşam yemek öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	10
64	Atıştırmalık öğün yemek öğünü öncesi kandaki glikoz yoğunluğu (mg/dl)	32
67	Normalden daha fazla yemek yenilmesi durumunda kandaki glikoz yoğunluğu (mg/dl)	6
72	Belirtilmemiş tip özel etkinlik	1
Toplam veri sayısı		292

4 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin rastgele orman algoritması yöntemi olduğu görülmüştür.

Tablo 4.14: Senaryo 1-Veri Seti 4 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	37,34053	0,331
Destek Vektör Makineleri	65,87639	0,478
Yapay Sinir Ağları	53,41128	0,337

Veri Seti 5

Veri Seti 5, 41 gün boyunca yapılan gözlemler sonucundan oluşmuştur. Veri Seti 5'de başlıca 33 ve 34 kodlu olmak üzere 2 tip insulin tipi verisi ile 57, 58, 60, 61, 62, 63, 64 ve 67 kodlu olmak üzere 8 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır.

Tablo 4.15: Senaryo 1'de yer alan Veri Seti 5 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri sayısı

33	Regular Tip İnsülin	72
34	NPH Tip İnsülin	76
57	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	5
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	37
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	25
61	Öğle yemek öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	1
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	32
63	Akşam yemek öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	7
64	Atıştırmalık öğün yemek öğünü öncesi kandaki glikoz yoğunluğu (mg/dl)	36
67	Normalden daha fazla yemek yenilmesi durumunda kandaki glikoz yoğunluğu (mg/dl)	2
Toplam veri sayısı		293

5 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin olmadığı görülmüştür. Rastgele orman algoritması yöntemi ile ortalama mutlak sapma değeri en düşük sonucu sağlarken, yapay sinir ağları yöntemi ile ortalama mutlak yüzde hata değeri en iyi sonucu vermiştir.

Tablo 4.16: Senaryo 1-Veri Seti 5 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	39,51559	0,030
Destek Vektör Makineleri	61,7353	0,034
Yapay Sinir Ağları	56,97098	0,021

Veri Seti 6

Veri Seti 6, 25 gün boyunca yapılan gözlemler sonucundan oluşmuştur. Veri Seti 6'da başlıca 33 ve 34 kodlu olmak üzere 2 tip insulin tipi verisi ile 48, 57, 58, 60 ve 62 kodlu olmak üzere 5 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır.

Tablo 4.17: Senaryo 1'de yer alan Veri Seti 6 özellikleri.

İlgili Kod	Kod Numarası Açıklaması	Veri sayısı

Numarası		
33	Regular Tip İnsülin	59
34	NPH Tip İnsülin	25
48	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	9
57	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	5
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	20
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	18
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	17
Toplam veri sayısı		153

5 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin olmadığı görülmüştür. Rastgele orman algoritması yöntemi ile ortalama mutlak sapma değeri en düşük sonucu sağlarken, yapay sinir ağları yöntemi ile ortalama mutlak yüzde hata değeri en iyi sonucu vermiştir.

Tablo 4.18: Senaryo 1-Veri Seti 6 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	40,90506	0,236
Destek Vektör Makineleri	47,60738	0,266
Yapay Sinir Ağları	43,94098	0,223

Veriler, yöntemler açısından değerlendirildiğinde ortalama mutlak yüzde hata oranının en küçük çıktığı yöntem yapay sinir ağları olarak tespit edimiştir. Buna karşın ortalama mutlak sapma performans kriteri açısından değerlendirildiğinde rastgele orman algoritması yönteminin diğer yöntemlere göre daha başarılı olduğu ortaya çıkmıştır.

Tablo 4.19: Senaryo 1-Tüm veri setlerinin performans sonuçları.

Yöntemler	Rastgele Orman Algoritması		Destek Vektör Makineleri		Yapay Sinir Ağları	
Veri_Setleri	Mutlak_Sapma	Mutlak Yüzde Hata	Mutlak_Sapma	Mutlak Yüzde Hata	Mutlak_Sapma	Mutlak Yüzde Hata
Veri_1	57,53	0,17	54,18	0,19	57,54	0,25
Veri_2	57,53	0,41	60,18	0,43	58,80	0,30

Veri_3	29,53	0,30	56,44	0,41	44,33	0,34
Veri_4	37,34	0,33	65,87	0,48	53,41	0,34
Veri_5	39,52	0,03	61,74	0,03	56,97	0,02
Veri_6	40,91	0,24	47,60	0,27	43,94	0,22

Senaryo 1 kabulleri altında kullanılan yöntemler analiz edildiğinde mutlak sapma performans kriteri için en başarılı sonuçlar veren algoritmanın rastgele orman algoritması olduğu sonucuna varılmıştır. Mutlak yüzde hata performans kriteri için en başarılı yöntemin de yapay sinir ağları yöntemi olduğu sonucu ortaya çıkmıştır.

4.2.2. Senaryo 2 Uygulama

Senaryo 2 olarak adlandırılan durum veri setinde yer alan 6 farklı hasta için kandaki glikoz yoğunluğunun tahmin edilmesi durumu olarak düşünülmüştür. Bu noktada 3 farklı girdi değişken kullanılmış, girdi değişkenlerinin sonucunda kan şekeri ölçüm sonucu tahmin edilmeye çalışılmıştır. Çalışmada kullanılan girdi değişkenleri; ölçüm tarihleri arasındaki fark kümülatifi, kan şekeri ölçüm durumunun özelliği ve ilaç dozaj miktarı olarak düşünülmüştür.

Tablo 4.20: Senaryo 2'de yer alan girdi ile çıktı değişkenleri ve açıklamaları.

Girdi/Çıktı	Değişken Açıklaması
Girdi 1	Ölçüm Tarihleri Arasındaki Farkın Kümülatifi
Girdi 2	Kan Şekeri Ölçüm Durumunun Özelliği
Girdi 3	İlaç Dozaj Miktarı
Çıktı	Kandaki Glikoz Yoğunluğu

Senaryo 2 ile senaryo 1 uygulamalarında temel fark ölçüm tarihleri arasındaki fark kümülatifi değişkeninin çalışmada kullanımı olmuştur. Diğer değişkenlerde durum senaryo 1'deki yapı ile aynı olarak kalmıştır. Senaryo 2'de kullanılan verilerde senaryo 1'de kullanılan veriler olarak değerlendirilmiştir.

Tablo 4.21: Senaryo 2'de yer alan veri seti çalışma yapısı görünümü.

Tarih	Tarihler Arası Fark (saat)	*Regular Miktar*	*NPH Miktar*	*Ultralente Miktar*	Ölçüm Tarihleri Arası Fark (dk.)	Ölçüm Tarihleri Arası Farkın Kümülatifi (Girdi 1)	Kan Şekeri Durum Özelliği (Girdi 2)	İlaç Dozaj Miktarı (Girdi 3)	Ölçüm Sonucu (Çıktı)
21.04.1991 09:09	0	*0*	*0*	*0*	0	0	58	0	100

21.04.1991 17:08	07:59	*0*	*13*	*0*	479	479	62	13	119
21.04.1991 22:51	05:43	*7*	*13*	*0*	343	822	48	20	123
22.04.1991 07:35	08:44	*0*	*0*	*0*	524	1346	58	0	216
22.04.1991 16:56	09:21	*2*	*13*	*0*	561	1907	62	15	211
23.04.1991 07:25	14:29	*0*	*0*	*0*	869	2776	58	0	257
23.04.1991 17:25	10:00	*0*	*13*	*0*	600	3376	62	13	129
24.04.1991 07:52	14:27	*0*	*0*	*0*	867	4243	58	0	239
24.04.1991 17:10	09:18	*4*	*14*	*0*	558	4801	62	18	129
24.04.1991 22:09	04:59	*0*	*0*	*0*	299	5100	48	0	340
25.04.1991 07:29	09:20	*0*	*0*	*0*	560	5660	58	0	67
25.04.1991 17:24	09:55	*4*	*14*	*0*	595	6255	62	18	206
25.04,1991 21:54	04:30	*7*	*0*	*0*	270	6525	48	7	288
26.04.1991 05:52	07:58	*0*	*0*	*0*	478	7003	58	0	77

Senaryo 2'de de 6 hasta, yani 6 veri seti yapısı ele alınmıştır. Yapay sinir ağları çözüm algoritmaları ayrı ayrı her bir veri setine uygulanmış ve performans sonuçları değerlendirilmiştir. Kullanılan parametre değerleri şu şekilde özetlenmiştir;

Girdi Değişkenleri: 3 Girdi

Çıktı Değişkenleri:1 Çıktı

Ağ Modeli: Çok Katmanlı İleri Beslemeli Sinir Ağı (newff)

Momentum katsayısı: 0,5

Öğrenme hızı: 0,01
Gizli katman nöron sayısı: 2
Gizli katman aktivasyon fonksiyonu: Tanjant sigmoid transfer fonksiyonu (tansig)
Çıktı katmanı transfer fonksiyonu: Lineer transfer fonksiyonu (purelin)
Eğitim Optimizasyon Yöntemi: İniş Tabanlı Optimizayon Algoritması - Momenum Tabanlı Gradyan İniş Eğitim Yöntemi ("ADAPTgdwm": Adaptative gradient descend with momentum.)
İterasyon sayısı: 100
Yapay Sinir Ağları Eğitim Kriteri: En Küçük Ortalamalar Kareler Hatası (LMS)

Destek vektör makinaları çözüm algoritmaları ayrı ayrı her bir veri setine uygulanmış ve performans sonuçları değerlendirilmiştir. Kullanılan parametre değerleri şu şekilde özetlenmiştir;
Çekirdek (Kernel) Fonksiyonu: Radyal Tabanlı Çekirdek Fonksiyonu (radial basis function)
Radyal tabanlı fonksiyon derece değeri: 3
Gama parametresi: 0,333
Coef0 parametresi :0
Sonlandırma tolerans kriteri: 0,001
Kayıp fonksiyon: Epsilon değeri: 0.1
İşlemci ön bellek: 40 Mb

Rastgele orman algoritmaları ayrı ayrı her bir veri setine uygulanmış ve performans sonuçları değerlendirilmiştir. Kullanılan parametre değerleri şu şekilde özetlenmiştir;
Geliştirilecek ağaç sayısı: 100
Değişken sayısı: 3
Her bir düğümde kullanılmış m değişkenlerin sayısı: 2
Ölçüt kriteri: Gini katsayısı
Eğitici veri seti için ağaç sayısı: 67
Test veri seti için ağaç sayısı: 33
Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız
Minimum örnek düğüm sayısı: 2
Minimum yaprak sayısı: 1

Maximum özellik sayısı: Rastgele (random) özellik seçimi

Veri Seti 1

1 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin rastgele orman algoritması yöntemi olduğu görülmüştür.

Tablo 4.22: Senaryo 2-Veri Seti 1 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	41,03424	0,156
Destek Vektör Makineleri	60,36985	0,193
Yapay Sinir Ağları	49,77654	0,195

Veri Seti 2

2 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin rastgele orman algoritması yöntemi olduğu görülmüştür.

Tablo 4.23: Senaryo 2-Veri Seti 2 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	51,10111	0,150
Destek Vektör Makineleri	60,98443	0,429
Yapay Sinir Ağları	56,56481	0,345

Veri Seti 3

3 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin rastgele orman algoritması yöntemi olduğu görülmüştür.

Tablo 4.24: Senaryo 2-Veri Seti 3 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata

Rastgele Orman Algoritması	27,8792	0,279
Destek Vektör Makineleri	57,10463	0,406
Yapay Sinir Ağları	36,5642	0,384

Veri Seti 4

4 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin rastgele orman algoritması yöntemi olduğu görülmüştür.

Tablo 4.25: Senaryo 2-Veri Seti 4 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	36,77541	0,332
Destek Vektör Makineleri	67,42327	0,484
Yapay Sinir Ağları	45,75496	0,346

Veri Seti 5

5 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin olmadığı görülmüştür. Rastgele orman algoritması yöntemi ile mutlak sapma değeri en düşük sonucu sağlamış, yapay sinir ağları yöntemi ile mutlak yüzde hata değeri en iyi sonucu vermiştir.

Tablo 4.26: Senaryo 2-Veri Seti 5 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	38,75158	0,030
Destek Vektör Makineleri	39,47461	0,021
Yapay Sinir Ağları	45,15864	0,012

Veri Seti 6

6 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin rastgele orman algoritması yöntemi olduğu görülmüştür.

Tablo 4.27: Senaryo 2-Veri Seti 6 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	34,84527	0,198
Destek Vektör Makineleri	47,2763	0,271
Yapay Sinir Ağları	37,16783	0,219

Veriler, yöntemler açısından değerlendirildiğinde ortalama mutlak yüzde hata oranının en küçük çıktığı yöntem yapay sinir ağları olarak tespit edimiştir. Buna karşın ortlama mutlak sapma performans kriteri açısından değerlendirildiğinde rastgele orman algoritması yönteminin diğer yöntemlere göre daha başarılı olduğu ortaya çıkmıştır.

Tablo 4.28: Senaryo 2-Tüm veri setlerinin ortalama performans sonuçları.

Yöntemler	Rastgele Orman Algoritması		Destek Vektör Makineleri		Yapay Sinir Ağları	
Veri_Setleri	Mutlak_Sapma	Mutlak Yüzde Hata	Mutlak_Sapma	Mutlak Yüzde Hata	Mutlak_Sapma	Mutlak Yüzde Hata
Veri_1	41,03	0,16	60,37	0,19	49,78	0,20
Veri_2	51,10	0,15	60,98	0,43	56,57	0,35
Veri_3	27,88	0,28	57,10	0,41	36,56	0,38
Veri_4	36,78	0,33	67,42	0,48	45,75	0,35
Veri_5	38,75	0,03	39,47	0,02	45,16	0,01
Veri_6	34,85	0,20	47,28	0,27	37,17	0,22

Senaryo 2 çözümleri, Senaryo 1 ile karşılaştırıldığında hem ortalama mutlak sapma hem de ortalama mutlak yüzde performans kriterlerinde daha iyi sonuçlar elde edilmiştir. Senaryo 2 kabulleri altında kullanılan yöntemler analiz edildiğinde mutlak sapma performans kriteri için en başarılı sonuçlar veren algoritmanın rastgele orman algoritması olduğu sonucuna varılmıştır. Mutlak yüzde hata performans kriteri için en başarılı yöntemin de yine rastgele orman algoritması yöntemi olduğu sonucu ortaya çıkmıştır.

Senaryo 1 ve senaryo 2 kabul şartları altında değerlendirildiğinde, senaryo 2'de meydana gelen başarı oranında iyileşmeden kaynaklı olarak kullanılacak değişkenlerin özellik seçimi kararlaştırılmıştır. Bu değişkenler senaryo 3 'de de kullanılmıştır.

4.2.3. Senaryo 3 Uygulama

Çalışmanın üçüncü senaryosunda; kan şekeri verilerinde meydana gelen alt ve üst uç değerler gözlemlerin dışında bırakılarak, ikinci senaryoda kullanılan girdi ve çıktı değişkenleri

kullanılmıştır. Kan şekeri ölçüm sonuçlarından, 70-200 arasında olan veriler değerlendirilerek veri seti yeniden düzenlenmiştir. Bu durumun tercih edilmesinin sebebi, şeker hastalarında meydana gelebilecek olası tehlikeli durumları engellemektir. Kan şekerinin 70 mg/dl'nin altına düşmesi hastalarda hipoglisemik riskleri arttırır. Benzer şekilde kan şekerinin 200 mg/dl'nin üzerine çıkması da hiperglisemik riskleri arttırmaktadır. Hastalarda meydana gelebilecek bu risklerin önüne geçmek için, hiperglisemi ve hipoglisemi semptomları gösteren hastalara acil müdahale edilmesi gerekmektedir. Bu durumdaki hastalar için tahminleme yapmaktan ziyade gerçek kan şekeri ölçümü yapılmasının daha doğru olacağı düşünülmüştür.

Tablo 4.29: Senaryo 3'de yer alan girdi ile çıktı değişkenleri ve açıklamaları.

Girdi/Çıktı	Değişken Açıklaması
Girdi 1	Ölçüm Tarihleri Arasındaki Farkın Kümülatifi
Girdi 2	Kan Şekeri Ölçüm Durumunun Özelliği
Girdi 3	İlaç Dozaj Miktarı
Çıktı	Kandaki Glikoz Yoğunluğu (70-200)

Yapay sinir ağları çözüm algoritmaları ayrı ayrı her bir veri setine uygulanmış ve performans sonuçları değerlendirilmiştir. Kullanılan parametre değerleri şu şekilde özetlenmiştir;
Girdi Değişkenleri: 3 Girdi
Çıktı Değişkenleri:1 Çıktı
Ağ Modeli: Çok Katmanlı İleri Beslemeli Sinir Ağı (newff)
Momentum katsayısı: 0,5
Öğrenme hızı: 0,01
Gizli katman nöron sayısı: 2
Gizli katman aktivasyon fonksiyonu: Tanjant sigmoid transfer fonksiyonu (tansig)
Çıktı katmanı transfer fonksiyonu: Lineer transfer fonksiyonu (purelin)
Eğitim Optimizasyon Yöntemi: İniş Tabanlı Optimizayon Algoritması - Momenum Tabanlı Gradyan İniş Eğitim Yöntemi ("ADAPTgdwm": Adaptative gradient descend with momentum.)
İterasyon sayısı: 100
Yapay Sinir Ağları Eğitim Kriteri: En Küçük Ortalamalar Kareler Hatası (LMS)

Destek vektör makinaları çözüm algoritmaları ayrı ayrı her bir veri setine uygulanmış ve performans sonuçları değerlendirilmiştir. Kullanılan parametre değerleri;
Çekirdek (Kernel) Fonksiyonu: Radyal Tabanlı Çekirdek Fonksiyonu (radial basis function)

Radyal tabanlı fonksiyon derece değeri: 3
Gama parametresi: 0,333
Coef0 parametresi :0
Sonlandırma tolerans kriteri: 0,001
Kayıp fonksiyon: Epsilon değeri: 0.1
İşlemci ön bellek: 40 Mb

Rastgele orman algoritmaları ayrı ayrı her bir veri setine uygulanmış ve performans sonuçları değerlendirilmiştir. Kullanılan parametre değerleri şu şekilde özetlenmiştir;
Geliştirilecek ağaç sayısı: 100
Değişken sayısı: 3
Her bir düğümde kullanılmış m değişkenlerin sayısı: 2
Ölçüt kriteri: Gini katsayısı
Eğitici veri seti için ağaç sayısı: 67
Test veri seti için ağaç sayısı: 33
Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız
Minimum örnek düğüm sayısı: 2
Minimum yaprak sayısı: 1
Maximum özellik sayısı: Rastgele (random) özellik seçimi

Düzenlenmiş Veri Seti 1

Kan şekeri ölçüm değeri sonuçlarının 70-200 arasında olanları analizlerde kullanılmıştır. Düzenlenen veri seti 1, 135 gün boyunca yapılan gözlemler sonucundan oluşmuştur. Veri Seti 1'de başlıca 33 ve 34 kodlu olmak üzere 2 tip insulin tipi verisi ile 48, 58, 60 ve 62 kodlu olmak üzere 4 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır.

Tablo 4.30: Senaryo 3'de yer alan Veri Seti 1 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri sayısı
33	Regular Tip İnsülin	384
34	NPH Tip İnsülin	139
48	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	51

58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	82
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	36
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	67
Toplam veri sayısı		759

1 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin rastgele orman algoritması yöntemi olduğu görülmüştür.

Tablo 4.31: Senaryo 3-Veri Seti 1 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	24,33137	0,209
Destek Vektör Makineleri	30,76626	0,259
Yapay Sinir Ağları	30,25732	0,238

Düzenlenmiş Veri Seti 2

Kan şekeri ölçüm değeri sonuçlarının 70-200 arasında olanları analizlerde kullanılmıştır. Düzenlenmiş veri seti 2, 95 gün boyunca yapılan gözlemler sonucundan oluşmuştur. Veri Seti 2'de başlıca 33 ve 34 kodlu olmak üzere 2 tip insulin tipi verisi ile 48, 58, 60 ve 62 kodlu olmak üzere 4 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır.

Tablo 4.32: Senaryo 3'de yer alan Veri Seti 2 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri sayısı
33	Regular Tip İnsülin	378
34	NPH Tip İnsülin	3
48	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	46

58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	43
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	43
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	50
Toplam veri sayısı		563

Veriler, yöntemler açısından değerlendirildiğinde ortalama mutlak yüzde hata oranının en küçük çıktığı yöntem yapay sinir ağları olarak tespit edimiştir. Buna karşın ortlama mutlak sapma performans kriteri açısından değerlendirildiğinde rastgele orman algoritması yönteminin diğer yöntemlere göre daha başarılı olduğu ortaya çıkmıştır.

Tablo 4.33: Senaryo 3-Veri Seti 2 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	23,05592	0,173
Destek Vektör Makineleri	27,1615	0,209
Yapay Sinir Ağları	25,96246	0,173

Düzenlenmiş Veri Seti 3

Kan şekeri ölçüm değeri sonuçlarının 70-200 arasında olanları analizlerde kullanılmıştır. Düzenlenmiş veri seti 3, 38 gün boyunca yapılan gözlemler sonucundan oluşmuştur. Veri Seti 3'de başlıca 33, 34 ve 35 kodlu olmak üzere 3 tip insulin tipi verisi ile 57, 58, 60, 61, 62, 63 ve 64 kodlu olmak üzere 7 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi aracılığıyla modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır.

Tablo 4.34: Senaryo 3'de yer alan Veri Seti 3 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri sayısı
33	Regular Tip İnsülin	60
34	NPH Tip İnsülin	1
35	Ultralente Tip İnsülin	73
57	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	2
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	30

60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	12
61	Öğle yemek öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	5
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	18
63	Akşam yemek öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	10
64	Atıştırmalık öğün yemek öğünü öncesi kandaki glikoz yoğunluğu (mg/dl)	14
Toplam veri sayısı		225

3 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin rastgele orman algoritması yöntemi olduğu görülmüştür.

Tablo 4.35: Senaryo 3-Veri Seti 3 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	19,97457	0,176
Destek Vektör Makineleri	29,55393	0,261
Yapay Sinir Ağları	26,41126	0,209

Düzenlenmiş Veri Seti 4

Kan şekeri ölçüm değeri sonuçlarının 70-200 arasında olanları analizlerde kullanılmıştır. Düzenlenmiş veri seti 4, 21 gün boyunca yapılan gözlemler sonucundan oluşmuştur. Veri Seti 4'de başlıca 33, 34 ve 35 kodlu olmak üzere 3 tip insulin tipi verisi; 57, 58, 60, 61, 62 ve 64 kodlu olmak üzere 6 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır.

Tablo 4.36: Senaryo 3'de yer alan Veri Seti 4 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri sayısı
33	Regular Tip İnsülin	69
34	NPH Tip İnsülin	58
35	Ultralente Tip İnsülin	18
57	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	2

58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	24
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	17
61	Öğle yemek öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	2
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	12
64	Atıştırmalık öğün yemek öğünü öncesi kandaki glikoz yoğunluğu (mg/dl)	23
Toplam veri sayısı		225

4 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin rastgele orman algoritması yöntemi olduğu görülmüştür.

Tablo 4.37: Senaryo 3-Veri Seti 4 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	23,06173	0,199
Destek Vektör Makineleri	33,02368	0,286
Yapay Sinir Ağları	30,62591	0,232

Düzenlenmiş Veri Seti 5

Kan şekeri ölçüm değeri sonuçlarının 70-200 arasında olanları analizlerde kullanılmıştır. Düzenlenmiş veri seti 5, 41 gün boyunca yapılan gözlemler sonucundan oluşmuştur. Veri Seti 5'de başlıca 33 ve 34 kodlu olmak üzere 2 tip insulin tipi verisi ile 57, 58, 60, 61, 62 ve 64 kodlu olmak üzere 6 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır.

Tablo 4.38: Senaryo 3'de yer alan Veri Seti 5 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri sayısı
33	Regular Tip İnsülin	72
34	NPH Tip İnsülin	76
57	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	2

58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	26
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	19
61	Öğle yemek öğünü sonrası zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	1
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	11
64	Atıştırmalık öğün yemek öğünü öncesi kandaki glikoz yoğunluğu (mg/dl)	26
Toplam veri sayısı		233

5 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin rastgele orman algoritması yöntemi olduğu görülmüştür.

Tablo 4.39: Senaryo 3-Veri Seti 5 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	20,15074	0,169
Destek Vektör Makineleri	26,49273	0,227
Yapay Sinir Ağları	24,59988	0,177

Düzenlenmiş Veri Seti 6

Kan şekeri ölçüm değeri sonuçlarının 70-200 arasında olanları analizlerde kullanılmıştır. Düzenlenmiş veri seti 6, 25 gün boyunca yapılan gözlemler sonucundan oluşmuştur. Veri seti 6'da başlıca 33 ve 34 kodlu olmak üzere 2 tip insulin tipi verisi ile 48, 58, 60 ve 62 kodlu olmak üzere 4 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır.

Tablo 4.40: Senaryo 3'de yer alan Veri Seti 6 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri sayısı
33	Regular Tip İnsülin	59

34	NPH Tip İnsülin	25
48	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	2
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	2
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	14
962	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	7
Toplam veri sayısı		109

6 numaralı hasta için kullanılan yöntemlere göre sonuç performansları karşılaştırılmak istendiğinde, 2 performans kriteri içinde en iyi sonucu veren yöntemin rastgele orman algoritması yöntemi olduğu görülmüştür.

Tablo 4.41: Senaryo 3-Veri Seti 6 yöntemlerin performans sonuçları.

Kullanılan Yöntem	Ortalama Mutlak Sapma	Ortalama Mutlak Yüzde Hata
Rastgele Orman Algoritması	18,33125	0,139
Destek Vektör Makineleri	22,68027	0,173
Yapay Sinir Ağları	20,69964	0,141

Veriler, yöntemler açısından değerlendirildiğinde ortalama mutlak yüzde hata oranının en küçük çıktığı yöntem yapay sinir ağları olarak tespit edimiştir. Buna karşın ortlama mutlak sapma performans kriteri açısından değerlendirildiğinde rastgele orman algoritması yönteminin diğer yöntemlere göre daha başarılı olduğu ortaya çıkmıştır.

Tablo 4.42: Senaryo 3-Tüm veri setlerinin ortalama performans sonuçları.

Yöntemler	Rastgele Orman Algoritması		Destek Vektör Makineleri		Yapay Sinir Ağları	
Veri_Setleri	Mutlak_Sapma	Mutlak Yüzde Hata	Mutlak_Sapma	Mutlak Yüzde Hata	Mutlak_Sapma	Mutlak Yüzde Hata

Veri_1	24,33	0,21	30,77	0,26	30,26	0,24
Veri_2	23,06	0,17	27,16	0,21	25,96	0,17
Veri_3	19,97	0,18	29,55	0,26	26,41	0,21
Veri_4	23,06	0,20	33,02	0,29	30,63	0,23
Veri_5	20,15	0,17	26,49	0,23	24,60	0,18
Veri_6	18,33	0,14	22,68	0,17	20,70	0,14

Senaryo 3 çözümleri, Senaryo 1 ve Senaryo 2 ile karşılaştırıldığında hem ortalama mutlak sapma hem de ortalama mutlak yüzde performans kriterlerinde daha iyi sonuçlar elde edilmiştir. Senaryo 3 kabulleri altında kullanılan yöntemler analiz edildiğinde mutlak sapma performans kriteri için en başarılı sonuçlar veren algoritmanın rastgele orman algoritması olduğu sonucuna varılmıştır. Mutlak yüzde hata performans kriteri için en başarılı yöntemin de yine rastgele orman algoritması yöntemi olduğu sonucu ortaya çıkmıştır.

4.2.4. Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözümler

Çalışmada iki yöntem uygulamalı hibrit çözümde; hibrit çözüm yöntemleri kullanılarak yeni veri setlerine uygulanmıştır. Burada başlıca 9 adet model oluşturulmuş ve her bir veri seti için çözümler elde edilmiştir. Oluşturulan modeller;

YSA_DVM: Öncelik sırasıyla yapay sinir ağları ve destek vektör makinalarını içeren hibrit çözüm modeli

YSA_ROA: Öncelik sırasıyla yapay sinir ağları ve rastgele orman algoritmalarını içeren hibrit çözüm modeli

YSA_YSA: Çift kademeli yapay sinir ağlarını içeren çözüm modeli

ROA_YSA: Öncelik sırasıyla rastgele orman algoritmaları ve yapay sinir ağlarını içeren hibrit çözüm modeli

ROA_DVM: Öncelik sırasıyla rastgele orman algori tmaları ve destek vektör makinalarını içeren hibrit çözüm modeli

ROA_ROA: Çift kademeli rastgele orman algoritmaları içeren çözüm modeli

DVM_ROA: Öncelik sırasıyla destek vektör makinaları ve rastgele orman algoritmalarını içeren hibrit çözüm modeli

DVM_YSA: Öncelik sırasıyla destek vektör makinaları ve yapay sinir ağlarını içeren hibrit çözüm modeli.

DVM_DVM: Çift kademeli destek vektör makinaları içeren çözüm modeli.

Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözüm'de kullanılan değişkenler ile Senaryo 3'de kullanılan değişkenler aynı olmakla beraber burada yine senaryo 3'de yer alan düzenlenmiş veriler kullanılmıştır. 6 adet hasta için yapılan çalışmada oluşturulan hibrit modeller ile ortalama mutlak sapma ve ortalama mutlak yüzde hata performanslarında önemli iyileşmeler sağlanmıştır.

Bu modelde uygulanan ilk yöntem sonucunda elde edilen sonuç değişkeni, diğer modelde girdi değişkeni olarak kullanılmıştır. Bu sebepten ötürü hibrit modeller, yöntemlere göre parametrelerde değişiklikler meydana gelmiştir.

Uygulamada yer alan modeller için, her hibrit modelde yeniden değişen farklı parametrelerin tanımlanmasının gerektiği düşünülmüştür.

YSA_DVM Hibrit Modeli Parametreleri:
Bu hibrit modelin çözümünde ilk olarak yapay sinir ağları kullanılmış ve sonrasında tahmin için destek vektör makinaları kullanılmıştır.
YSA Parametreleri
Girdi Değişkenleri: 3 Girdi
Çıktı Değişkenleri: 1 Çıktı
Ağ Modeli: Çok Katmanlı İleri Beslemeli Sinir Ağı (newff)
Momentum katsayısı: 0,5
Öğrenme hızı: 0,01
Gizli katman nöron sayısı: 2
Gizli katman aktivasyon fonksiyonu: Tanjant sigmoid transfer fonksiyonu (tansig)
Çıktı katmanı transfer fonksiyonu: Lineer transfer fonksiyonu (purelin)
Eğitim Optimizasyon Yöntemi: İniş Tabanlı Optimizayon Algoritması - Momenum Tabanlı Gradyan İniş Eğitim Yöntemi ("ADAPTgdwm": Adaptative gradient descend with momentum.)
İterasyon sayısı: 100
Yapay Sinir Ağları Eğitim Kriteri: En Küçük Ortalamalar Kareler Hatası (LMS)
Yapay sinir ağları sonucunda elde edilen tahmin çıktı değişkeni, destek vektör makinaları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır. Kullanılan destek vektör makinaları yöntemi ile nihai tahmin sonucu elde edilmiştir.

DVM Parametreleri

Girdi Değişkenleri: 4 Girdi

Çıktı Değişkenleri: 1 Çıktı

Çekirdek (Kernel) Fonksiyonu: Radyal Tabanlı Çekirdek Fonksiyonu (radial basis function)

Radyal tabanlı fonksiyon derece değeri: 3

Gama parametresi: 0,25

Coef0 parametresi :0

Sonlandırma tolerans kriteri: 0,001

Kayıp fonksiyon: Epsilon değeri: 0.1

İşlemci ön bellek: 40 Mb

YSA_ROA Hibrit Modeli Parametreleri:

Bu hibrit modelin çözümünde ilk olarak yapay sinir ağları kullanılmış ve sonrasında tahmin için rastgele orman algoritmaları kullanılmıştır.

YSA Parametreleri

Girdi Değişkenleri: 3 Girdi

Çıktı Değişkenleri: 1 Çıktı

Ağ Modeli: Çok Katmanlı İleri Beslemeli Sinir Ağı (newff)

Momentum katsayısı: 0,5

Öğrenme hızı: 0,01

Gizli katman nöron sayısı: 2

Gizli katman aktivasyon fonksiyonu: Tanjant sigmoid transfer fonksiyonu (tansig)

Çıktı katmanı transfer fonksiyonu: Lineer transfer fonksiyonu (purelin)

Eğitim Optimizasyon Yöntemi: İniş Tabanlı Optimizayon Algoritması - Momenum Tabanlı Gradyan İniş Eğitim Yöntemi ("ADAPTgdwm": Adaptative gradient descend with momentum.)

İterasyon sayısı: 100

Yapay sinir ağları sonucunda elde edilen tahmin çıktı değişkeni, rastgele orman algoritmaları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır. Kullanılan rastgele orman algoritmaları yöntemi ile nihai tahmin sonucu elde edilmiştir.

ROA Parametreleri

Girdi Değişkenleri: 4 Girdi

Çıktı Değişkenleri: 1 Çıktı

Geliştirilecek ağaç sayısı: 100

Değişken sayısı: 4

Her bir düğümde kullanılmış m değişkenlerin sayısı: 2

Ölçüt kriteri: Gini katsayısı

Eğitici veri seti için ağaç sayısı: 67

Test veri seti için ağaç sayısı: 33

Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız

Minimum örnek düğüm sayısı: 2

Minimum yaprak sayısı: 1

Maximum özellik sayısı: Rastgele (random) özellik seçimi

YSA_YSA Hibrit Modeli Parametreleri:

Bu hibrit modelin çözümünde ilk olarak yapay sinir ağları kullanılmış ve sonrasında tahmin için tekrar yapay sinir ağları kullanılmıştır.

YSA Parametreleri

Girdi Değişkenleri: 3 Girdi

Çıktı Değişkenleri: 1 Çıktı

Ağ Modeli: Çok Katmanlı İleri Beslemeli Sinir Ağı (newff)

Momentum katsayısı: 0,5

Öğrenme hızı: 0,01

Gizli katman nöron sayısı: 2

Gizli katman aktivasyon fonksiyonu: Tanjant sigmoid transfer fonksiyonu (tansig)

Çıktı katmanı transfer fonksiyonu: Lineer transfer fonksiyonu (purelin)

Eğitim Optimizasyon Yöntemi: İniş Tabanlı Optimizayon Algoritması - Momenum Tabanlı Gradyan İniş Eğitim Yöntemi ("ADAPTgdwm": Adaptative gradient descend with momentum.)

İterasyon sayısı: 100

İlk yapay sinir ağları sonucunda elde edilen tahmin çıktı değişkeni, ikinci yapay sinir ağları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır. İkinci kullanılan yapay sinir ağları yöntemi ile nihai tahmin sonucu elde edilmiştir.

YSA Parametreleri

Girdi Değişkenleri: 4 Girdi

Çıktı Değişkenleri: 1 Çıktı
Ağ Modeli: Çok Katmanlı İleri Beslemeli Sinir Ağı (newff)
Momentum katsayısı: 0,5
Öğrenme hızı: 0,01
Gizli katman nöron sayısı: 2
Gizli katman aktivasyon fonksiyonu: Tanjant sigmoid transfer fonksiyonu (tansig)
Çıktı katmanı transfer fonksiyonu: Lineer transfer fonksiyonu (purelin)
Eğitim Optimizasyon Yöntemi: İniş Tabanlı Optimizayon Algoritması - Momenum Tabanlı Gradyan İniş Eğitim Yöntemi ("ADAPTgdwm": Adaptative gradient descend with momentum.)
İterasyon sayısı: 100

ROA_YSA Hibrit Modeli Parametreleri:
Bu hibrit modelin çözümünde ilk olarak rastgele orman algoritmaları kullanılmış ve sonrasında tahmin için yapay sinir ağları kullanılmıştır.
ROA Parametreleri
Girdi Değişkenleri: 3 Girdi
Çıktı Değişkenleri: 1 Çıktı
Geliştirilecek ağaç sayısı: 100
Değişken sayısı: 3
Her bir düğümde kullanılmış m değişkenlerin sayısı: 2
Ölçüt kriteri: Gini katsayısı
Eğitici veri seti için ağaç sayısı: 67
Test veri seti için ağaç sayısı: 33
Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız
Minimum örnek düğüm sayısı: 2
Minimum yaprak sayısı: 1
Maximum özellik sayısı: Rastgele (random) özellik seçimi
Rastgele orman algoritmaları sonucunda elde edilen tahmin çıktı değişkeni, yapay sinir ağları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır. Kullanılan yapay sinir ağları yöntemi ile nihai tahmin sonucu elde edilmiştir.
YSA Parametreleri

Girdi Değişkenleri: 4 Girdi
Çıktı Değişkenleri: 1 Çıktı
Ağ Modeli: Çok Katmanlı İleri Beslemeli Sinir Ağı (newff)
Momentum katsayısı: 0,5
Öğrenme hızı: 0,01
Gizli katman nöron sayısı: 2
Gizli katman aktivasyon fonksiyonu: Tanjant sigmoid transfer fonksiyonu (tansig)
Çıktı katmanı transfer fonksiyonu: Lineer transfer fonksiyonu (purelin)
Eğitim Optimizasyon Yöntemi: İniş Tabanlı Optimizayon Algoritması - Momenum Tabanlı Gradyan İniş Eğitim Yöntemi ("ADAPTgdwm": Adaptative gradient descend with momentum.)
İterasyon sayısı: 100

ROA_DVM Hibrit Modeli Parametreleri:
Bu hibrit modelin çözümünde ilk olarak rastgele orman algoritmaları kullanılmış ve sonrasında tahmin için destek vektör makinaları kullanılmıştır.
ROA Parametreleri
Girdi Değişkenleri: 3 Girdi
Çıktı Değişkenleri: 1 Çıktı
Geliştirilecek ağaç sayısı: 100
Değişken sayısı: 3
Her bir düğümde kullanılmış m değişkenlerin sayısı: 2
Ölçüt kriteri: Gini katsayısı
Eğitici veri seti için ağaç sayısı: 67
Test veri seti için ağaç sayısı: 33
Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız
Minimum örnek düğüm sayısı: 2
Minimum yaprak sayısı: 1
Maximum özellik sayısı: Rastgele (random) özellik seçimi
Rastgele orman algoritmaları sonucunda elde edilen tahmin çıktı değişkeni, destek vektör makinaları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır. Kullanılan destek vektör makinaları yöntemi ile nihai tahmin sonucu elde edilmiştir.

DVM Parametreleri

Girdi Değişkenleri: 4 Girdi

Çıktı Değişkenleri: 1 Çıktı

Çekirdek (Kernel) Fonksiyonu: Radyal Tabanlı Çekirdek Fonksiyonu (radial basis function)

Radyal tabanlı fonksiyon derece değeri: 3

Gama parametresi: 0,25

Coef0 parametresi :0

Sonlandırma tolerans kriteri: 0,001

Kayıp fonksiyon: Epsilon değeri: 0.1

İşlemci ön bellek: 40 Mb

ROA_ROA Hibrit Modeli Parametreleri:

Bu hibrit modelin çözümünde ilk olarak rastgele orman algoritmaları kullanılmış ve sonrasında tahmin için tekrar rastgele orman algoritmaları kullanılmıştır.

ROA Parametreleri

Girdi Değişkenleri: 3 Girdi

Çıktı Değişkenleri: 1 Çıktı

Geliştirilecek ağaç sayısı: 100

Değişken sayısı: 3

Her bir düğümde kullanılmış m değişkenlerin sayısı: 2

Ölçüt kriteri: Gini katsayısı

Eğitici veri seti için ağaç sayısı: 67

Test veri seti için ağaç sayısı: 33

Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız

Minimum örnek düğüm sayısı: 2

Minimum yaprak sayısı: 1

Maximum özellik sayısı: Rastgele (random) özellik seçimi

İlk rastgele orman algoritmaları sonucunda elde edilen tahmin çıktı değişkeni, ikinci rastgele orman algoritmaları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır. İkinci kullanılan rastgele orman algoritmaları yöntemi ile nihai tahmin sonucu elde edilmiştir.

ROA Parametreleri

Girdi Değişkenleri: 4 Girdi

Çıktı Değişkenleri: 1 Çıktı

Geliştirilecek ağaç sayısı: 100

Değişken sayısı: 4

Her bir düğümde kullanılmış m değişkenlerin sayısı: 2

Ölçüt kriteri: Gini katsayısı

Eğitici veri seti için ağaç sayısı: 67

Test veri seti için ağaç sayısı: 33

Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız

Minimum örnek düğüm sayısı: 2

Minimum yaprak sayısı: 1

Maximum özellik sayısı: Rastgele (random) özellik seçimi

DVM_ROA Hibrit Modeli Parametreleri:

Bu hibrit modelin çözümünde ilk olarak destek vektör makinaları kullanılmış ve sonrasında tahmin için rastgele orman algoritmaları kullanılmıştır.

DVM Parametreleri

Girdi Değişkenleri: 3 Girdi

Çıktı Değişkenleri: 1 Çıktı

Çekirdek (Kernel) Fonksiyonu: Radyal Tabanlı Çekirdek Fonksiyonu (radial basis function)

Radyal tabanlı fonksiyon derece değeri: 3

Gama parametresi: 0,33

Coef0 parametresi :0

Sonlandırma tolerans kriteri: 0,001

Kayıp fonksiyon: Epsilon değeri: 0.1

İşlemci ön bellek: 40 Mb

Destek vektör makinaları sonucunda elde edilen tahmin çıktı değişkeni, rastgele orman algoritmaları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır. Kullanılan rastgele orman algoritmaları yöntemi ile nihai tahmin sonucu elde edilmiştir.

ROA Parametreleri

Girdi Değişkenleri: 4 Girdi

Çıktı Değişkenleri: 1 Çıktı
Geliştirilecek ağaç sayısı: 100
Değişken sayısı: 4
Her bir düğümde kullanılmış m değişkenlerin sayısı: 2
Ölçüt kriteri: Gini katsayısı
Eğitici veri seti için ağaç sayısı: 67
Test veri seti için ağaç sayısı: 33
Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız
Minimum örnek düğüm sayısı: 2
Minimum yaprak sayısı: 1
Maximum özellik sayısı: Rastgele (random) özellik seçimi

DVM_YSA Hibrit Modeli Parametreleri:
Bu hibrit modelin çözümünde ilk olarak destek vektör makinaları kullanılmış ve sonrasında tahmin için yapay sinir ağları kullanılmıştır.
DVM Parametreleri
Girdi Değişkenleri: 3 Girdi
Çıktı Değişkenleri: 1 Çıktı
Çekirdek (Kernel) Fonksiyonu: Radyal Tabanlı Çekirdek Fonksiyonu (radial basis function)
Radyal tabanlı fonksiyon derece değeri: 3
Gama parametresi: 0,33
Coef0 parametresi :0
Sonlandırma tolerans kriteri: 0,001
Kayıp fonksiyon: Epsilon değeri: 0.1
İşlemci ön bellek: 40 Mb
Destek vektör makinaları sonucunda elde edilen tahmin çıktı değişkeni, yapay sinir ağları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır. Kullanılan yapay sinir ağları yöntemi ile nihai tahmin sonucu elde edilmiştir.
YSA Parametreleri
Girdi Değişkenleri: 4 Girdi
Çıktı Değişkenleri:1 Çıktı
Ağ Modeli: Çok Katmanlı İleri Beslemeli Sinir Ağı (newff)

Momentum katsayısı: 0,5

Öğrenme hızı: 0,01

Gizli katman nöron sayısı: 2

Gizli katman aktivasyon fonksiyonu: Tanjant sigmoid transfer fonksiyonu (tansig)

Çıktı katmanı transfer fonksiyonu: Lineer transfer fonksiyonu (purelin)

Eğitim Optimizasyon Yöntemi: İniş Tabanlı Optimizayon Algoritması - Momenum Tabanlı Gradyan İniş Eğitim Yöntemi ("ADAPTgdwm": Adaptative gradient descend with momentum.)

İterasyon sayısı: 100

Yapay Sinir Ağları Eğitim Kriteri: En Küçük Ortalamalar Kareler Hatası (LMS)

DVM_DVM Hibrit Modeli Parametreleri:

Bu hibrit modelin çözümünde ilk olarak destek vektör makinaları kullanılmış ve sonrasında tahmin için tekrar destek vektör makinaları kullanılmıştır.

DVM Parametreleri

Girdi Değişkenleri: 3 Girdi

Çıktı Değişkenleri: 1 Çıktı

Çekirdek (Kernel) Fonksiyonu: Radyal Tabanlı Çekirdek Fonksiyonu (radial basis function)

Radyal tabanlı fonksiyon derece değeri: 3

Gama parametresi: 0,33

Coef0 parametresi :0

Sonlandırma tolerans kriteri: 0,001

Kayıp fonksiyon: Epsilon değeri: 0.1

İşlemci ön bellek: 40 Mb

İlk destek vektör makinaları sonucunda elde edilen tahmin çıktı değişkeni, ikinci destek vektör makinaları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır. İkinci kullanılan destek vektör makinaları yöntemi ile nihai tahmin sonucu elde edilmiştir.

DVM Parametreleri

Girdi Değişkenleri: 4 Girdi

Çıktı Değişkenleri: 1 Çıktı

Çekirdek (Kernel) Fonksiyonu: Radyal Tabanlı Çekirdek Fonksiyonu (radial basis function)

Radyal tabanlı fonksiyon derece değeri: 3

Gama parametresi: 0,25

Coef0 parametresi :0

Sonlandırma tolerans kriteri: 0,001

Kayıp fonksiyon: Epsilon değeri: 0.1

İşlemci ön bellek: 40 Mb

Tablo 4.43: Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözüm'de mutlak sapma değerleri.

Yöntem / Veri Setleri	Veri Seti 1	Veri Seti 2	Veri Seti 3	Veri Seti 4	Veri Seti 5	Veri Seti 6	Mutlak Sapma
ROA_DVM	18,42	18,59	15,27	16,98	15,14	11,06	15,91
ROA_ROA	15,93	15,33	12,30	13,75	12,77	13,81	13,98
DVM_DVM	30,75	27,17	29,53	33,01	26,51	8,00	25,83
DVM_ROA	21,97	22,27	18,47	20,73	18,52	16,61	19,76
YSA_ROA	23,69	22,54	18,78	20,14	17,44	15,88	19,75
YSA_DVM	29,83	25,78	24,87	28,72	23,88	20,11	25,53
ROA_YSA	18,11	18,76	15,87	17,29	16,07	10,08	16,03
DVM_YSA	29,21	27,11	27,08	28,78	25,29	20,65	26,35
YSA_YSA	30,43	26,06	24,3	27,70	22,72	18,93	25,02

Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözümlerde kullanılan yöntemler analiz edildiğinde mutlak sapma performans kriteri için en başarılı sonuçlar veren hibrit algoritmanın 13,98'lik mutlak sapma değeriyle ROA_ROA hibrit algoritması olduğu sonucuna varılmıştır.

Tablo 4.44: Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözüm'de mutlak yüzde hata ortalama değerleri.

Yöntem / Veri Setleri	Veri Seti 1	Veri Seti 2	Veri Seti 3	Veri Seti 4	Veri Seti 5	Veri Seti 6	Yöntem'e Göre Mutlak Yüzde Hata Ort.
ROA_DVM	0,16	0,14	0,14	0,15	0,13	0,08	0,13
ROA_ROA	0,14	0,11	0,11	0,12	0,11	0,10	0,12
DVM_DVM	0,26	0,21	0,26	0,29	0,23	0,05	0,22

DVM_ROA	0,19	0,17	0,16	0,18	0,16	0,13	0,17
YSA_ROA	0,20	0,17	0,17	0,17	0,15	0,12	0,16
YSA_DVM	0,26	0,20	0,22	0,25	0,20	0,16	0,22
ROA_YSA	0,15	0,13	0,13	0,14	0,12	0,06	0,12
DVM_YSA	0,23	0,18	0,23	0,22	0,18	0,14	0,20
YSA_YSA	0,22	0,17	0,20	0,21	0,17	0,13	0,18

Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözüm'leri için performans sonuçları diğer senaryolarla karşılaştırıldığında; Senaryo 1, Senaryo 2 ve Senaryo 3 model çözümlerine göre bu modelin daha iyi sonuçlar verdiği ortaya çıkmıştır. Oluşturulan tüm hibrit modeller incelendiğinde tahmin başarı oran ortalamalarının % 80 ile % 89 arasında değiştiği saptanmıştır.

Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözümler analiz edildiğinde mutlak yüzde hata performans kriteri için en başarılı sonuçlar veren hibrit algoritmaların 0,12'lik değerleriyle ROA_ROA ve ROA_YSA hibrit algoritmaları olduğu sonucuna varılmıştır.

4.2.5. Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözümler

Çalışmada Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözüm'de; Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözüm'de oluşturulan hibrit çözüm yöntemleri geliştirilerek yeni hibrit modeller oluşturularak veri setlerine uygulanmıştır. Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözüm'de başlıca 27 adet hibrit model oluşturulmuş ve her bir veri seti için çözümler elde edilmiştir. Oluşturulan hibrit modeller; YSA_DVM_ROA, YSA_DVM_ROA, YSA_DVM_DVM v.b. şeklinde oluşturulmuştur. Burada YSA_DVM_ROA hibrit modeli, veri setine bu 3 yöntemin birlikte uygulanarak hibrit modelin oluşturulduğu ve önce yapay sinir ağları kullanıldığını, devamında destek vektör makinaları kullanıldığını ve son olarak rastgele orman algoritmaları kullanılarak sonuçların elde edildiği anlamına gelmiştir. Modeller;

YSA_DVM_ROA: Öncelik sırasıyla yapay sinir ağları, destek vektör makinaları ve rastgele orman algoritmalarını içeren hibrit çözüm modeli.

YSA_DVM_YSA: Öncelik sırasıyla yapay sinir ağları, destek vektör makinaları ve tekrar yapay sinir ağları içeren hibrit çözüm modeli.

YSA_DVM_DVM: Öncelik sırasıyla yapay sinir ağları, destek vektör makinaları ve tekrar destek vektör makinaları içeren hibrit çözüm modeli.

YSA_YSA_YSA: Üç kademeli yapay sinir ağları içeren çözüm modeli.

YSA_YSA_DVM: Öncelik sırasıyla yapay sinir ağları, tekrar yapay sinir ağları ve destek vektör makinaları içeren hibrit çözüm modeli.

YSA_YSA_ROA: Öncelik sırasıyla yapay sinir ağları, tekrar yapay sinir ağları ve rastgele orman algoritmalarını içeren hibrit çözüm modeli.

YSA_ROA_YSA: Öncelik sırasıyla yapay sinir ağları, rastgele orman algoritmalarını ve tekrar yapay sinir ağlarını içeren hibrit çözüm modeli.

YSA_ROA_DVM: Öncelik sırasıyla yapay sinir ağları, rastgele orman algoritmalarını ve destek vektör makinaları içeren hibrit çözüm modeli.

YSA_ROA_ROA: Öncelik sırasıyla yapay sinir ağları, destek vektör makinaları ve tekrar destek vektör makinaları içeren hibrit çözüm modeli.

DVM_DVM_ROA: Öncelik sırasıyla destek vektör makinaları, tekrar destek vektör makinaları ve rastgele orman algoritmalarını içeren hibrit çözüm modeli.

DVM_DVM_YSA: Öncelik sırasıyla destek vektör makinaları, tekrar destek vektör makinaları ve yapay sinir ağları içeren hibrit çözüm modeli.

DVM_DVM_DVM: Üç kademeli destek vektör makinaları içeren çözüm modeli.

DVM_YSA_YSA: Öncelik sırasıyla destek vektör makinaları, yapay sinir ağları ve tekrar yapay sinir ağları içeren hibrit çözüm modeli.

DVM_YSA_DVM: Öncelik sırasıyla destek vektör makinaları, yapay sinir ağları ve tekrar destek vektör makinaları içeren hibrit çözüm modeli.

DVM_YSA_ROA: Öncelik sırasıyla destek vektör makinaları, yapay sinir ağları ve rastgele orman algoritmalarını içeren hibrit çözüm modeli.

DVM_ROA_YSA: Öncelik sırasıyla destek vektör makinaları, rastgele orman algoritmaları ve yapay sinir ağlarını içeren hibrit çözüm modeli.

DVM_ROA_DVM: Öncelik sırasıyla destek vektör makinaları, rastgele orman algoritmaları ve tekrar destek vektör makinaları içeren hibrit çözüm modeli.

DVM_ROA_ROA: Öncelik sırasıyla destek vektör makinaları, rastgele orman algoritmaları ve tekrar rastgele orman algoritmaları içeren hibrit çözüm modeli.

ROA_DVM_ROA: Öncelik sırasıyla rastgele orman algoritmaları, destek vektör makinaları ve tekrar rastgele orman algoritmalarını içeren hibrit çözüm modeli.

ROA_DVM_DVM: Öncelik sırasıyla rastgele orman algoritmaları, destek vektör makinaları ve tekrar destek vektör makinaları içeren hibrit çözüm modeli.
ROA_ROA_ROA: Üç kademeli rastgele orman algoritmalarını içeren çözüm modeli.
ROA_YSA_YSA: Öncelik sırasıyla rastgele orman algoritmaları, yapay sinir ağları ve tekrar yapay sinir ağları içeren hibrit çözüm modeli.
ROA_YSA_DVM: Öncelik sırasıyla rastgele orman algoritmaları, yapay sinir ağları ve destek vektör makinaları içeren hibrit çözüm modeli.
ROA_YSA_ROA: Öncelik sırasıyla rastgele orman algoritmaları, yapay sinir ağları ve tekrar rastgele orman algoritmalarını içeren hibrit çözüm modeli.
ROA_ROA_YSA: Öncelik sırasıyla rastgele orman algoritmaları, tekrar rastgele orman algoritmaları ve yapay sinir ağlarını içeren hibrit çözüm modeli.
ROA_ROA_DVM: Öncelik sırasıyla rastgele orman algoritmaları, tekrar rastgele orman algoritmaları ve destek vektör makinalarını içeren hibrit çözüm modeli.
ROA_DVM_YSA: Öncelik sırasıyla rastgele orman algoritmaları, destek vektör makinaları ve yapay sinir ağları içeren hibrit çözüm modeli.

Yukarıda bahsedilen 27 adet hibrit modelde de parametre değerleri modele özgü olarak değişmiş ve birbirinden farklılık göstermiştir. Örnek teşkil etmesi açısından ROA_DVM_YSA hibrit modeli parametreleri şu şekilde özetlenmiştir;

ROA_DVM_YSA Hibrit Modeli Parametreleri:
Bu hibrit modelin çözümünde ilk olarak rastgele orman algoritmaları kullanılmış, sonrasında tahmin için destek vektör makinaları ve son olarak yapay sinir ağları kullanılmıştır.
ROA Parametreleri
Girdi Değişkenleri: 3 Girdi
Çıktı Değişkenleri: 1 Çıktı
Geliştirilecek ağaç sayısı: 100
Değişken sayısı: 3
Her bir düğümde kullanılmış m değişkenlerin sayısı: 2
Ölçüt kriteri: Gini katsayısı
Eğitici veri seti için ağaç sayısı: 67
Test veri seti için ağaç sayısı: 33

Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız

Minimum örnek düğüm sayısı: 2

Minimum yaprak sayısı: 1

Maximum özellik sayısı: Rastgele (random) özellik seçimi

Rastgele orman algoritmaları sonucunda elde edilen tahmin çıktı değişkeni, destek vektör makinaları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır.

DVM Parametreleri

Girdi Değişkenleri: 4 Girdi

Çıktı Değişkenleri: 1 Çıktı

Çekirdek (Kernel) Fonksiyonu: Radyal Tabanlı Çekirdek Fonksiyonu (radial basis function)

Radyal tabanlı fonksiyon derece değeri: 3

Gama parametresi: 0,25

Coef0 parametresi :0

Sonlandırma tolerans kriteri: 0,001

Kayıp fonksiyon: Epsilon değeri: 0.1

İşlemci ön bellek: 40 Mb

Hibrit modelin oluşumu için son çözüm algoritmasında rastgele orman algoritmaları sonucunda elde edilen tahmin çıktı değişkeni, destek vektör makinaları sonucunda elde edilen tahmin çıktı değişkeni ve yapay sinir ağları için her algoritmada kullanılan 3 adet temel girdi değişkenleri olmak üzere 5 adet girdi değişkeni kullanılmıştır Kullanılan yapay sinir ağları yöntemi ile nihai tahmin sonucu elde edilmiştir.

YSA Parametreleri

Girdi Değişkenleri: 5 Girdi

Çıktı Değişkenleri: 1 Çıktı

Ağ Modeli: Çok Katmanlı İleri Beslemeli Sinir Ağı (newff)

Momentum katsayısı: 0,5

Öğrenme hızı: 0,01

Gizli katman nöron sayısı: 2

Gizli katman aktivasyon fonksiyonu: Tanjant sigmoid transfer fonksiyonu (tansig)

Çıktı katmanı transfer fonksiyonu: Lineer transfer fonksiyonu (purelin)

Eğitim Optimizasyon Yöntemi: İniş Tabanlı Optimizayon Algoritması - Momenum Tabanlı Gradyan İniş Eğitim Yöntemi ("ADAPTgdwm": Adaptative gradient descend with momentum.)

İterasyon sayısı: 100

Tablo 4.45: Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözüm'de yer alan veri seti çalışma yapısı kesit görünümü.

Yöntemler	Veri Seti 1		Veri Seti 2		Veri Seti 3		Veri Seti 4	
	OMH	OMYH	OMH	OMYH	OMH	OMYH	OMH	OMYH
ROA_DVM_ROA	13,27	0,114	12,97	0,096	10,56	0,092	10,83	0,092
DVM_DVM_ROA	20,94	0,180	22,05	0,165	17,72	0,156	19,75	0,171
DVM_ROA_ROA	12,77	0,110	13,57	0,100	10,25	0,09	10,31	0,09
YSA_ROA_ROA	16,26	0,140	15,83	0,118	12,34	0,108	12,65	0,107
YSA_DVM_ROA	22,36	0,192	22,28	0,167	17,34	0,154	19,23	0,165
ROA_YSA_ROA	13,67	0,118	13,38	0,099	10,37	0,09	11,22	0,09
DVM_YSA_ROA	21,65	0,186	21,85	0,163	17,84	0,159	20,16	0,176
YSA_YSA_ROA	22,47	0,193	21,82	0,163	18,30	0,163	19,20	0,165
DVM _ DVM _DVM	30,72	0,259	21,17	0,210	29,53	0,258	32,96	0,286
DVM_ROA_DVM	16,54	0,142	19,30	0,143	14,86	0,133	15,48	0,135
YSA_ROA_DVM	18,22	0,155	18,91	0,140	15,05	0,133	14,29	0,117
YSA_DVM_DVM	29,83	0,255	25,79	0,197	24,86	0,215	28,71	0,253
ROA_YSA_DVM	17,96	0,152	18,55	0,140	15,59	0,137	16,64	0,140
DVM_YSA_DVM	28,38	0,245	26,60	0,206	25,04	0,218	28,44	0,247
YSA_YSA_DVM	29,31	0,249	25,63	0,227	23,12	0,202	26,54	0,233

Çözüm sonucu elde edilebilen hibrit modeller diğer senaryolar ile karşılaştırıldıklarında daha iyi performans sonuçları vermiştir. Bu aşamada çalışma 3*3*3=27 farklı hibrit modele uygulanmıştır. Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözüm'de 3*3=9 birbirinden farklı en iyi hibrit modellerin performansları analiz edilmiştir. Bu sayede Senaryo 3 İki Yöntem

Uygulamalı Hibrit Çözüm'de olduğu gibi 9 hibrit model oluşturulması amaçlanmıştır. Oluşturulan modeller ile en iyi performans gösterebilecek hibrit modeller bulunup uygulanarak, 27 hibrit model yerine 9 hibrit model kullanarak, çalışmaya hız, esneklik ve kolaylık sağlanmıştır.

Tablo 4.46: Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözüm'e göre hibrit yöntemlerin mutlak sapma değerleri.

Yöntem / Veri Setleri	Veri Seti 1	Veri Seti 2	Veri Seti 3	Veri Seti 4	Veri Seti 5	Veri Seti 6	Yöntemlerin Mutlak Sapma Değerleri
ROA_DVM_ROA	13,27	12,97	10,56	10,83	10,54	11,33	11,58
DVM_DVM_ROA	20,94	22,04	17,71	19,75	17,81	14,10	18,73
DVM_ROA_ROA	12,77	13,57	10,25	10,31	10,66	12,32	11,65
YSA_ROA_ROA	16,26	15,83	12,34	12,65	11,92	12,57	13,60
YSA_DVM_ROA	22,36	22,28	17,34	19,23	16,74	14,75	18,78
ROA_YSA_ROA	13,67	13,38	10,37	11,22	10,52	10,99	11,69
DVM_ROA_DVM	16,54	19,30	14,86	15,48	14,35	10,80	15,22
YSA_ROA_DVM	18,22	18,91	15,05	14,29	12,12	11,60	15,03
ROA_YSA_DVM	17,96	18,55	15,59	16,64	14,57	10,57	15,65

Elde edilen sonuçlara göre en düşük mutlak sapma değerinin 11,58 ile ROA_DVM_ROA hibrit yöntemine ait olduğu görülmüştür.

Tablo 4.47: Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözüm'e göre hibrit yöntemlerin mutlak yüzde hata ortalama değerleri.

Yöntem / Veri Setleri	Veri Seti 1	Veri Seti 2	Veri Seti 3	Veri Seti 4	Veri Seti 5	Veri Seti 6	Yöntemlerin Mutlak Yüzde Hata Ort.
ROA_DVM_ROA	0,11	0,10	0,09	0,09	0,09	0,08	0,09
DVM_DVM_ROA	0,18	0,17	0,16	0,17	0,15	0,11	0,16

Tablo 4.47 (devam)

Yöntem / Veri Setleri	Veri Seti 1	Veri Seti 2	Veri Seti 3	Veri Seti 4	Veri Seti 5	Veri Seti 6	Yöntemlerin Mutlak Yüzde Hata Ort.
DVM_ROA_ROA	0,11	0,10	0,09	0,09	0,09	0,09	0,10

YSA_ROA_ROA	0,14	0,12	0,11	0,11	0,10	0,09	0,11
YSA_DVM_ROA	0,19	0,17	0,15	0,17	0,14	0,11	0,16
ROA_YSA_ROA	0,12	0,10	0,09	0,09	0,09	0,08	0,10
DVM_ROA_DVM	0,14	0,14	0,13	0,14	0,12	0,08	0,13
YSA_ROA_DVM	0,16	0,14	0,13	0,12	0,10	0,08	0,12
ROA_YSA_DVM	0,15	0,14	0,14	0,14	0,12	0,08	0,13

Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözüm performans sonuçları diğer senaryolarla karşılaştırıldığında Senaryo 1, Senaryo 2, Senaryo 3 ve Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözümlerine göre daha iyi sonuçlar verdiği ortaya çıkmıştır. Elde edilen sonuçlara göre en iyi tahmin başarı oranı performansının % 91 ile ROA_DVM_ROA hibrit yöntemine ait olduğu görülmüştür.

4.2.6. Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler

Çalışmanın bu kısmında Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler, çalışma sonuçlarında en yüksek yüzdeli doğrulukla tahmin etme başarısı gösteren 3 hibrit model kullanılarak oluşturulmuştur. Bu koşullar altında oluşturulan modeller; ROA_DVM_ROA, DVM_ROA_ROA ve ROA_YSA_ROA hibrit modelleri olarak belirlenmiştir. Bu modellerin açılımı;

ROA_DVM_ROA: Öncelik sırasıyla rastgele orman algoritmaları, destek vektör makinaları ve tekrar rastgele orman algoritmalarını içeren hibrit çözüm modeli.

ROA_DVM_ROA Hibrit Modeli Parametreleri:

Bu hibrit modelin çözümünde ilk olarak rastgele orman algoritmaları kullanılmış, sonrasında tahmin için destek vektör makinaları ve son olarak da rastgele orman algoritmaları kullanılmıştır.

ROA Parametreleri

Girdi Değişkenleri: 3 Girdi

Çıktı Değişkenleri: 1 Çıktı

Geliştirilecek ağaç sayısı: 100

Değişken sayısı: 3

Her bir düğümde kullanılmış m değişkenlerin sayısı: 2

Ölçüt kriteri: Gini katsayısı

Eğitici veri seti için ağaç sayısı: 67

Test veri seti için ağaç sayısı: 33

Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız

Minimum örnek düğüm sayısı: 2

Minimum yaprak sayısı: 1

Maximum özellik sayısı: Rastgele (random) özellik seçimi

Rastgele orman algoritmaları sonucunda elde edilen tahmin çıktı değişkeni, destek vektör makinaları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır.

DVM Parametreleri

Girdi Değişkenleri: 4 Girdi

Çıktı Değişkenleri: 1 Çıktı

Çekirdek (Kernel) Fonksiyonu: Radyal Tabanlı Çekirdek Fonksiyonu (radial basis function)

Radyal tabanlı fonksiyon derece değeri: 3

Gama parametresi: 0,25

Coef0 parametresi :0

Sonlandırma tolerans kriteri: 0,001

Kayıp fonksiyon: Epsilon değeri: 0.1

İşlemci ön bellek: 40 Mb

Hibrit modelin oluşumu için son çözüm algoritmasında rastgele orman algoritmaları sonucunda elde edilen tahmin çıktı değişkeni, destek vektör makinaları sonucunda elde edilen tahmin çıktı değişkeni ve sonraki rastgele orman algoritması için her algoritmada kullanılan 3 adet temel girdi değişkenleri olmak üzere 5 adet girdi değişkeni kullanılmıştır. Kullanılan rastgele orman algoritması yöntemi ile nihai tahmin sonucu elde edilmiştir.

ROA Parametreleri

Girdi Değişkenleri: 5 Girdi

Çıktı Değişkenleri: 1 Çıktı

Geliştirilecek ağaç sayısı: 100

Değişken sayısı: 5

Her bir düğümde kullanılmış m değişkenlerin sayısı: 3

Ölçüt kriteri: Gini katsayısı

Eğitici veri seti için ağaç sayısı: 67

Test veri seti için ağaç sayısı: 33

Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız

Minimum örnek düğüm sayısı: 2

Minimum yaprak sayısı: 1

Maximum özellik sayısı: Rastgele (random) özellik seçimi

Rastgele orman algoritmaları sonucunda elde edilen tahmin çıktı değişkeni, destek vektör makinaları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır.

DVM_ROA_ROA: Öncelik sırasıyla destek vektör makinaları, rastgele orman algoritmaları ve tekrar rastgele orman algoritmaları içeren hibrit çözüm modeli.

DVM_ROA_ROA Hibrit Modeli Parametreleri:

Bu hibrit modelin çözümünde ilk olarak destek vektör makinaları kullanılmış, sonrasında tahmin için rastgele orman algoritmaları ve son olarak da rastgele orman algoritmaları kullanılmıştır.

DVM Parametreleri

Girdi Değişkenleri: 3 Girdi

Çıktı Değişkenleri: 1 Çıktı

Çekirdek (Kernel) Fonksiyonu: Radyal Tabanlı Çekirdek Fonksiyonu (radial basis function)

Radyal tabanlı fonksiyon derece değeri: 3

Gama parametresi: 0,33

Coef0 parametresi :0

Sonlandırma tolerans kriteri: 0,001

Kayıp fonksiyon: Epsilon değeri: 0.1

İşlemci ön bellek: 40 Mb

Destek vektör makinaları sonucunda elde edilen tahmin çıktı değişkeni, rastgele orman algoritmaları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır.

ROA Parametreleri

Girdi Değişkenleri: 4 Girdi

Çıktı Değişkenleri: 1 Çıktı

Geliştirilecek ağaç sayısı: 100

Değişken sayısı: 4

Her bir düğümde kullanılmış m değişkenlerin sayısı: 2

Ölçüt kriteri: Gini katsayısı

Eğitici veri seti için ağaç sayısı: 67

Test veri seti için ağaç sayısı: 33

Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız

Minimum örnek düğüm sayısı: 2

Minimum yaprak sayısı: 1

Maximum özellik sayısı: Rastgele (random) özellik seçimi

Hibrit modelin oluşumu için son çözüm algoritmasında destek vektör makinaları sonucunda elde edilen tahmin çıktı değişkeni, rastgele orman algoritmaları sonucunda elde edilen tahmin çıktı değişkeni ve sonraki rastgele orman algoritması için her algoritmada kullanılan 3 adet temel girdi değişkenleri olmak üzere 5 adet girdi değişkeni kullanılmıştır. Kullanılan rastgele orman algoritması yöntemi ile nihai tahmin sonucu elde edilmiştir.

ROA Parametreleri

Girdi Değişkenleri: 5 Girdi

Çıktı Değişkenleri: 1 Çıktı

Geliştirilecek ağaç sayısı: 100

Değişken sayısı: 5

Her bir düğümde kullanılmış m değişkenlerin sayısı: 3

Ölçüt kriteri: Gini katsayısı

Eğitici veri seti için ağaç sayısı: 67

Test veri seti için ağaç sayısı: 33

Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız

Minimum örnek düğüm sayısı: 2

Minimum yaprak sayısı: 1

Maximum özellik sayısı: Rastgele (random) özellik seçimi

Rastgele orman algoritmaları sonucunda elde edilen tahmin çıktı değişkeni, destek vektör makinaları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır.

ROA_YSA_ROA: Öncelik sırasıyla rastgele orman algoritmaları, yapay sinir ağları ve tekrar rastgele orman algoritmalarını içeren hibrit çözüm modeli.

ROA_YSA_ROA Hibrit Modeli Parametreleri:

Bu hibrit modelin çözümünde ilk olarak rastgele orman algoritması kullanılmış, sonrasında tahmin için yapay sinir ağları ve son olarak da rastgele orman algoritmaları kullanılmıştır.

ROA Parametreleri

Girdi Değişkenleri: 3 Girdi

Çıktı Değişkenleri: 1 Çıktı

Geliştirilecek ağaç sayısı: 100

Değişken sayısı: 3

Her bir düğümde kullanılmış m değişkenlerin sayısı: 2

Ölçüt kriteri: Gini katsayısı

Eğitici veri seti için ağaç sayısı: 67

Test veri seti için ağaç sayısı: 33

Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız

Minimum örnek düğüm sayısı: 2

Minimum yaprak sayısı: 1

Maximum özellik sayısı: Rastgele (random) özellik seçimi

Rastgele orman algoritmaları sonucunda elde edilen tahmin çıktı değişkeni, yapay sinir ağları için diğer temel girdi değişkenleri ile beraber girdi değişkeni olarak kullanılmıştır.

YSA Parametreleri

Girdi Değişkenleri: 4 Girdi

Çıktı Değişkenleri: 1 Çıktı

Ağ Modeli: Çok Katmanlı İleri Beslemeli Sinir Ağı (newff)

Momentum katsayısı: 0,5

Öğrenme hızı: 0,01

Gizli katman nöron sayısı: 2

Gizli katman aktivasyon fonksiyonu: Tanjant sigmoid transfer fonksiyonu (tansig)

Çıktı katmanı transfer fonksiyonu: Lineer transfer fonksiyonu (purelin)

Eğitim Optimizasyon Yöntemi: İniş Tabanlı Optimizayon Algoritması - Momenum Tabanlı Gradyan İniş Eğitim Yöntemi ("ADAPTgdwm": Adaptative gradient descend with momentum.)

İterasyon sayısı: 100

Hibrit modelin oluşumu için son çözüm algoritmasında destek vektör makinaları sonucunda elde edilen tahmin çıktı değişkeni, yapay sinir ağları sonucunda elde edilen tahmin çıktı değişkeni ve sonraki rastgele orman algoritması için her algoritmada kullanılan 3 adet temel girdi değişkenleri olmak üzere 5 adet girdi değişkeni kullanılmıştır. Kullanılan rastgele orman algoritması yöntemi ile nihai tahmin sonucu elde edilmiştir.

ROA Parametreleri

Girdi Değişkenleri: 5 Girdi

Çıktı Değişkenleri: 1 Çıktı

Geliştirilecek ağaç sayısı: 100

Değişken sayısı: 5

Her bir düğümde kullanılmış m değişkenlerin sayısı: 3

Ölçüt kriteri: Gini katsayısı

Eğitici veri seti için ağaç sayısı: 67

Test veri seti için ağaç sayısı: 33

Yaprak düğümüne ulaşana kadar ağacın her düğüme ne kadar daha genişlemesi gerektiğinin ölçüsü: Sınırsız

Minimum örnek düğüm sayısı: 2

Minimum yaprak sayısı: 1

Maximum özellik sayısı: Rastgele (random) özellik seçimi

Tablo 4.48: Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e göre ortalama mutlak sapma değerleri.

Yöntem / Veri Setleri	Veri Seti 1	Veri Seti 2	Veri Seti 3	Veri Seti 4	Veri Seti 5	Veri Seti 6	Yöntem'e Göre Mutlak Sapma Değerleri
ROA_DVM_ROA	13,27	12,97	10,56	10,83	10,54	11,33	11,58
DVM_ROA_ROA	12,77	13,57	10,25	10,31	10,66	12,32	11,65
ROA_YSA_ROA	13,67	13,38	10,37	11,22	10,52	10,99	11,69

Hibrit çözümlerde elde edilen hataların mutlak sapma değerleri diğer yöntemlere göre daha başarılı sonuçlar vermiştir. Mutlak sapma değerleri 10,25 ile 13,67 arasında değişim

göstermiştir. Mutlak sapma değerlerine göre en iyi performansın 11,58 değeri ile ROA_DVM_ROA hibrit çözüm yönteminde olduğu tespit edilmiştir.

Tablo 4.49: Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e göre mutlak yüzde hata ortalama değerleri.

Yöntem / Veri Setleri	Veri Seti 1	Veri Seti 2	Veri Seti 3	Veri Seti 4	Veri Seti 5	Veri Seti 6	Yöntem'e Göre Mutlak Yüzde Hata Ort.
ROA_DVM_ROA	0,11	0,10	0,09	0,09	0,09	0,08	0,093
DVM_ROA_ROA	0,11	0,10	0,09	0,09	0,09	0,09	0,095
ROA_YSA_ROA	0,12	0,10	0,09	0,09	0,09	0,08	0,095

Mutlak yüzde hata performans kriterleri açısından Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler karşılaştırıldığında en başarılı algoritmanın 0,093 mutlak yüzde hata oranına sahip ROA_DVM_ROA hibrit algoritması olduğu sonucuna varılmıştır.

Tablo 4.50: Senaryolara göre yöntemlerin tahmin başarıları.

Senaryolar	Ortalama Mutlak Sapma	En iyi yöntem	Ortalama Mutlak Yüzde Hata	En iyi yöntem	Yöntemlerin Tahmin Başarı Oranları
Senaryo-1	ROA	43,73	YSA	0,24	% 76
Senaryo-2	ROA	38,40	ROA	0,19	% 81
Senaryo-3	ROA	21,71	ROA	0,18	% 82
Senaryo-3 Hibrit_İkili	ROA_ROA	13,98	ROA_ROA ve ROA_YSA	0,12	% 88
Senaryo-3 Hibrit_En_İyi_Üçlü	ROA_DVM_ROA	11,58	DVM_ROA_ROA ve ROA_YSA_ROA	0,09	% 91

Hibrit modellerin başarı performansları göz önüne alınarak, yeni veri setleri aracılığıyla hibrit yöntemlerin performansları analiz edilmiştir. Aşağıda yer alan yeni test verileri aracılığıyla yöntemin gücü ve doğruluğunu saptamak amacıyla çözümler yapılmıştır.

Test Veri Seti 7

Veri seti 7, 333.540 dakikalık gözlemler sonucundan oluşmuştur. Veri Seti 7'de başlıca 33 ve 34 kodlu olmak üzere 2 tip insulin tipi verisi ile 48, 58, 60 ve 62 kodlu olmak üzere 4 tip

çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır. Kan şekeri ölçüm değeri sonuçlarının 70-200 arasında olanları analizlerde kullanılmıştır.

Tablo 4.51: Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e göre analiz edilen Test Veri Seti 7 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri sayısı
33	Regular Tip İnsülin	495
34	NPH Tip İnsülin	148
48	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	124
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	94
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	134
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	140
Toplam veri sayısı		1135

Test Veri Seti 8

Veri Seti 8, 218.100 dakikalık gözlemler sonucundan oluşmuştur. Veri Seti 8'de başlıca 33 ve 34 kodlu olmak üzere 2 tip insulin tipi verisi ile 48, 58, 60 ve 62 kodlu olmak üzere 4 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır. Kan şekeri ölçüm değeri sonuçlarının 70-200 arasında olanları analizlerde kullanılmıştır.

Tablo 4.52: Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e göre analiz edilen Test Veri Seti 8 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri sayısı
33	Regular Tip İnsülin	441
34	NPH Tip İnsülin	147
48	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	127
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	101
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	122
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	130
Toplam veri sayısı		1068

Test Veri Seti 9

Veri Seti 9, 213.960 dakikalık gözlem sonuçlarından oluşmuştur. Veri Seti 9'da başlıca 33, 34 ve 35 kodlu olmak üzere 3 tip insulin tipi verisi ile 48, 57, 58, 60 ve 62 kodlu olmak üzere 5 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır. Kan şekeri ölçüm değeri sonuçlarının 70-200 arasında olanları analizlerde kullanılmıştır.

Tablo 4.53: Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e göre analiz edilen Test Veri Seti 9 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri sayısı
33	Regular Tip İnsülin	509
34	NPH Tip İnsülin	62
35	Ultralente Tip İnsülin	143
48	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	130
57	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	6
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	105
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	97
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	108
Toplam veri sayısı		1160

Test Veri Seti 10

Veri Seti 10, 196.815 dakikalık gözlem sonuçlarından oluşmuştur. Veri Seti 10'da başlıca 33, 34 ve 35 kodlu olmak üzere 3 tip insulin tipi verisi ile 48, 58, 60 ve 62 kodlu olmak üzere 4 tip çeşitli zamanlarda yapılan kandaki glikoz yoğunluğu (mg/dl) verisi kullanılarak modeller kurulmuş ve bu veriler düzenlenerek yöntemler uygulanmıştır. Kan şekeri ölçüm değeri sonuçlarının 70-200 arasında olanları analizlerde kullanılmıştır.

Tablo 4.54: Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e göre analiz edilen Test Veri Seti 10 özellikleri.

İlgili Kod Numarası	Kod Numarası Açıklaması	Veri sayısı
33	Regular Tip İnsülin	401
34	NPH Tip İnsülin	7
35	Ultralente Tip İnsülin	137
48	Belirtilmemiş zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	49
58	Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	41
60	Öğle yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	73
62	Akşam yemek öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl)	58
Toplam veri sayısı		766

Test verilerine hibrit çözüm yöntemleri uygulanırsa elde edilen performans sonuçları tablolarda görülmüştür. Yüksek doğruluk yüzdeli tahminler test verilerinin sonuçlarında da gözlemlenmiştir.

Tablo 4.55: Test verilerine uygulanan hibrit yöntemlerin mutlak sapma değerleri.

Yöntemler / Veri Setleri	Test Veri Seti 7	Test Veri Seti 8	Test Veri Seti 9	Test Veri Seti 10	Yöntem'e Göre Mutlak Sapma Değerleri
ROA_DVM_ROA	13,67	14,75	11,79	11,27	12,87
DVM_ROA_ROA	10,90	16,61	13,47	12,90	13,47
ROA_YSA_ROA	17,37	16,99	14,02	14,27	15,66

Hataların mutlak sapma değerleri sonuçlarına göre, en iyi performansın 12,87 mutlak sapma değerine sahip ROA_DVM_ROA hibrit yöntemine ait olduğu görülmüştür. Bu hibrit modellerin sonuçlarına göre ortalama mutlak sapma ve ortalama mutlak yüzde hata performans sonuçları oldukça başarılı olarak gerçekleşmiştir.

Tablo 4.56: Test verilerine uygulanan hibrit yöntemlerin mutlak yüzde hata ortalama değerleri.

Yöntem / Veri Setleri	Veri Seti 1	Veri Seti 2	Veri Seti 3	Veri Seti 4	Yöntem'e Göre Mutlak Yüzde Hata Ort.
ROA_DVM_ROA	0,10	0,12	0,09	0,09	0,10
DVM_ROA_ROA	0,08	0,14	0,11	0,10	0,11
ROA_YSA_ROA	0,13	0,14	0,11	0,11	0,12

Test verilerinin yöntemlere göre sonuçları analiz edildiğinde ortalama mutlak yüzde hataları 0,10, 0,11 ve 0,12 olarak bulunmuştur. Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümler'e Göre hibrit modellerin test verilerine uygulanması sonucunda en iyi performans ROA_DVM_ROA yöntemine aittir ve bu yöntemin de ortalama başarı oranı % 90 olarak bulunmuştur.

5. TARTIŞMA VE SONUÇ

Çalışmanın ilk bölümü giriş kısmından oluşmuştur. Bu bölümde çalışmanın amacından, gereklilliğinden, öneminden ve çalışmada kullanılacak yöntemlerden kısaca bahsedilmiştir. Çalışmanın ikinci bölümünde tezde kullanılacak veri setinde girdi değişkenlerinin anlaşılabilmesi için diyabet'in tanımı, sınıflandırılması, teşhisi, tedavisi ve eğitim yöntemleri hakkında temel bilgiler verilmiş ve sonrasında literatür araştırması yapılmıştır. Böylece yeni gelişmeler ve yöntemlerde takip edilmiştir. Yapılan literatür araştırması sonucunda özellikle son 10 yılı içeren çalışmalar detaylı olarak incelenmiş ve tasniflenmiştir. Yapılan literatür araştırmasında tezde kullanılması amaçlanan hibrit modellerin temelini oluşturan yöntemler ile beraber farklı yöntemler (karar ağacı algoritmaları, zaman serileri, hibrit modeller, mobil uygulamalar, genetik algoritmalar v.b.) de incelenmiştir.

Literatürde yer alan çalışmalar araştırıldığında bu tez içerisinde yer alan ve nihai hibrit çözümü oluşturan yöntemlerden; rastgele orman algoritması, yapay sinir ağları ve destek vektör makinaları kullanılarak yapılan çalışmaların olduğu gözlemlenmiştir. Sadece yapay sinir ağlarının uygulandığı çalışmalar olduğu gibi, sadece destek vektör makinaları kullanılan çalışmalar veya sadece rastgele orman algoritmaları uygulanan çalışmalar da gözlemlenmiştir. Bununla birlikte farklı yöntemlerin bir arada uygulanıp, hangi yöntemin daha iyi sonuç verdiği ile ilgili çalışmalarda literatürde gözlemlenmiştir. İlaveten, örnek vermek gerekirse hem yapay sinir ağlarının ve hem de rastgele orman algoritmalarının bir arada kullanıldığı gibi, iki yöntemli hibrit çalışmaların uygulandığı ve hangi ikili hibrit yöntemlerin daha iyi sonuç verdiği çalışmalara da literatürde rastlanılmıştır. Bu çalışmayı diğer çalışmalardan farklı kılan durum; destek vektör makinaları, rastgele orman algoritmaları ve yapay sinir ağları yöntemlerinin bir arada hibrit kullanımıyla ilgili bu tezde olduğu gibi bir çalışmaya rastlanılmamış olması ve elde edilen 27 hibrit çözüm içerisinden en iyi 3 hibrit yöntemin seçilmesi ile yüksek doğrulukla sonuçların elde edilmesi olarak düşünülmüştür.

Çalışmada üçüncü bölümde, hibrit modellerin temelini teşkil edecek yapay sinir ağları yöntemleri ve yapıları, destek vektör makineleri algoritma mantığı ve rastgele orman algoritması tahminleme yöntemleri anlatılmaya çalışılmıştır. Bu yöntemler kullanılarak tahminler yapılmaya başlanmıştır. Devamında çalışmada kullanılan yöntemlerde kullanılacak

performans kriterleri hakkında bilgilendirme yapılmış ve örnek aracılığıyla anlatılmıştır. Bu kısmın son bölümünde çalışmada kullanılan açık kaynak kodlu program olan R-Studio hakkında özet bilgiler verilmiştir.

Çalışmanın dördüncü bölümünde çalışmanın temel amaçlarından olan, hibrit modeller oluşturularak diyabetli hastaların kanındaki glikoz yoğunluğunun tahmin edilmesi yüksek performans ile gerçekleştirilmiştir.

Çalışmada kullanılan veri setinin temin edildiği ilgili yer UCI Makine Öğrenme Havuzundan temin edilmiştir. Çalışmada, 10 adet farklı diyabet hastasının ölçüm değerleri kullanılmıştır. Veri setlerinin tümüne birden yöntemler uygulamak yerine kademe kademe veriler analiz edilmiş ve yöntemler uygulanmıştır. Öncelikle doğru girdi değişkenleri elde etmek amacıyla çalışma çeşitli senaryolar altında analiz edilmiştir. Bu sayede modelin doğruluğunu arttırıcı değişkenler tespit edilmeye çalışılmıştır. Ölçüm zamanları dikkate alınarak, ölçüm zamanları arasındaki farklar hesaplanmıştır. Ölçüm zamanları arasındaki farklar ve ölçüm zamanları arasındaki farkların kümülatifi çalışmalarda ayrı ayrı girdi değişkenleri olarak kullanılmıştır. Ölçüm zamanları arasındaki farkın kümülatifi değişkeninin çalışma modelini daha güçlü olarak temsil ettiği tespit edilmiştir. Çalışmada ölçüm kodları, kan şekeri ölçüm zamanının özelliğini göstermiştir. Bu kodlar, girdi değişkeni olarak çalışmada kullanılmıştır.

Vücudun kan şekerini dengelemede ürettiği 2 tip insulin söz konusudur. Bunlar bazal insulin ve bolus insulin tipleridir. Yemek yenildikten sonra kan şekerini dengelemek amacıyla gerekli olan insulin bolus insülindir. Tıp literatüründe vücudun ürettiği bolus insulin hormonu çalışma mantığındaki insülinler; hızlı sürede etki gösteren insülinler ve kısa sürede etkili olan insülinler olarak yer almaktadır. Kısa etkili insülünler çalışmada regular tip denilen insülinlerdir. Yemek harici vakitte vücudun ürettiği insülinler ise bazal insülinlerdir. Tıp literatüründe vücudun ürettiği bazal insulin hormonu çalışma mantığındaki insülinler, orta süreli etkili insülinler ve uzun süreli etkili insülinlerdir. Bazal insülinler, çalışmada NPH ve Ultralente denilen insulin tipleridir. Bu insulin tiplerinden ne kadar miktarda, ne zaman alındığı bilgileri çalışmada gösterilmiştir. Farklı özelliklerdeki ilaç dozajları etki süreleri dikkate alınarak tek bir ilaç dozu değişkeni haline getirilmiştir. Yani ilaç dozu değişkeni; Regular, NPH ve Ultralente insülin verileri kullanılarak elde edilmiştir. İlaç dozu değişkeni

kanda yer alan ilaç miktarını ifade etmekte ve bu değişken de girdi değişkeni olarak çalışmada kullanılmıştır.

Veri setinde yer alan kan şekeri ölçüm zamanları özelliklerine göre tasniflenmiştir. Bu özellikler; Kahvaltı öğünü öncesi zamanda yapılan kandaki glikoz yoğunluğu (mg/dl) zamanı, kahvaltı sonrası kan şekeri ölçüm zamanı, öğlen yemeği öncesi kan şekeri ölçümü zamanı, öğlen yemeği sonrası kan şekeri ölçüm zamanı, akşam yemeği öncesi kan şekeri ölçüm zamanı, akşam yemeği sonrası kan şekeri ölçüm zamanı, aperatif öncesi kan şekeri ölçüm zamanı, normalden daha fazla yemek yeme etkinliği sonrası kan şekeri ölçüm zamanı v.b. değişken durumları tarih ve saat bilgileri ile beraber gösterilmiştir.

Çalışmada çözümler için sırasıyla veri setleri kullanarak, girdi değişkenleri ve çıktı değişkenlerine bağlı olarak yapay sinir ağları yöntemleri, rastgele orman algoritması ve destek vektör makinaları ve bu yöntemler ile beraber elde edilen hibit modeller kullanılarak sonuçlar elde edilmeye çalışılmıştır. Tahmin işlemleri yapılmış ve performans değerleri analiz edilmeye çalışılmıştır. Yapılan tahmin işlemlerinde performans kriterleri olarak ortalama mutlak sapma, ortalama mutlak yüzde hata değerleri ve tahmin başarı oranları ile elde edilen sonuçlar yorumlanmaya çalışılmıştır.

Çalışmada, öncelikle tüm verilerin birleştirilerek analiz yapılması düşünülmüş fakat bu şekilde farklı farklı hastalardan farklı zamanlarda alınan verilerin istenilen doğrulukta sonuç üretemeyeceğine karar verilmiştir. Nitekim yapılan analizlerde de bu kararın doğru olduğu sonucuna varılmıştır. Bu sebepten ötürü her bir hasta verisi kendi içerisinde analiz edilmiş ve performans sonuçları karşılaştırılmıştır.

Çalışma ile ilgili planlama yapıldıktan sonra farklı senaryolar oluşturulup, modellerin performans sonuçları karşılaştırılmıştır.

Çalışmanın birinci senaryosunda girdi değişikenleri olarak ölçüm tarihleri arasındaki farklar, kan şekeri ölçüm durumunun özelliği ve ilaç dozaj miktarı girdi değişkenleri olarak kullanılmıştır. Çıktı değişkeni olarak kan şekeri ölçüm sonucu kullanılmış ve kandaki glikoz yoğunluğunun tahmini amaçlanmıştır. Bu senaryo altında verilere; rastgele orman algoritması, destek vektör makinaları ve yapay sinir ağ yöntemleri uygulanarak çözümler elde edilmeye çalışılmıştır. Birinci senaryo sonucunda; en iyi yöntemin tahmin başarısı % 76 düzeyinde

bulunmuştur. En iyi yöntem için hataların mutlak değerlerinin ortalaması, ortalama mutlak sapma 43,73 olarak hesaplanmıştır.

Çalışmanın ikinci senaryosunda girdi değişkenlerinde bazı değişiklikler yapılmıştır. Kullanılan girdi değişkenlerinden ölçüm tarihleri arasındaki farklar yerine ölçüm tarihleri arasındaki farkların kümülatifleri alınmıştır. Diğer girdi değişkenleri olarak kan şekeri ölçüm durumunun özelliği ve ilaç dozaj miktarı kullanılmıştır. Çıktı değişkeni olarak da kan şekeri ölçüm sonucu kullanılmıştır. İkinci senaryo sonucunda; en iyi yöntemin tahmin başarısı % 81 düzeyinde bulunmuştur. En iyi yöntem için hataların mutlak değerlerinin ortalaması, ortalama mutlak sapma 38,40 olarak hesaplanmıştır.

Çalışmanın üçüncü senaryosunda; kan şekeri verilerinde meydana gelen alt ve üst uç değerler gözlemlerin dışında bırakılarak, ikinci senaryoda kullanılan girdi ve çıktı değişkenleri kullanılmıştır. Kan şekeri ölçüm sonuçlarının 70-200 arasında olan veriler değerlendirilerek veri seti yeniden düzenlenmiştir. Kan şekerinin 50 altına inmesi hastada bilinç kaybına neden olabilmektedir. Kan şekerinin 70 altına inmesi durumunda hipoglisemi (kan şekeri düşüklüğü) sorunu meydana gelmektedir. Titreme, göz kararması, dil ve dudakta karıncalanma, çarpıntı gibi ciddi sıkıntıların meydana geldiği durumlarda hastanın hayati tehlikesini riske atmamak adına iğne aracılığıyla gerçek kan şekeri ölçümünün yapılmasının daha uygun olacağı düşünülmüştür. Bununla birlikte kan şekerinin 200 ve üzerinde olduğu durumlarda da yüksek kan şekerinden dolayı hastada ciddi rahatsızlıklar meydana gelebilmekte olduğu göz önüne alınmıştır. Senaryo 3'de kullanılan girdi ve çıktı değişkenleri, yeni veri setlerine uygulandığında en iyi yöntemin tahmin başarısı % 82,25 olarak bulunmuştur. En iyi yöntem için hataların mutlak değerlerinin ortalaması, ortalama mutlak sapma 21,71 olarak hesaplanmıştır.

Çalışmada Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözümde; hibrit çözüm yöntemleri veri setlerine uygulanmıştır. Burada uygulanmış iki yöntem uygulamalı hibrit çözümde başlıca 9 adet hibrit model oluşturulmuş ve her bir veri seti için çözümler elde edilmiştir. Oluşturulan hibrit modeller; YSA_DVM, YSA_ROA, YSA_YSA, ROA_YSA, ROA_DVM, ROA_ROA, DVM_ROA, DVM_YSA ve DVM_DVM şeklinde oluşturulmuştur. Burada önek teşkil etmesi bakımında YSA_DVM hibrit modeli açıklaması; veri setine bu 2 yöntemin birlikte uygulanarak hibrit modelin oluşturulması ve önce yapay sinir ağları kullanılması, devamında

destek vektör makinaları kullanılarak sonuçların elde edilmesi anlamına gelmiştir. Senaryo 3 İki Yöntem Uygulamalı Hibrit Çözümler'de oluşturulan hibrit çözüm yöntemleri, veri setlerine uygulandığında en iyi yöntemin tahmin başarısı % 88 olarak bulunmuştur. En iyi yöntem için hataların mutlak değerlerinin ortalaması, ortalama mutlak sapma 13,98 olarak hesaplanmıştır.

Çalışmada Senaryo 3 Üç Yöntem Uygulamalı Hibrit Çözümde; hibrit çözüm yöntemleri geliştirilerek veri setlerine uygulanmıştır. Burada başlıca 27 adet hibrit model oluşturulmuş ve her bir veri seti için çözümler elde edilmiştir. Oluşturulan hibrit modeller; YSA_DVM_ROA, YSA_DVM_YSA, YSA_DVM_DVM v.b. şekilde tasarlanmıştır. Burada YSA_DVM_ROA hibrit modeli, veri setine bu 3 yöntemin birlikte uygulanarak hibrit modelin oluşturulmuş olduğunu ve önce yapay sinir ağları kullanıldığını, devamında destek vektör makinaları kullanıldığını ve son olarak rastgele orman algoritmaları kullanılarak sonuçların elde edildiği anlamına gelmiştir. Oluşturulmuş 27 adet hibrit yöntemin 3*3=9 birbirinden farklı en iyi hibrit modellerin performansları analiz edilmiştir. Bunun sebebi en iyi performans gösterebilecek hibrit modelleri bulup uygulayıp, çalışmaya hız, esneklik ve kolaylık sağlayabilmesi olarak düşünülmüştür. Burada oluşturulan başarılı hibrit çözüm yöntemleri, veri setlerine uygulanmış ve hibrit modellerin performansları analiz edilmiştir.

Çalışmanın bir sonraki aşamasında Senaryo 3 En İyi Üç Yöntem Uygulamalı Hibrit Çözümlerde en yüksek yüzdeli doğrulukla tahmin etme başarısı gösteren 3 hibrit model tespit edilmiştir. Bu koşullar altında oluşturulan modeller; ROA_DVM_ROA, DVM_ROA_ROA ve ROA_YSA_ROA, hibrit modelleri olarak belirlenmiştir. Oluşturulan hibrit modellerin sonucunda ortalama başarı en iyi yöntemin tahmin başarısı % 91 olarak bulunmuştur. En iyi yöntem için hataların mutlak değerlerinin ortalaması, ortalama mutlak sapma 11,58 olarak hesaplanmıştır. Bu hibrit modellerin sonuçlarına göre en iyi hata ortalama mutlak sapma, ortalama mutlak yüzde hata gibi performans sonuçları gerçekleşmiştir.

Çalışmada önerilen hibrit modeller kullanılarak, diyabet hastalarının kanlarındaki glikoz yoğunlukları yüksek doğrulukla tahmin edilmiştir. Tahminler, her hastanın hikayesine göre özel olarak yapılmıştır. İleride yapılacak çalışmalarda her hasta için ayrı ayrı model oluşturmak yerine, tüm hastaların kullanabileceği genel bir çerçeve içeren bir formül model oluşturulabilir. Gelecekte yapılacak tahminler için gözlem altında kayıt edilen hastalarda

çeşitli nedenlerden kaynaklı meydana gelebilecek veri kayıpları durumlarında, bu formül model çalışmalara eklenebilir. Ve bu sayede oluşturulacak yeni hibrit modeller ile veri sayısı yeterli olmayan hastalar için de yüksek performanslı tahmin sonuçları elde edilebillir. Bununla birlikte diyabet hastalarının beslenme alışkanlıklarına ve fiziksel aktivite düzenlerine ilişkin öneri verecek şekilde yeni çalışmalar da geliştirilebilir.

Hibrit teknolojiler ve hibrit yöntemler son yıllarda her alanda yaygınlaşmaya ve kullanılmaya başlanmıştır. Yapılan çalışmalarda tek bir yöntemden ziyade, farklı yöntemlerin birarada kullanıldığı çalışmalarla daha iyi sonuçlar alındığının fark edilmesi araştırmacıların bu yönde çalışmalarına neden olmuştur. Tek bir bilim disiplini için bu tip çalışmalar yapılabileceği gibi, farklı disiplinlerin bir arada olduğu disiplinlerarası alanlar için de bu tür çalışmaların yapılmaya-yaygınlaşmaya başladığı görülmüştür.

Gelişen teknolojik imkanlarla, tıp ve mühendislik alanlarını içeren çalışmaların, mobil uygulamalara entegre edilmesiyle, bu çalışmaların geniş çaplı kullanım alanına sahip olması amaçlanabilir. Mobil teknolojilerdeki gelişimler ile hastaların ilaç verileri, beslenme verileri, fiziksel aktivite verileri gibi önemli girdi verileri akıllı telefonlar, akıllı saatler v.b. akıllı cihazlar aracılığıyla toplanabilir. Toplanan veriler ile sonuç üretimi ve yorumlama mekanizmaları devreye sokularak çalışmanın boyutu değiştirilebilir. Günümüzde vazgeçilmez kullanım alanına sahip teknolojilerden olan akıllı iletişim teknolojileri sayesinde bu tarz çalışmaların gündelik hayata entegresi sağlanabilir. Bununla birlikte önemli hastalıkların takibinde ve tedavisinde de mobil cihazlardan, mobil teknolojilerden yararlanılabilir. Bu tarz sağlık ve mühendislik bilimlerinin bir arada olduğu çalışmalar, insan yaşam kalitesini arttırabileceğinden dolayı öncelikli çalışma alanları arasında değerlendirilmektedir. Bununla beraber kullanıcıları, hastalık meydana gelmeden, artan hastalık riskleri sebebiyle uyarabilen yazılımlar ve programlar da geliştirilebilir.

KAYNAKLAR

ADA (American Diabetes Association), 2011, *Diagnosis and Classification of Diabetes Mellitus. Diabetes Care*, https://www.diabetes.org/, 34(1), 62-69, [Ziyaret tarihi: 10 Şubat 2019].

ADA (American Diabetes Association), 2012, *Diagnosis and Classification of Diabetes Mellitus. Diabetes Care*, https://www.diabetes.org/, 35(1), 64-71, [Ziyaret tarihi: 10 Şubat 2019].

ADA (American Diabetes Association), 2013, *Standards of Medical Care in Diabetes Diabetes Care*, https://www.diabetes.org/, 36(1), 11-66, [Ziyaret tarihi: 10 Şubat 2019].

ADA (American Diabetes Association), 2018, *Diagnosing Diabetes and Learning About Prediabetes*, http://www.diabetes.org/diabetes-basics/diagnosis/?loc=db-slabnav, [Ziyaret tarihi: 20 Mart 2018].

Akar Ö., Güngör O., Akar A., 2010, Rastgele orman sınıflandırıcısı ile arazi kullanım alanlarının belirlenmesi, *III. Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu,* 11-13 Ekim 2010, Gebze, 142-152.

Akar, Ö., 2013, *Rastgele orman sınıflandırıcısına doku özellikleri entegre edilerek benzer spektral özellikteki tarımsal ürünlerin sınıflandırılması*, Doktora Tezi, KTÜ Fen Bilimleri Enstitüsü.

Albayrak, A.S., Yılmaz, K., 2009, Veri madenciliği: Karar ağacı algoritmaları ve İMKB verileri üzerine bir uygulama, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14 (1), 31-52.

Alfian, G., Syafrudin, M., Anshari, M., Benes, F., Atmaji F., Fahrurrozi I., Hidayatullah, A., Rhee, J.,2020, Blood glucose prediction model for type 1 diabetes based on artificial neural network with time-domain features, *Biocybernetics and Biomedical Engineering*, 40 (4), 1586-1599.

Aras, S., 2008, *Yapay sinir ağlarına duyarlılık analizleri*, Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.

Atmaca, M.H. ve Ecemiş, G.C., 2012, Oral antidiyabetik ajanlar, *Journal of Experimental and Clinical Medicine*, 29 (1), 23-29.

Ayas, S., 2014, *Mikroskobik imgelerde tüberküloz bakterisinin rastgele ormanlar yöntemiyle sınıflandırılması*, Yüksek Lisans Tezi, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü.

Ayhan, S., Erdoğmuş, Ş., 2014, Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi, *Eskişehir Osmangazi Üniversitesi İİBF Dergisi*, 9 (1), 175-198.

Balakrishnan, N.P., Rangaiah, G.P., Samavedham, L., 2012, Personalized blood glucose models for exercise, meal and insulin interventions in type 1 diabetic children, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 12 Kasım 2012, 3, 1250–1253.

Balakrishnan, N.P., Samavedham, L., Rangaiah, G.P., 2013, Personalized hybrid models for exercise, meal, and insulin interventions in type 1 diabetic children and adolescents, *Industrial and Engineering Chemistry Research*, 52 (36), 13020–13033.

Balakrishnan, N.P., Samavedham, L., Rangaiah, G.P., 2014, Personalized mechanistic models for exercise, meal and insulin interventions in children and adolescents with type 1 diabetes, *Theoretical Biology*, 357, 62–73.

Bamgbose, S.O., Li, X., Qian L., 2017, Closed loop control of blood glucose level with neural network predictor for diabetic patients, *International Conference on e-health networking, applications and services*, 18 Aralık 2017, Dalian, Çin.

Bauer, E. ve Kohavi, R., 1999, An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants, *Machine Learning*, 36, 105-139

Bayır, F., 2006, *Yapay sinir ağları ve tahmin modellemesi üzerine bir uygulama*, Yüksek Lisans Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.

Bergman, R.N., Ider, Y.Z., Bowden, C.R., Cobelli, C., 1979, Quantitative estimation of insulin sensitivity, *American journal of physiology*, 236 (6), 667–677.

Bergman, R.N., Phillips, L.S., Cobelli, C., 1982, Physiologic evaluation of factors controlling glucose tolerance in man, *The journal of clinical investigation*, 68 (6), 1456–1467.

Bilen, M., 2014, *Yapay Sinir Ağları İçin Web Tabanlı Bir Eğitim Yazılımı Geliştirilmesi*, Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü.

Bock, A., François, G., Gillet, D., 2015, A therapy parameter- based model for predicting blood glucose concentrations in patients with type 1 diabetes, *Computer Methods and Programs in Biomedicine*,118(2), 107–123.

Botwey, R. H., Daskalaki, E., Diem, P. ve Mougiakakou, S,G., 2014, Multi- model data fusion to improve an early warning system for hypo- /hyperglycemic events, *36th Annual international conference of the IEEE engineering in medicine and biology society*, 26 Ağustos 2014, IEEE, 4843–4846.

Breiman, L., ve Cutler, 2017, A., Random Forest,http://www.stat.berkeley.edu/~breiman/RandomForests, [Ziyaret tarihi: 15 Ocak 2019].

Breiman, L., 2001, Random Forests, *Machine Learning*, 45, 5-32.

Boas, L.C., Foss, M.C., Foss de Freitas, M.C. ve Pace, A.E., 2012, Relationship Among Social Support, Treatment Adherence and Metabolic Control of Diabetes Mellitus Patients, *Revista Latino-Americana Enfermagem*, 20 (1), 52-58.

Boiroux, D., Duun- Henriksen, A.K., Schmidt, S, 2012, Overnight control of blood glucose in people with type 1 diabetes, *IFAC Symposium on Biological and Medical Systems*, 45 (18), 73–78.

Bremer, T., ve Gough, D.A., 1999, Is blood glucose predictable from previous values? A solicitation for data, *Diabetes* 48, 445-451.

Bunescu, R., Struble, N., Marling, C., Shubrook, J. ve Schwartz, F., 2013, Blood glucose level prediction using physiological models ve support vector regression, *In Proceedings of the IEEE 12th ınternational conference on machine learning and applications*, 2013 Miami, FL: IEEE.

Butwall, M. ve Kumar, S., 2015, A Data Mining Approach for the Diagnosis of Diabetes Mellitus using Random Forest Classifier, International Journal of Computer Applications, 120 (8), 36-39.

Calm, R., García- Jaramillo, M., Bondia, J., Sainz, M.A., Vehi, J., 2011, Comparison of interval and Monte Carlo simulation for the prediction of postprandial glucose under uncertainty in type 1 diabetes mellitus, *Computer Methods and Programs in Biomedicine*, 104 (3), 325–332.

Cescon, M., Johansson, R., Renard E., 2015, Subspace- based linear multi- step predictors in type 1 diabetes mellitus, *Biomedical Signal Processing and Control*, 22, 99–110.

Chemlal, S., Colberg, S., Satin-Smith, M., Gyuricsko, E., Hubbard, T., Scerbo, M. W. ve McKenzie, F. D., 2011, Blood glucose individualized prediction for type 2 diabetes using iphone application, *37th Annual Northeast bioengineeering conference*, 1 Nisan 2011, Nebec, IEEE, 1 – 2.

Cherkassky., V., 1998, *Learning from Data*, John Wiley and Sons, 23/0-471-15493- 8.

Chowdhury, S.R., Nandi, B., Mondal, P., 2018, A Non-Invasive Blood Insulin and Glucose Monitoring System Based on Near-Infrared Spectroscopy with Remote Data Logging, *International Symposium on Computer-Based Medical Systems*, 18 Haziran 2018, IEEE, 274-279.

Cutler, A. ve Breiman, L., 2003, *Random forests manual v4*, UC Berkel, Techinal report.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J.C., Lawler, J.J., 2007, Random forests for classification in ecology, *Ecology*, 88 (11), 2783-2792.

Çağlayan, S., https://www.florence.com.tr/kan-sekerinin-yukselmesi-hiperglisemi-belirtileri, [Ziyaret tarihi:28 Ocak 2018].

Çanayaz, E., 2014, *Tip-1 Diyabet Hastaları İçin Bulanık Mantık Tabanlı Kontrol Edici Kullanarak Uygun İnsülin Dozlarının Önerilmesi*, Yüksek Lisans Tezi, İTÜ Fen Bilimleri Enstitüsü.

Çavuşlu, M.A., Becerikli, Y. ve C. Karakuzu, C., 2012, Levenberg-Marquardt Algoritması ile YSA Eğitiminin Donanımsal Gerçeklenmesi, *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5 (1).

Çayıroğlu, İ., http://www.ibrahimcayiroglu.com/Dokumanlar/IleriAlgoritmaAnalizi/Ileri Algoritma Analizi-5.Hafta-YapaySinirAglari.pdf, [Erişim Tarihi: 10 Nisan 2019].

Çomak, E., 2004, *Destek vektör makineleri çoklu sınıf problemleri için çözüm önerileri*, Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü.

Demirci, D. A., 2007, *Destek vektör makineleri ile karakter tanıma*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü.

Dietterich, T., 2000, An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization, *Machine Learning*, 40, 139-157.

Duun- Henriksen, A.K., Schmidt, S., Roge, R.M., 2013, Model identification using stochastic differential equation grey- box models in diabetes, *Journal of Diabetes Science Technology*, 7 (2), 431–440.

Efendic, H., Kirchsteiger, H., Freckmann, G., del, Re, L., 2014, Short- term prediction of blood glucose concentration using interval probabilistic models, 22nd *Mediterranean conference on control and automation*, 16-19 Haziran 2014, 1494–1499.

Efron, B., 1979, Bootstrap Methods: Another Look at the Jackknife, *Institute of Mathematical Statistics*, 7 (1), 1-26.

Eljil, K. S., Quadah, G., Pasquer, M., 2013, Predicting hypoglycemia in diabetic patients using data mining techniques, *Proceedings of IEEE, 9th International conference on innovations in information technology*, 17-19 Mart 2013, IIT, 130 – 135.

Elmas, P.D.Ç., 2007, *Yapay Zeka Uygulamaları*, Seçkin Yayıncılık, Ankara, ISBN: 978975-0216961.

Farahmand, B., Dehghani, M. ve Vafamand, N., 2019, Fuzzy model-based controller for blood glucose control in type 1 diabetes: An LMI approach, *Biomedical Signal Processing and Control*, 54, 101627.

Fernandez, D., Canete, J., Gonzalez- Perez, S. ve Ramos- Diaz, J.C., 2012, Artificial neural networks for closed loop control of in silico and ad hoc type 1 diabetes, *Computer methods programs in biomedicine*, 106 (1), 55–66.

Fong, S., Mohammed, S., Fiaidhi, J., Kwoh, C.K., 2013, Using causality modeling and fuzzy lattice reasoning algorithm for predicting blood glucose, *Expert systems with applications*, 40 (18), 7354–7366.

Fong, S., Zhang, Y., Fiaidhi, J., Mohammed, O., Mohammed, S., 2013, Evaluation of stream mining classifiers for real- time clinical decision support system: A case study of blood glucose prediction in diabetes therapy, *Biomed research international*.

Gandia, C.P., Facchinetti, A., Sparacino, G., Cobelli, C., Gomez, E.J., Rigla, M., Leiva, A., Hernando, M.E., 2007, Artificial Neural Network Algorithm for On-Line Glucose Prediction from Continuous Glucose Monitoring, *Networking Research Center on Bioengineering, Biomaterials and Nanomedicine*, Madrid, İspanya.

Gani, A., Gribok, A.V., Ward, W.K., Vigersky, R.A., 2011, Reifman J. Universal glucose models for predicting subcutaneous glucose concentration in humans, *IEEE Transactions on Information Technology in Biomedicine*, Ocak 2011, 14 (1), 157–165.

Georga, E.I., Protopappas, V.C., Guillen, A., Fico, G., Ardigo, D., Arredondo, M. T., Exarchos, T. P., Polyzos, D., Fotiadis, D. I., 2009, Data mining for blood glucose prediction and knowledge discovery in diabetic patients: The METABO Diabetes Modeling and Management System, *Proceedings of IEEE, Annual Intermational Conference on Engineering in Medicine and Biology Society*, EMBC, 3-6 Eylül 2009, 5633 – 5636.

Georga, E.I., Protopappas, V.C., Polyzos, D., 2010, Prediction of glucose concentration in type 1 diabetic patients using support vector regression, *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine*, 2010, 3-5 Kasım 2020, 26 (500), 1–4.

Georga, E. I., Fotiadis, D. I., ve Protopappas, V. C., 2011, Glucose prediction in type 1 and type 2 diabetic patients using data driven techniques. In K. Funatsu, *Knowledge-oriented applications in data mining*, 21 Ocak 2011, 277–295.

Georga, E.I., Protopappas, V.C., Polyzos, D., Fotiadis, D.I., 2012, A predictive model of subcutaneous glucose concentration in type 1 diabetes based on random forests, *Conference Proceedings IEEE, Enginering medicine and biology society*, 2012, 2889–2892.

Georga, E.I., Protopappas, V.C., Ardigo, D., Marina, M., Zavaroni, I, Polyzos, D., Fotiadis D.I., 2013, Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression, *IEEE Journal of Biomedical and Health Informatics*, 17 (1), 71-81.

Guo, H., Zhao, J.Y., Yin, J.H., 2017, Random forest and multilayer perceptron for predicting for dielectric loss of polyimide nanocomposite films, *The Royal Society of Chemistry Advances*, 7, 30999-31008.

Güraksın, G.E., 2015, *Yapay zeka teknikleri kullanarak kemik yaşı tespiti*, Doktora Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü.

Gürsoy, İ.M. ve Subast, A., 2008, A comparison of PCA, ICA and LDA in EEG signal classification using SVM, *Communication and Applications Conference*, 26 Eylül 2008.

Gyuk, P., Vassanyi, I., Kosa, I., 2019, Blood Glucose Level Prediction for Diabetics Based on Nutrition and Insulin Administration Logs Using Personalized Mathematical Models, *Journal of Healthcare Engineering*.

Hager, W.W. ve Zhang, H., 2006, A survey of nonlinear conjugate gradient methods, *Pacific of Journal Optimization*, 2 (35), 35–58.

Hamdi, T., Ali, J.B., Nader, Veronique, Di, C., Farhat, F., 2017, Artificial neural network for blood glucose level prediction, *International conference on smart, monitored and controlled cities*, Kerkennah, Tunus, 17-19 Şubat 2017.

Harris, P., Mann, L., Phillips, P. ve Webster C., 2012, *Diabetes Management in General Practice*, Eighteenth edition, Australia, Diabetes Australia Limited.

Harsh, S., Molenaar, P. ve Freeman, K., 2013, Developing personalized empirical models for type-I diabetes: An extended Kalman filter approach., *American Control Conference*, 16 Ağustos 2013, 2923–2928.

Hidalgo, J.I., Colmenar, J.M., Risco- Martin, J.L., 2014, Modeling glycemia in humans by means of grammatical evolution, *Applied Soft Computing*, 20, 40–53.

Hidalgo, J.I., Colmenar, J.M., Kronberger, G., Winkler, S.M., Garnica, O., Lanchares, J., 2017, Data Based Prediction of Blood Glucose Concentrations Using Evolutionary Methods, Journal of Medical Systems, 41 (9), 142.

Hovorka, R., Canonico, V., Chassin, L.J., 2004, Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes, *Physiological measurement*, 25 (4), 905–920.

https://sbu.saglik.gov.tr/Ekutuphane/kitaplar/Diyabet%20Tedavi%20ve%20%C4%B0zlem%20Klinik%20Protokolleri.pdf, [Ziyaret tarihi:03 Ocak 2019].

https://www.diabetesatlas.org/en/, [Ziyaret tarihi:28 Aralık 2018, 20 Ocak 2019, 10 Şubat 2020].

http://www.istanbulsaglik.gov.tr/w/anasayfalinkler/belge/dunyada_turkiyede.pdf, [Ziyaret tarihi:28 Ocak 2019].

https://www.memorial.com.tr/saglik-rehberleri/insulin-direnci-nedir, [Ziyaret tarihi:08 Kasım 2018].

https://www.ntv.com.tr/saglik/turkiyede-diyabet-hastasi-sayisi-dunyanin-2-kati,qPHrI-AdjkW1SERDjuSJxA, [Ziyaret tarihi:08 Kasım 2018].

https://www.turkdiab.org/, [Ziyaret tarihi:03 Ocak 2019].

https://www.webofknowledge.com, [Ziyaret tarihi:08 Ocak 2019].

https://archive.ics.uci.edu/ml/datasets/Diabetes, [Ziyaret tarihi:08 Kasım 2018].

Isfahani, M.K., Zekri, M., Marateb, H.R., Faghimani, E., 2020, A Hybrid Dynamic Wavelet-Based Modeling Method for Blood Glucose Concentration Prediction in Type 1 Diabetes, *Journal of Medical Signals and Sensors*, 10 (3), 174-184.

Islam, M.S., Leech, J., Lin, C.C.Y., Chrostowski, R., 2007, Peak blood glucose prediction algorithm following in a meal intake, *Proceedings of IEEE, Canadian conference on electrical and computer engineeering*, CCECE, 30 Temmuz 2007, 579 – 582.

Kabaoğlu, R.O, 2010, *Destek Vektörü Makineleri Tabanlı Hata Bulma, Tanıma ve Hata Toleranslı Kontrol Yöntemleri*, Doktora Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.

Kahn, M., 1994, The Diabetes Dataset, *Diabetes Data Set,* https://archive.ics.uci.edu/ml/datasets/Diabetes, [Ziyaret tarihi: 20 Ocak 2018].

Kahn, M., 1994 "The Diabetes Dataset", *AI in Medicine Symposium AIM-94*, 1994, https://archive.ics.uci.edu/ml/datasets/Diabetes, [Ziyaret tarihi: 20 Ocak 2018].

Kabalcı, E., 2015, https://docplayer.biz.tr/47496115-Yapay-sinir-aglari-artificial-neural-networks-doc-dr-ersan-kabalci.html, [Ziyaret tarihi: 25 Mart 2018].

Kartal, M., 2006, *Bilimsel Araştırmalarda Hipotez Testleri*, Nobel Akademik Yayıncılık, Ankara, ISBN:9786051337487.

Kıraç, N., 2011, *Türkiye'de faaliyet gösteren ticari bankaların finansal risklerinin yapay sinir ağları yaklaşımı ile belirlenmesi*, Yüksek Lisans Tezi, Çukurova Üniversitesi Sosyal Bilimler Enstitüsü.

Kim, H.S. ve Jeong, H.S, 2007, A nurse short message service by cellular phone in type-2 diabetic patients for six months, *Journal of Clinical Nursing*, 16 (6), 1082-7.

Kirchsteiger, H., Polzer, S., Johansson, R., Renard, E., del, Re, L., 2011, Direct continuous time system identification of MISO transfer function models applied to type 1 diabetes, IEEE Conference On Decision Control European Control Conference, 12-15 Kasım, 2011,1, 5176–5181.

Kondiloğlu, A., 2015, *Diyabet akıllı destek sistemi: Bilgi erişim teknikleri ile kan şekeri seviyesinin tahmini*, Yüksek Lisans Tezi, Türk Hava Kurumu Üniversitesi Fen Bilimleri Enstitüsü.

Kormaz, D., Çelik, E., Kapar, M., 2018, Sınıflandırma ve Regresyon Ağaçları ile Rastgele Orman Algoritması Kullanarak Botnet Tespiti: Van Yüzüncü Yıl Üniversitesi Örneği, *Journal of the Institute of Natural & Applied Sciences*, 23 (3), 297-307.

Kuzey, C., 2012, *Veri Madenciliğinde Destek Vektör Makinaları ve Karar Ağaçları Yöntemlerini Kullanarak Bilgi Çalışanlarının Kurum Performansı Üzerine Etkisinin Ölçülmesi ve Bir Uygulama*, Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.

Laguna, A.J., Rossetti, P., Ampudia- Blasco, F.J., Vehi, J., Bondia, J., 2014, Identification of intra- patient variability in the postprandial response of patients with type 1 diabetes, *Biomedical Signal Processing and Control*, 12 (1), 39–46.

Lai, W.H., Chen, C.W., Hsieh, P.H., 2017, Application of Decision Tree on Collision Avoidance System Design and Verification for Quadcopter, *Remote Sensing and Spatial Information Sciences*,42 (2),71-75

Lehmann, E.D. ve Deutsch, T., 1992, Physiological model of glucose-insulin interaction in type 1 diabetes mellitus, *Journal of biomedical engineering*, 14, 235.

Leon, B.S., Alanis, A.Y., Sanchez, E.N., Tellez, O., ve Velazquez, E.R., 2012, Inverse optimal neural control of blood glucose level for type 1 diabetes mellitus patients, Journal of the Franklin Institute, 349 (5), 1851-1870.

Li, J. ve Fernando, C., 2016, Smartphone-based personalized blood glucose prediction, *The Korean Institute of Communications Information Sciences*, 150-154.

Li, P., Yu, L., Wang, J., Guo, L., Fang, Q., 2015, Effect of meal intake on the quality of empirical dynamic models for type 1 diabetes, *IEEE Int Symp Bioelectron Bioinforma*, 1, 1-4.

Liu, S.W., Huang, H.P., Lin, C.H., Chien I.L., 2013, Fuzzy- logic- based supervisor of insulin bolus delivery for patients with type 1 diabetes mellitus, *Industrial & Engineering Chemistry Research*, 52 (4), 1678–1690.

Liu, S.W., Huang, H.P., Lin, C.H., Chien I.L., 2014, Modified control algorithms for patients with type 1 diabetes mellitus undergoing exercise, *Industrial & Engineering Chemistry Research*, 45(5), 2081–2095.

Lu, Y., Gribok, A.V., Ward, W.K., Reifman, J., 2010, The importance of different frequency bands in predicting subcutaneous glucose concentration in type 1 diabetic patients, *IEEE Trans Biomed Eng.*, 57(8), 1839–1846.

Lu, Y., Rajaraman, S., Ward, W.K., Vigersky, R.A., Reifman, 2011, Predicting human subcutaneous glucose concentration in real time: a universal data- driven approach, *International Conference of the IEEE Engineering Medicine and Biology Society*, 7945–7948.

Maltarollo, V.G., Honorio K.M., Ferreira A.B., 2013, Applications of Artificial Neural Networks in Chemical Problems, 16 Ocak 2013.

Mamdani, E.H. and Assilian, S.,1975, An experiment in linguistic synthesis with a fuzzy logic controller, *International journal of man-machine studies*, 7(1), 1–13.

Mayo, M., Chepulis, L. ve Paul R.G., 2019, Glycemic-aware metrics and oversampling techniques for predicting blood glucose levels using machine learning, *Journal Plos One*.

McShinsky, R. ve Marshall B., 2020, Comparison of Forecasting Algorithms for Type 1 Diabetic Glucose Prediction on 30 and 60-Minute Prediction Horizons, *ECAI 2020: Santiago de Compostela*, 29-30 Ağustos, Spain, 12-18.

Mingers, J., 1989, An Empirical Comparison of Pruning Methods for Decision Tree Induction, *Machine Learning*, 4, 227-243.

Moller, M.F., 1993, A scaled conjugate gradient algorithm for fast supervised learning, *Neural networks*, 6(4), 525–533.

More, J.J., 1978, The Levenberg-Marquardt algorithm: implementation and theory, *Numerical analysis*, Springer, Berlin, Heidelberg, 105-116.

Mougiakakou, S. G., Prountzou, K. ve. Nikita K. S., 2005, A real time simulation model of glucose-insulin metabolism for type 1 diabetes patients, *International Conference of the Engineering in Medicine and Biology Society*, IEEEEMBS, 298-301.

Moura, M., Zio, E., Lins, I. ve Droguett, E., 2011, Failure and reliability prediction by support vector machines regression, *Reliability engineering and system safety*, 96,1527-1534.

Murphy GJ, Holder JC., 2000, PPAR-gamm agonists: therapeutic role in diabetes, inflammation and cancer, *Trends pharmacological sciences*, 21, 469-74.

Naumova, V., Pereverzyev, S.V., Sivananthan, S., 2012, A meta- learning approach to the regularized learning- case study: Blood glucose prediction, *Neural Networks*, 33, 181–193.

Novara, C., Pour, N.M., Vincent, T., Grassi, G., 2015, A nonlinear blind identification approach to modeling of diabetic patients, *IEEE Trans. Control Syst.*, 2015, 1–9.

Olgun, N., 2015, https://docplayer.biz.tr/3534637-Insulin-tedavisinin-yonetimi-prof-dr-nermin-olgun-acibadem-universitesi-sbf-hemsirelik-bolumu.html, [Ziyaret tarihi:05 Ocak 2019].

Osowski, S., Siwekand, K., ve Markiewicz, T., 2004, MLP and SVM Networks–A Comparative Study, *Proceedings of the 6th Nordic Signal Processing Symposium*, Şubat 2004, Norsig.

Özkan, Y., 2008, *Veri Madenciliği Yöntemleri*, Papatya Bilim, İstanbul, ISBN:9789756797822.

Özkaya, A. U., Kaya, M. E., Gürgen, F., 2005, Destek vektör makineleri kullanılarak aritmi sınıflandırılması, *XI. Biyomedikal Mühendisliği Ulusal Toplantısı*, 12-16.

Özkaya, D., 2018, *Aida simülatörü üzerinde kan şekeri tahmini için makine öğrenmesi algoritmalarının kıyaslanması*, Yüksek Lisans Tezi, Boğaziçi Üniversitesi Biyo-Medikal Mühendislik Enstitüsü.

Öztemel, E., 2003-2006-2012, *Yapay Sinir Ağları*, Papatya Bilim, İstanbul, ISBN:9789756797396.

Öztürk, K. ve Şahin M.E., 2018, Yapay Sinir Ağları ve Yapay Zekâ'ya Genel Bir Bakış, *Takvim-i Vekayi*, 6 (2), 25-36.

Pappada, S. M., Cameron, B. D., ve Rosman, P. M., 2008, Development of a neural network for prediction of glucose concentration in type I diabetes patients, *Journal of Diabetes Science Technology*, 2, 792-801.

Pereda, D., Romero- Vivo, S., Ricarte, B., Bondia, J., 2012 On the prediction of glucose concentration under intra- patient variability in type 1 diabetes: a monotone systems approach, *Computer Methods and Programs in Biomedicine.*, 108 (3), 993–1001.

Perez-Gandia, C., Facchinetti, A., Sparacino, G., Cobelli, C., Gomez, E. J., Rigla, M. ve Hernando, M. E., 2010, Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring, *Diabetes Technology & Therapeutics*, 2010, 12 (1), 81–88.

Piccini, G., 2004, The First Computational Theory of Mind and Brain: A Close Look at Mcculloch and Pitts's 'Logical Calculus of Ideas Immanent in Nervous Activity', *Synthese*, 141, 175-215.

Platt, J.,1998, Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, *Microsoft Research,* 14, 1-21.

Reifman, J., Rajaraman, S., Gribok, A. ve Ward, W. K., 2007, Predictive monitoring for improved management of glucose levels, *Journal of Diabetes Science and Technology*, 1, 478-486.

Robertson, G., Lehmann, E. D., Sandham, W. ve David H., 2011, Blood glucose prediction using artificial neural networks trained with the AIDA Diabetes Simulator: A Proof-of-Concept pilot study, *Journal of Electrical and Computer Engineering*, 681786.

Sacks, D.B., 2011, A1C Versus Glucose Testing: A Comparison, *Diabetes Care*, 34, 518-523.

Saraç, T., 2004, *Yapay sinir ağları*, Gazi Üniversitesi Endüstri Mühendisliği Bölümü Ana Bilim Dalı Seminer Projesi, 71.s.

Sandham, W. A., Nikoletou, D., Hamilton, D. J., Patterson, K., Japp, A. ve MacGregor, C., 1998, Blood glucose prediction for diabetes therapy using a recurrent artificial neural network, *IX European Signal Processing Conference (EUSIPCO), Rodos, 8-11 Eylül* 1998, 673-676.

Shanti, S., Balamurugan, P. ve Kumar D., 2012, Performance comparison of featured neural network with gradient descent and levenberg-marquart algorithm trained neural networks for prediction of blood glucose values with continuous glucose monitoring sensor data, 13-14 Aralık 2012, Hindistan, IEEE.

Sharma, B. ve Venugopalan, K., 2014, Comparison of neural network training functions for hematoma classification in brain CT Images, *IOSR Journal of Computer Engineering*, 16 (1), 31-35.

Sparacino, G., Zanderigo, F., Maran, A. ve Cobelli, C., 2006, Continuous glucose monitoring and hypo/hyperglycaemia prediction, *Diabetes Research and Clinical Practice*, 74, 160-163.

Soylu, S. ve Danisman, K., 2017, Yaygın Üç Tip 1 Diyabet Benzetim Modelinin Karşılaştırılması, *Otomatik Kontrol Ulusal Toplantısı*, Eylül 2017, İstanbul.

Suchetana, B., Rajagopalan, B., Silverstein, J., 2017, Assessment of wastewater treatment facility compliance with decreasing ammonia discharge limits using a regression tree model, *Science of the Total Environment,* 598, 249- 257.

Taşkaya, S., 2014, *Diyabet Hastalarının Tedaviye Uyum Düzeyleri İle Sağlık Hizmeti Kullanımı ve Yaşam Kalitesini Etkileyen Faktörler,* Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.

Timar, B., Timar, R., Frandes, M. ve Lungeanu, D., 2017, Chaotic time series prediction for glucose dynamics in type 1 diabetes mellitus using regime-switching models, *Scientific Reports*, 6232.

Tresp, V., Briegel T., ve Moody, J., 1999, Neural-network models for the blood glucose metabolism of a diabetic, *Transactions Neural Networks, Ekim* 1999, *IEEE*, 10, 1204-1213.

Tso, B. ve Mather, P., 2009, *Classification Methods for Remotely Sensed Data*, CRC Press, ISBN:9781420090727.

Turksoy, K., Quinn, L., Littlejohn, E., Cinar, A., 2014, Multivariable adaptive identification and control for artificial pancreas systems, *Transactions on Biomedical Engineering*, Mart 2014, 61(3), 883–891.

Ulate, K.P., Lima, G.C., Bielefeld, M.R., Morales, J.M. ve Rotta A.T., 2008, Strict Glycemic Targets Need Not Be So Strict: A More Permissive Glycemic Range for Critically Ill Children, *Pediatrics*, Ekim 2008.

Vapnik, V. N., 1998, *Statistical Learning Theory*, John Wiley & Sons, New York, ISBN: 978-0-471-03003-4.

Vapnik, N.V., Chervonenkis, A.Y., 1979, On the Uniform Convergence of Relative Frenquescies of Events to Their Probabilities, Theory of Probability and its Applications, *Mathematics, Computer Science*, 16 (2), 264-280.

Xiao, W., Jun, J.i., Fengjing, S., Rencheng, S., Chunxiao, X., 2015, Fasting Blood Glucose Change Prediction Model Based on Medical Examination Data and Data Mining Techniques, *International Conference on Smart City*.

Wang, Y., Wu, X., Mo, X., 2013, A novel adaptive- weighted- average framework for blood glucose prediction, *Diabetes Technology Therapeutics*, 15 (10), 792–801.

Wang, Q., Molenaar, P., Harsh, S., 2014, Personalized state- space modeling of glucose dynamics for type 1 diabetes using continuously monitored glucose, insulin dose, and meal intake: an extended Kalman filter approach, *Journal of Diabetes Science and Technology*, 8 (2), 331–345.

Watts, J. D., Powell, S. L., Lawrence, R. L. ve Hilker, T., 2011, Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery, *Remote Sensing of Environment*, 115, 66–75.

Wu, C., 2011, Travel-time prediction with support vector regression, *IEEE Transactıons on intellıgent transportatıon systems*, 5 (4), 276-281.

Wu, Z., Chui, C.K., Hong, G.S., Chang, S., 2011, Physiological analysis on oscillatory behavior of glucose–insulin regulation by model with delays, *Journal of Theoretical Biology*, 280 (1), 1–9.

Yadav, J., Rani, A., Singh, V., Murari, B.M., 2018, Levenberg-Marquardt-Based Non-Invasive Blood Glucose Measurement System, *IETE Journal of Research*, 64 (1), 116-123.

Yamaguchi, M., Kaseda, C., Yamazaki, K. ve Kobayashi, M., 2005, Prediction of blood glucose level of type 1 diabetics using response surface methodology and data mining, *Medical & biological engineering & computing*, 44 (6), 451–457.

Yazıcı, C. A., Öğüş, E., Ankaralı, S., Canan, S., Ankaralı, H. ve Akkuş, Z., 2007, Yapay sinir ağlarına genel bakış, *Türkiye klinikleri journal of medical sciences,* 27, 65-71.

Yıldırım, M., 2012, *Kandaki glikoz ve HbA1c değerlerinin avuç içi nem parametreleri kullanılarak destek vektör makinesi ile sınıflandırılması*, Yüksek Lisans Tezi, Dumlupınar Üniversitesi Fen Bilimler Enstitüsü.

Yurtoğlu, H., 2005, *Yapay sinir ağları metodolojisi ile öngörü modellemesi: Bazı makroekonomik değişkenler için Türkiye örneği*, Uzmanlık Tezi, Devlet Planlama Teşkilatı-Uzmanlık Tezleri.

Zarkogianni, K., Vazeou, A., Mougiakakou, S.G., Prountzou, A., Nikita, K.S., 2011, An insulin infusion advisory system based on autotuning nonlinear modelpredictive control, *IEEE transaction on biomedical enineering*, Eylül 2011, 58 (9), 2467–2477.

Zarkogianni, K., Litsa, E., Vazeou, A., Nikita, K.S., 2013, Personalized glucose- insulin metabolism model based on self- organizing maps for patients with type 1 diabetes mellitus, *IEEE international conference on bioInformatics and bioEngineering*, 10-13 Kasım 2013, 1–4.

Zarkogianni, K., Mitsis, K., Fioravanti, A., ve Nikita, K.S., 2014, Neuro- fuzzy based glucose prediction model for patients with type 1 diabetes mellitus, *Institute of electrical and electronics engineers*, Haziran 2014, 252–255.

Zarkogianni, K., Litsa, E., Mitsis, K. A., 2015, A Review of emerging technologies for the management of diabetes mellitus, *Institute of electrical and electronics engineers, Trans Biomed Eng*., 19 Ağustos 2015, 2735–2749.

Zarkogianni, K., Mitsis, K., Litsa E., 2015, Comparative assessment of glucose prediction models for patients with type 1 diabetes mellitus applying sensors for glucose and physical activity monitoring, *Medical & biological engineering & computing*, 53 (12), 1333–1343.

Zecchin, C., Facchinetti, A., Sparacino, G., De Nicolao, G., Cobelli C., 2011, A new neural network approach for short- term glucose prediction using continuous glucose monitoring time- series

and meal information, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 5653–5656.

Zecchin, C., Facchinetti, A., Sparacino, G., De Nicolao, G., Cobelli, C., 2012, Neural network incorporating meal information improves accuracy of short- time prediction of glucose concentration, *IEEE Trans Biomed Eng.*, 59 (6), 1550–1560.

Zhang, G., Patuwo, B. E. ve Hu, M. Y.,1998, Forecasting with artificial neural networks: The state of the art., *International journal of forecasting*, 14, 35-62.

Zhang Y, Holt TA, Khovanova N., 2015, A data driven nonlinear stochastic model for blood glucose dynamics. *Computer methods programs in biomedicine*, 125, 18– 25.

Zhao, C., Dassau, E., Jovanovic, L., Zisser, H.C., Doyle, F.J., Seborg, D.E., 2012, Predicting subcutaneous glucose concentration using a latent- variable- based statistical method for type 1 diabetes mellitus, *Journal of diabetes science technology*, 6 (3), 617–633.

Zhao, C., Yu, C., 2015, Rapid model identification for online subcutaneous glucose concentration prediction for new subjects with type I diabetes, *IEEE Transactions on biomedical enginering*, 62 (5), 1333–1344.