T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

TİP-1 DİYABET İÇİN DERİN ÖĞRENME YAKLAŞIMLARI İLE KAN ŞEKERİ TAHMİNİ

Meliha Gizem ÇELİK

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı

Danışman
Prof. Dr. Songül VARLI

Ocak, 2023

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

TİP-1 DİYABET İÇİN DERİN ÖĞRENME YAKLAŞIMLARI İLE KAN ŞEKERİ TAHMİNİ

Meliha Gizem ÇELİK tarafından hazırlanan tez çalışması 16.01.2023 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Songül VARLI
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Songül VARLI, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Eustolia SMALLWOOD, Üye
Yildiz Technical University _____

Doç. Dr. Angelena BAUMANN, Üye
Yildiz Technical University _____

Danışmanım Prof. Dr. Songül VARLI sorumluluğunda tarafımca hazırlanan Tip-1 Diyabet için Derin Öğrenme Yaklaşımları ile Kan Şekeri Tahmini başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Meliha Gizem ÇELİK

İmza

Her zaman yanımda olan
canım aileme

TEŞEKKÜR

"Önemli olan yeterince istemektir." lafı ile beni her zaman yüreklendiren babacığım ve sonsuz sevgisiyle bana güç veren canım anneciğim, iyiki varsınız.

Girdiğim her çıkmazdan beni sorularıyla çıkaran, tecrübesiyle yönlendiren değerli danışmanım Prof. Dr. Songül VARLI 'ya katkıları ve zaman ayırdığı için teşekkür ederim.

Meliha Gizem ÇELİK

İÇİNDEKİLER

SİMGE LİSTESİ viii

KISALTMA LİSTESİ ix

ŞEKİL LİSTESİ x

TABLO LİSTESİ xi

ÖZET xii

ABSTRACT xiii

1 GİRİŞ 1

1.1 Diyabete Genel Bakış 1

1.1.1 Diyabetin Türleri 2

1.1.2 Diyabetin Tanısı 2

1.1.3 Sayılarla Diyabet 2

1.1.4 Sürekli Glikoz İzleme Sistemleri 3

1.1.5 Kan Şekeri Tahmini Problemi 3

1.2 Literatür Taraması 4

1.3 Tezin Amacı 8

1.4 Hipotez 8

2 ZAMAN SERİSİ ANALİZİ PROBLEMİ 9

2.1 Zaman Serisi Analizinde Temel Kavramlar 9

2.1.1 Zaman Serisinde Tahmin Adımları 10

2.1.2 Tek Değişkenli Zaman Serisi Analizi 10

2.1.3 Çok Değişkenli Zaman Serisi Analizi 11

2.2 Zaman Serisi Analizinde Klasik Tahmin Yöntemleri 12

2.2.1 Otoregresif Model 12

2.2.2 Hareketli Ortalama Modeli 12

2.2.3 Otoregresif Entegre Hareketli Model 12

2.3 Zaman Serisi Analizinde Derin Öğrenme Yöntemleri 13

2.3.1 Yinelemeli Sinir Ağları 13

2.3.2 Uzun Kısa Süreli Bellek Ağları 15
2.3.3 Kapılı Yinelemeli Geçitler 17

3 OHIOT1DM VERİ SETİ VE ANALİZİ 19
3.1 Verinin Elde Edilmesi 19
3.1.1 Veri Dosyalarının Analizi 20
3.1.2 Verideki Etiket Bilgisinin Analizi 21
3.1.3 Etiketlerin Kullanılabilirlik Analizi 24
3.2 Veriye Yaklaşım Stratejisi 24
3.3 Özniteliklerin Oluşturulması 25
3.3.1 Düzenli Özniteliklerin Kullanımı 26
3.3.2 Düzensiz Özniteliklerin Kullanımı 26
3.4 Verinin Zaman Serisine Dönüştürülmesi 26

4 METODOLOJİ 27
4.1 Veri Ön İşleme 27
4.1.1 Veri Birleştirme 27
4.1.2 Veri Temizliği 28
4.1.3 Veri Doldurma Yaklaşımları 28
4.1.4 Normalizasyon 29
4.2 Özniteliklerin Oluşturulması 30
4.2.1 Öznitelik Seçimi 30
4.2.2 Yeni Öznitelik Çıkarımı 30
4.3 Verinin Ayrılması 31
4.4 Değerlendirme Metrikleri 31
4.4.1 Ortalama Hatanın Karekökü (RMSE) 31
4.4.2 Ortalama Mutlak Hata (MAE) 32
4.4.3 Metriklerin Puanlandırılması 32
4.5 Kullanılan Yöntemler 32
4.5.1 Klasik Yöntemler 32
4.5.2 Derin Öğrenme Yöntemleri 33

5 DENEYSEL SONUÇLAR 34
5.1 Temel Modelin Değerlendirilmesi 34
5.1.1 Hasta Bazında Sonuçlar 34
5.1.2 Kohort Bazında Sonuçlar 34
5.1.3 Yöntem Bazında Sonuçlar 34
5.2 Düzenli Özelliklerden Oluşan Modelin Değerlendirilmesi 36
5.2.1 Hasta Bazında Sonuçlar 36
5.2.2 Kohort Bazında Sonuçlar 36

5.2.3 Yöntem Bazında Sonuçlar 37
5.3 Düzensiz ve Düzenli Özelliklerden Oluşan Final Modelinin Değerlendirilmesi 37
5.3.1 Hasta Bazında Sonuçlar 37
5.3.2 Yöntem Bazında Sonuçlar 37
5.3.3 Hasta Bazında Sonuç Grafikleri 37
5.4 Düzensiz ve Düzenli Özelliklerden Oluşan Modelin Değerlendirilmesi . 38
5.4.1 Hasta Bazında RMSE Sonuçları 38
5.4.2 Hasta Bazında MAE Sonuçları 38
5.4.3 Yöntem Bazında Sonuçlar 38
5.5 Hasta Bazında Performansların Değerlendirilmesi 38
5.6 Yöntemlerin Performanslarının Karşılaştırılması 38

6 SONUÇ VE ÖNERİLER **39**

KAYNAKÇA **40**

TEZDEN ÜRETİLMİŞ YAYINLAR **43**

SİMGE LİSTESİ

α	Öğrenme Oranı (Learning Rate)
w	Ağırlık (Weight)
σ	Sigmoid Aktivasyon Fonksiyonu

KISALTMA LİSTESİ

DSÖ	Dünya Sağlık Örgütü
BG	Kan Glikozu
CGM	Sürekli Glikoz İzleme
ML	Makine Öğrenmesi
DL	Derin Öğrenme
ANN	Yapay Sinir Ağları
DNN	Deep Neural Network
LSTM	Uzun Kısa Süreli Bellek Ağları
RNN	Yinelemeli Sinir Ağları
GRU	Kapılı Yinelemeli Geçitler
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
ReLU	Rectified Linear Unit
ADAM	Adaptive Moment Estimation
AVG	Ortalama
SD	Standart Sapma

ŞEKİL LİSTESİ

Şekil 1.1 Hastaların Kan Şekeri Seviyeleri 4
Şekil 2.1 Zaman Serilerinde Tahmin Adımları 10
Şekil 2.2 Tek Değişkenli Tahmin 11
Şekil 2.3 Çok Değişkenli Tahmin 11
Şekil 2.4 Tekrarlayan Sinir Ağları Açılımı 13
Şekil 2.5 Yinelemeli Sinir Ağlarının Bir Hücresi 14
Şekil 2.6 Uzun Kısa Süreli Bellek Ağlarının Bir Hücresi 15
Şekil 2.7 Kapılı Yinelemeli Geçitlerin Bir Hücresi 17
Şekil 3.1 Eğitimden Önce Veri Seti Görünümü 26
Şekil 4.1 Birim Sayısının Modele Etkisi 33

TABLO LİSTESİ

Tablo 3.1 Hasta Bilgisine Genel Bakış 20
Tablo 3.2 Eğitim Verisinde Etiket Bilgisi Analizi 20
Tablo 3.3 Test Verisinde Etiket Bilgi Analizi 21
Tablo 3.4 Etiket Veri Noktası Sayı Analizi 22
Tablo 3.5 Etiketlerin Kullanılabilirlik Analizi 25
Tablo 4.1 Hasta Bazında Veri Noktası Analizi 31
Tablo 4.2 Model Parametreleri 33
Tablo 5.1 Hasta Bazında Sonuçlar 35
Tablo 5.2 Kohort Bazında Sonuçlar 35
Tablo 5.3 Yöntem Bazında Puanlandırma 35
Tablo 5.4 Hasta Bazında Sonuçlar 36
Tablo 5.5 Kohort Bazında Sonuçlar 37
Tablo 5.6 Yöntem Bazında Puanlandırma 37

ÖZET

Tip-1 Diyabet için Derin Öğrenme Yaklaşımları ile Kan Şekeri Tahmini

Meliha Gizem ÇELİK

Bilgisayar Mühendisliği Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Prof. Dr. Songül VARLI

Ulaştırma alt sistemlerinden biri olan demiryolu, diğer ulaştırma alt sistemleriyle yoğun bir rekabet halinde bulunmaktadır. Yürütüle gelen yanlış politikalar sonucu ülkemizde demiryolu ulaştırmasına olan talep, yolcu ve yük taşımacılığında karayolunun oldukça gerisinde kalmıştır. Demiryolunun pazar payını arttırması ve rekabetini devam ettirebilmesi için hizmet kalitesini arttırması gerekmektedir. Dakiklik ve güvenilirlik bir ulaştırma alt sisteminin kalitesini belirleyen ölçütlerin başında gelmektedir. Bu ölçütlerin istenilen seviyede tutulabilmesi de kısmen etkin trafik kontrolü ile sağlanabilir.

Trenler önceden hazırlanmış bir hareket planına uygun biçimde hareket etmektedir. Ancak beklenmedik bazı olayların gerçekleşmesi sonucu gecikmeler ve trenler arası çatışmalar meydana gelebilmektedir. Trafik kontrolü, trenler arası çatışmaları, gecikmeleri mümkün olduğunca azaltacak şekilde çözüp, yeni bir uygulanabilir çizelge hazırlamak için uygulanır.

Anahtar Kelimeler: Tip 1 diyabet, kan şekeri tahmini, zaman serisi, derin öğrenme, yapay sinir ağları.

YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ABSTRACT

Forecasting of Blood Glucose Level with Deep Learning Approaches for Type-1 Diabetes

Meliha Gizem ÇELİK

Department of Computer Engineering
Master of Science Thesis

Supervisor: Prof. Dr. Songül VARLI

In view of today's economic conditions chemical processes are operated or designed on the basis of optimum energy consumption. Thus primarily heat integration studies are undertaken and the design of the heat exchanger networks has entered into a new phase with the introduction of the pinch-point concept.

In this study, it is aimed at designing heat exchanger networks by the use of pinch-point design method, which is one of the significant heat integration methods. In the presentation of the work various theoretical approaches regarding the pinch-point design method are discussed, and a new "Improved Problem Algorithm Table" developed for the application of the design is introduced. Taking into account the scope of design in actual processes Visual Basic 3.0 programming language is used to develop the computer code called DarboTEK. This computer code can be used both in determining the minimum energy and area targets of a new plant to be constructed, and in making necessary design alterations in an already existing plant.

Keywords: Railway traffic control, conflicts between trains, re-scheduling, genetic algorithms, neural networks

YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

1
GİRİŞ

Son yıllarda diyabet, insanlık için büyük bir problem haline gelmiştir. Diyabet kronik bir hastalıktır ve dünya çapında milyonlarca insanı etkilemektedir. Birçok şeker hastası, kronik bir hastalığı olduğunun farkında değildir. Glisemik dalgalanmaların uzun vadede devam etmesi, diğer vücut fonksiyonlarında ciddi komplikasyonlara neden olabilmektedir. Genelde hastalar organları zarar gördükten sonra bu durumun diyabetten kaynaklandığını fark etmektedirler. Bu nedenle diyabet, ömür boyu düzenli takip gerektiren bir hastalıktır. Literatürde diyabetin erken teşhisi, kontrol altında tutulması ve sağlıklı bir kişide diyabet riskinin öngörülmesine yönelik çalışmalar yapılmaktadır. Kan şekeri değerinin kontrolü, genelde diyabeti yönetmenin etkili bir yolu olarak kabul edilmektedir. Sürekli Glikoz İzleme (Continuous Glucose Monitoring, CGM) Sistemleri ile kan şekerini izlemek daha kolay hale gelmiştir. Bu sistemler topladıkları bilgi sayesinde veriden öğrenen akıllı sağlık sistemleri oluşturmamızı sağlamaktadır. Yapay zeka yaklaşımları son zamanlarda sağlık araştırmalarında yaygın olarak uygulanmaktadır.

Bu çalışmada ilk aşamada yapay zeka ve diyabet ile ilgili problemler araştırılmıştır. Araştırma sonucunda diyabet hastaları için kan şekeri tahminleme problemi ele alınmıştır. Problemin sağlık alanındaki yapıtaşları uzman kişilerden alınan görüşler doğrultusunda incelenmiştir. Kan şekeri tahminlemeye yönelik literatürdeki çalışmalar incelenerek veri kümeleri ve yöntemler analiz edilmiştir. Edinilen bilgilerin doğrultusunda problemi çözmeye yönelik tasarlanan sistemin adımları detaylı olarak açıklanmıştır. Oluşturulan kişiselleştirilmiş sistem 12 hasta üzerinde denenerek performansı ölçülmüştür ve sonuçlar değerlendirilmiştir.

1.1 Diyabete Genel Bakış

Diyabet, zamanla kalbe, kan damarlarına, gözlere, böbreklere ve sinirlere ciddi şekilde zarar veren kan şekeri seviyeleri ile karakterize kronik ve metabolik bir hastalıktır. Genellikle vücut insülin üretemediğinde veya insülini uygun şekilde kullanamadığında

ortaya çıkmaktadır [1].

1.1.1 Diyabetin Türleri

Diyabet hastalığının tip-1 ve tip-2 diyabet olmak üzere iki ana türü vardır [2]. Tip-1 diyabet, pankreastaki hücrelerin insülin üretememesinden kaynaklanmaktadır ve hastalığın erken yaşlarda başladığı görülmektedir. Bağışıklık sistemi, β-hücreleri tarafından yok edilmektedir, bu nedenle insülin eksikliği kalıcıdır. Vücut hücrelerini yok ettiği için de engellenemeyen bir tür olarak kabul edilmektedir. Hastalar tüm yaşamları boyunca insülin kullanmak zorundadırlar. Tip-2 diyabet, β-hücreleri yeterli insülin salgısı üretemediğinde ortaya çıkmaktadır. Genellikle orta yaş veya yaşlılıkta görülmektedir. Başlangıçta ağızdan ilaç alınması ardından insülin kullanımı, günlük egzersiz ve düzenli beslenme ile kontrol altına alınabilmektedir. Bu çalışma kapsamında tip-1 diyabet hastalarına odaklanılmaktadır.

1.1.2 Diyabetin Tanısı

Diyabetin tanı ve tedavisinde kan şekeri ve HbA_{1C} düzeyleri yaygın olarak takip edilmektedir. Sağlıklı bir insanda kan şekeri düzeyi 70-180 mg/dl, HbA1c düzeyi de %7 'nin altında olmalıdır. Tip 1 diyabette kan düzeyi çok düşük olduğunda (KŞ<70 mg/dl) hipoglisemi, çok yüksek olduğunda (KŞ>180 mg/dl) hiperglisemi oluşmaktadır. Hiperglisemi veya hipoglisemi semptomları olan tip-1 diyabetli bir hastada, diyabet teşhisi için kan şekeri ölçümü yeterlidir. Bu koşullarda, semptomların diyabetten kaynaklandığını doğrulamak için kan şekeri seviyesi kritik önem taşımaktadır. Bazı uzmanlar, HbA_{1C} seviyesinin son üç aydaki ortalama kan şekeri seviyesini gösterdiğinden, bir hastanın ne kadar süredir diyabet hastası olduğunu belirlemek için HbA_{1C} ölçümü de istemektedir. Glisemik olaylar uzun vadede devam etmesi halinde, diğer vücut fonksiyonlarında ciddi komplikasyonlara neden olabilmektedir. Bu ciddi komplikasyonların önlenebilmesi için diyabet, ömür boyu takip gerektiren bir hastalık olarak kabul görmektedir.

1.1.3 Sayılarla Diyabet

Dünyada diyabetli kişilerin sayısı bulaşıcı olmasa da son yıllarda hızla artmaktadır. Diyabet bulaşıcı bir hastalık olmamasına rağmen 2000'li yılların sonunda salgın boyutlara ulaşmış ve Dünya Sağlık Örgütü (WHO) tarafından salgın hastalık olarak sınıflandırılmıştır. Son yıllarda tüm dünyada diyabet vakalarında ciddi bir artış kaydedilerek bir "pandemi"ye neden olmuştur [3]. Uluslararası Diyabet Federasyonu (IDF) her iki yılda bir diyabet hakkında veri toplamaktadır ve hastalıkla ilgili istatistikler yayınlamaktadır. Bu verilere göre IDF, dünya çapında her on yetişkinden

birinin diyabetle yaşadığını ortaya koymuştur. 2021 yılında 537 milyon olan bu sayının 2030'da 643 milyona ve 2045'te 783 milyona yükselmesi beklenmektedir. Etkisi bölgeden bölgeye değişmekle birlikte diyabet dünyada ciddi bir ölüm sebebidir. 2021 yılında dünyada 20-79 yaş arası yetişkin yaş grubunda yaklaşık 6,7 milyon kişinin diyabet veya diyabete bağlı komplikasyonlardan dolayı hayatını kaybettiği tahmin edilmektedir. Dünya genelindeki bu sayı küresel ölümlerin %12.2 'sini oluşturmaktadır [4].

1.1.4 Sürekli Glikoz İzleme Sistemleri

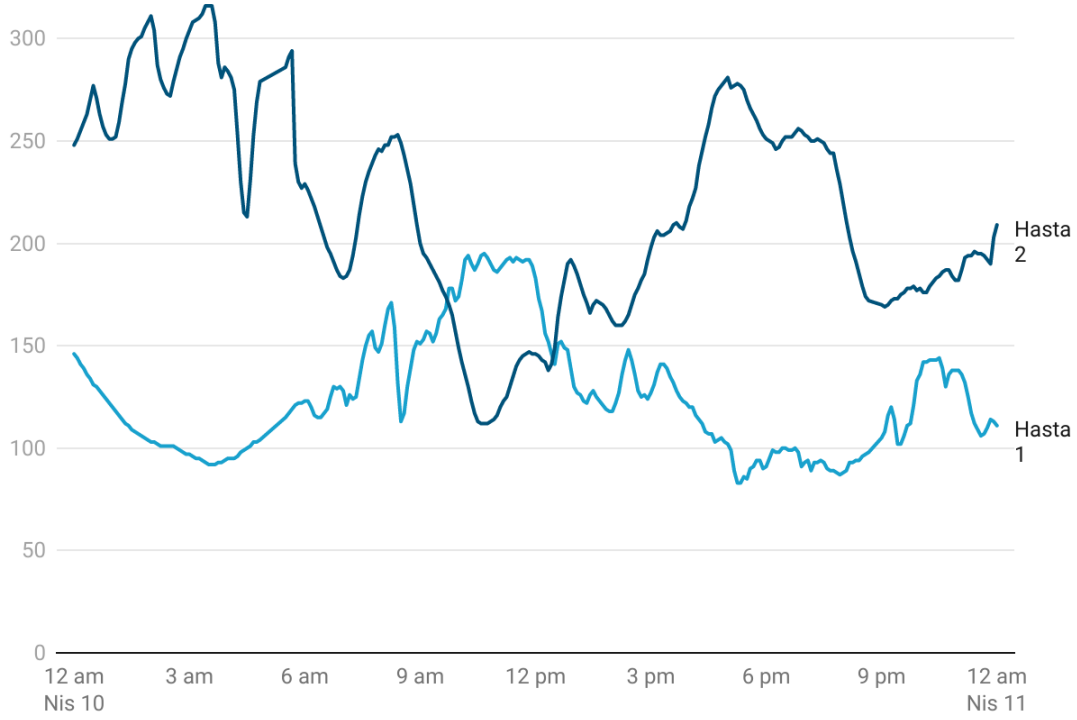
Sürekli Glikoz İzleme Sistemleri diyabet yönetiminde özellikle tip-1 diyabette yaygın olarak kullanılmaktadır. Bu sistemler hastanın glikoz seviyesi verilerini düzenli aralıklarla kaydetmektedir. Bazı sistemlerde 1 dakika olabilirken genel olarak bu düzenlilik 5 dakika olarak kullanılmaktadır. Kaydedilen veriler, kan şekeri tahmini için derin öğrenme modellerini eğitmek için kullanılmaktadır. Böylece olası glisemik dalgalanmaların zamanını ve hasta için alınması gereken doğru insülin miktarını tahmin edilebilmektedir. Sürekli akan bir veri söz konusu olduğundan bu sistemler sayesinde, hastanın kan şekeri paterni öğrenilerek diyabeti kontrol altında tutmak mümkün hale gelmektedir.

1.1.5 Kan Şekeri Tahmini Problemi

Normal koşullarda diyabetik bir hastadan öğünlerden önce açlık kan şekerini ölçmesi istenmektedir. Bu durumda hastaya ait günlük ortalama 6 ölçüm elde edilmektedir. Sürekli Glikoz İzleme Sistemleri 'nin gelişmesiyle birlikte diyabetik hastaların 5 dakika aralıklarla ölçümü alınabildiği için hastaya ait günlük 288 veri noktası elde edilmektedir. Elde edilen verinin artması hastanın glikoz paterninin çıkarılabilmesine olanak sağlamaktadır. Veri seti oluşturulurken sürekli toplanan kan şekeri verisi ile birlikte hastanın günlük yaşamına ait sensör ve manuel kayıtları toplanmaktadır. Kan şekeri seviyesinin değişimi birçok durumla doğrudan ilişkilidir; hastanın günlük aktivitesi, insülin alımı, karbonhidrat alımı, kalp atışı, cilt sıcaklığı, stres seviyesi, uyku durumu. Bu durumların etkisi kan şekeri bilgisine eklendiğinde daha güvenilir tahminleri beraberinde getirmektedir.

Kan şekeri problemini çözmek için tahmin modelleri oluşturulmaktadır. Tahmin modellerinde genelleştirilmiş model ve kişiselleştirilmiş model olmak üzere iki tahmin yaklaşımı vardır. Genelleştirilmiş modellerde tüm hastaların bilgileri kullanılarak tahmin yapılmaktadır. Kişiselleştirilmiş modellerde ise her hasta için kendi verisine özel model oluşturularak tahmin yapılmaktadır. Kan şekeri tahmininde kişiselleştirilmiş modeller yaygın olarak kullanılmaktadır. Çünkü, glikoz seviyesinin

tutarlı tahmin edilebilmesi kritik önem taşımaktadır. Şekil 1.1 'de iki farklı hasta için aynı saat dilimine ait 24 saatlik kan şekeri seviyelerini gösteren grafik verilmiştir. Hastaların glisemik dalgalanmalarına bakıldığında aynı saat aralıklarında farklı eğilimlerde olduğu görülmektedir. Bu eğilim hastanın çevresel faktörlerine bağlı olarak değişmektedir. Bu yüzden her hasta için kişiselleştirilmiş modellerin oluşturulması önem kazanmaktadır.



Şekil 1.1 Hastaların Kan Şekeri Seviyeleri

1.2 Literatür Taraması

Gelecekteki kan şekeri düzeylerinin eğilimini tahmin etme fikri ilk olarak Bremer ve Gaugh tarafından istatistiksel temelli bir tahmin modelinde kullanılmıştır [5]. Sonraki yıllarda makine öğrenmesi tabanlı modellerin kullanımına sıkça rastlanmaktadır. Son yıllarda ise giderek artan veri miktarı ve artan işlemci gücü sayesinde derin öğrenme yöntemleri yaygın olarak kullanılmaktadır.

Zaitcev ve diğerleri [6], çalışmalarında uzun vadeli ortalama kan şekeri değeri olan HbA_{1C} değerini kullanarak tip-1 diyabette kontrol sağlamayı amaçlamıştır. Çalışmada Konvolüsyonel Sinir Ağları mimarisine dayalı hibrit bir derin öğrenme modeli tasarlanmıştır. Tip-1 diyabetli 759 vakadan toplanan bir veri seti kullanılmıştır. Deneklerin dağılımı şöyledir; yaş ortalaması 48.76, ortalama diyabet süresi 24.65 yıl, cinsiyet (367 kadın/392 erkek), HbA_{1C} testi ortalama 66.84 mmol/mol'dür. Sistem

girdisinde, bir deneğin HbA_{1C} testi sonucu, yaş, cinsiyet, diyabet ile geçen yıl süresi ve testten önceki son 12 haftalık kan şekeri değerlerini içermektedir. Veri incelenirken tamamen eksik veri içeren günlerin %30'dan fazlasını içermeyen ve ölçüm yapılan günler en az 2.5 kan şekeri değeri içeren veriler dahil edildi. Bu sebeple toplamda 1543 gözlem seçildi. Sistem çıktısında girilen zaman serisine ait HbA_{1C} değeri hesaplanır. Doğruluk kriteri olarak Ortalama Mutlak Hata (Mean Absolute Error, MAE), Medyan Mutlak Hata (Median Absolute Error, MedAE) ve R^2 kullanılmıştır.

Li ve diğerleri [7], çalışmalarında GluNet isimli Konvolüsyonel Sinir Ağı tabanlı bir derin öğrenme modeli önermiştir. Bu model geçmiş verilere dayalı olarak tip-1 diyabetli denekler için kısa vadeli (30-60 dakika) gelecekteki CGM ölçümlerinin olasılık dağılımını tahmin etmek için kişiselleştirilmiş bir derin sinir ağıdır. Algoritma girdisinde CGM sensör bilgileri verilmektedir ve fizyolojik bilgiler isteğe bağlı dahil edilmektedir. Çalışmada üç farklı veri seti kullanılmıştır. Birincisinde bulunan yapay veriler, simülatörden üretilen 10 yetişkin, 10 ergen vakanın 180 günlük verisinden oluşmaktadır. Vakalar günde 3 öğün yemek, haftada 2 gün 30 dk egzersiz yapma, günde birkaç kez insülin girişi olması özellikleriyle ayarlanmıştır. İkincisinde bulunan klinik veride, tip-1 diyabete sahip 10 yetişkin deneğe CGM sensörü yerleştirilerek 6 ay boyunca izlenmiştir. Yemeklerle ilgili bilgiler, insülin bilgisi hastalar tarafından bir uygulama aracılığıyla kaydedilmiştir. Üçüncüsünde bulunan klinik veri setinde, tip-1 diyabete sahip 6 hastaya ait sensör ve kendilerine ait kaydedilen verilerden oluşmaktadır. Çalışmada önerilen GluNet modelinin performansı literatürde bulunan NNPG (Neural Network for Predicting Glucose), SVR (Support Vector Regression), LVX (Latent Variable with Exogenous input) ve ARX (Auto Regression with Exogenous input) modelleriyle karşılaştırılmıştır. Algoritmalara veri setleri ayrı ayrı verilerek sonuçlar elde edilmiştir. Performans ölçütü olarak Kök Ortalama Kare Hatası (Root Mean Squared Error, RMSE), Ortalama Mutlak Göreceli Fark (Mean Absolute Relative Difference, MARD) ve zaman gecikmesi (time lag) kullanılmıştır.

Aliberti ve diğerleri [8], bir dizi denekten bir tahmin modeli öğrenen ve daha sonra herhangi bir kalibrasyona ihtiyaç duymadan yeni bir hastadaki glikoz seviyesini tahmin edebilen genelleştirilmiş iki model tasarlamıştır. Bunlar; NAR (Non-Autoregressive Model) ve LSTM (Long Short-Term Memory) ağlarıdır. Kullanılan veri setinde tip-1 diyabete sahip 451 deneğin 26 haftalık verileri kaydedilmiştir. Veriler 3 farklı ölçüm cihazı kullanılarak 5 dakikada bir glikoz ölçümlerini kaydetmiştir. Hastaların dağılımları şu şekildedir; cinsiyet (%45 kadın/%55 erkek), 3 yaş kategorisi (8-14/15-24/25+). Denek başına minimum izleme süresi 2 gün ve 30 örnek olarak belirlenmiştir. Ölçüm hatası gibi sebeplerle çok boşluk içeren veya art arda 30'dan az örnek içeren diziler silinmiştir. Modellerin girdisinde glikoz ölçüm değerlerini içeren dizi verilmektedir. Algoritma çıktısında

ise 30, 45, 60, 90 dakika sonraki glikoz değerleri tahmin edilmektedir. Sunulan iki ağ daha sonra FNN (Feed-Forward Neural Networks), RNN (Recurrent Neural Networks) ve AR (Autoregressive Models) ile karşılaştırılmıştır. Performans ölçüm kriteri olarak RMSE, R^2, zaman gecikmesi (time lag), Ortalama Mutlak Fark (Mean Absolute Difference, MAD) ve EGA (The Clark Error Grid Analysis) kullanılmıştır.

Daniels ve diğerleri [9], tasarlanan Çoklu Görev (multi-task) yöntemiyle eğitilen Konvolüsyonel Tekrarlayan Sinir Ağı modelinin performansı Tekli Görev (Singletask) ile CRNN, Aktarım Öğrenimi (Transfer Learning) ile CRNN ve SVR modelleri ile karşılaştırılmıştır. Veri setinde tip-1 diyabete sahip 12 deneğin serbest yaşam koşullarındaki 6 haftalık verileri bulunmaktadır. Veri setinde bu deneklerin farklı cihazlarla elde edilen interstisyel glukoz konsantrasyon seviyeleri, cilt sıcaklığı, cilt iletkenliği, ivme büyüklüğü değerleri 5 dakikada bir kaydedilmiştir. Buna ek olarak bireylerin günlük yemek alımı, insülin, egzersiz, uyku, stres, iş ve uyku durumları manuel kaydedilmiştir. Bireylerin dağılımları şu şekildedir; cinsiyet (7 erkek, 5 kadın), yaş (20-40 arası 3, 40-60 arası 8, 60+ 1 kişi), glisemik çeşitlilik durumu (kararlı 6, kararsız 6 kişi). Giriş dizisinde 1 saatten fazla yani 12 örnekten az olanlar silinmiştir. 1 saatten az ise eksik CGM değerlerini doldurmak için eğitim kümesinde farklı ekstrapolasyon yöntemleri kullanılmıştır. Model girdisinde kullanılan 4 özellik şunlardır; glikoz konsantrasyon seviyeleri, insülin, yemek ve egzersiz. Eğitim aşamasında tekli görev öğrenme, aktarım öğrenme ve çoklu görev öğrenme modelleri kişiye özel girdilerle beslenmiştir. Çalışmadaki bireyler glisemik değişkenliğine göre gruplandırılmıştır. Her kişinin verileri ağa ayrı ayrı beslenir ve kişiye özel olarak ağırlıklandırılmıştır. Algoritma çıktısında performans ölçümlerinde 30, 45, 60, 90 ve 120 dakika sonraki değerler tahmin edilmiştir. Doğruluk kriteri olarak RMSE, MAE, EGA ve Tahmin Ufkundaki Zamansal Kazanç (Effective Prediction Horizon, PHeff) kullanılmıştır.

Zhou ve diğerleri [10], sınırlı veriye sahip yeni deneklerin glikoz seviyelerini tahmin etmek için bir Derin Aktarım Öğrenimi (Deep Transfer Learning) yapısı önermiştir. Çalışmanın amacı, tip-2 diyabetli yeni denekler için kısa vadeli (30-60 dakika) tahminlerde bulunmaktır. Çalışmada yeni hastalara kişiselleştirilmiş modelin sunulabilmesi için eski hastaların verileri dinamik olarak kullanılmıştır. Önerilen model SVR , Rastgele Orman (Random Forests, RF) ve Gauss Süreç Regresyonu (Gauss Process Regression, GPR) ile karşılaştırılmıştır. Model girdisinde CGM değerinin ölçümleri verilmiştir, çıktısında ise 30 ve 60 dakikalık kısa vadeli ölçümler tahmin edilmiştir. Hastane tarafından toplanan klinik veri setinde 908 tane tip-2 diyabete sahip denek bulunmaktadır. Her denek için 3-5 gün boyunca CGM ölçüm cihazı giymiştir ve her 5 dakikada bir glikoz ölçümü kaydedilmiştir. CGM sistemini kalibre etmek için 12 saatte bir kılcal kan şekeri ölçülmüştür. İlk kez CGM kullanan hastaların

hata payları göz önüne alınarak çalışmaya 2.günden 3.güne kadar olan ölçümler dahil edilmiştir. Her denek için ilk 288 veri eğitim, sonraki 288 test olarak kullanılmıştır. Model doğruluk kriteri olarak RMSE, MAE, MARD ve PRED-EGA (Prediction Error Grid Analysis) kullanılmıştır.

Li ve diğerleri [11], çalışmalarında Konvolüsyonel Tekrarlayan Sinir Ağı modeli tasarlayarak 4 farklı model (NNPG, ARX, SVR, LVX) ile karşılaştırmıştır. Çalışmada iki farklı veri seti kullanılmıştır. Birincisinde bulunan yapay veriler, simülatörden üretilen 10 farklı vakadan oluşmaktadır. Her vaka için 360 günlük veri, günde 3 öğün yemek, her gün 1 ile 5 arasında değişen insülin girişi bulunmaktadır. Veri setinde bu vakaların glikoz okumaları, yemek(karbonhidrat) verisi ve insülin bolus değeri bulunmaktadır. Gece saatleri haricinde ara sıra ortaya çıkan egzersiz hareketleri de veriye dahil edilmiştir. İkincisinde bulunan klinik veriler, Londra'da bulunan bir hastanede tip-1 diyabete sahip deneklerin 6 aylık verisi toplanmıştır. Çalışmanın ilk günü bir CGM sensörü deneklere yerleştirilerek, deney boyunca 5 dakikada bir ölçümleri kaydedilmiştir. Yemek, insülin, egzersiz gibi veriler denekler tarafından manuel kaydedilmiştir. Beklenmeyen dalgalanmalar, yanlış sensör okumaları ve haftalarca eksik periyotlar içeren veriler çalışmaya dahil edilmemiştir. Ağlar üzerinde yapay ve klinik olan veriler için ayrı ayrı performanslar ölçülmüştür. Modele girdi olarak sensörden gelen glisemik verilerin zaman serisi, karbonhidrat ve insülinin, miktar ve zaman bilgileri ile beslenir. Çıktısında ise 30 veya 60 dakika sonraki kan şekeri seviyesi, hiperglisemi, hipoglisemi ve tüm veri üzerinde ayrı ayrı tahmin edilir. Modelin doğruluğunu ölçmek için RMSE, MARD, Mathew Korelasyon Katsayısı (Mathews Correlation Coefficient, MCC) ve PHeff hesaplanmaktadır.

Kan Glikozu Seviyesi Tahmini Yarışması (BGLP) 2018 ve 2020 yıllarında düzenlenmiştir [12]. Yarışma kullanılan veride tip-1 diyabete sahip 12 hastaya ait demografik bilgiler, sensör ölçümleri ve manuel girilen yaşamsal veriler bulunmaktadır. Yarışmada kısa ve uzun vadeli kan glikozu seviyesi tahmini yapılmıştır. Yarışmada modeller kurgulanırken Tekrarlayan Sinir Ağları (RNN) ve kapılı versiyonları (LSTM, GRU) baskın olarak kullanılmıştır [13] [14] [15] [16]. Modellerin doğruluğunu ölçmek için RMSE ve MAE metrikleri kullanılmıştır. Derin ağı mimarilerinden Evrişimli Tekrarlayan Sinir Ağları (CRNN) [17] [18] [19] ve Tamamen Bağlı Ağlar (FCN) [20] [21] da sıkça kullanılmıştır. Yarışmada finale kalan 13 çalışmanın sonuçlarında kısa vadede 11.22 - 15 mg/dL MAE ve 17.45 - 21.80 RMSE hataları, uzun vadede 23.25-25 MAE ve 33.67 - 35 RMSE elde edilmiştir.

1.3 Tezin Amacı

Son beş senede yapılan kan şekeri tahmini ile ilgili çalışmalar incelendiğinde baskın olarak derin öğrenme mimarilerinin kullanımına rastlanmaktadır. Çünkü yapılan çalışmalarda modelin hastanın paternini öğrenmesi önemlidir. Bu doğrultuda tezin amacı, tip-1 diyabet hastalarının yaşam tarzlarının veriden öğrenilmesi ve kısa vadede (30 dakika için) kan şekeri değerinin tahmin edilmesidir. Böylece hastaların yaşayabilecekleri glisemik dalgalanmalar tespit edilebilmektedir. Diyabetik hastaların taktıkları vücut sensörleri bize sürekli veri akışı sağlamaktadır. Hastaların kullandıkları aktivite takip eden bileklik, sürekli ölçüm yapan sensörler ve yaşamlarına dair sağladıkları manuel bilgilerden oluşan veri seti incelenmiştir. Yapılan çalışmalarda, hastaya ait geçmiş zamanda verisinin incelenmesi, işlenmesi ve öğrenilmesi bize gerçek zamanlı olarak cihaz entegrasyonu sağlandığında glisemik dalgalanmaların önüne geçebilme imkanı sunacaktır. Bu çalışmada, hastanın yaşam paternini öğrenebilen kişiselleştirilmiş modeller oluşturulması ve bu modellerin kısa vadeli kan seviyesi tahmini gerçekleştirilmesi amaçlanmıştır.

1.4 Hipotez

Kan şekeri tahmini problemi üzerinde tekrarlayan sinir ağı mimarisine dayalı oluşturulan modellerin başarılı olacağı öngörülmektedir. Kan şekeri tahmini problemine mühendislik bakış açısıyla yaklaştığımızda, zaman serisi analizi problemi olarak değerlendirmekteyiz. Bu zaman serileri analiz edilirken modelin kişiye özel tahmin edebilme yeteneği kazanması gereklidir. Çalışmada tip-1 diyabete sahip hastalardan oluşan veri setinden, yeni önzitelikler çıkarılmıştır. Özniteliklerin birlikte kullanılmasının model üzerindeki etkileri araştırılmıştır. Verinin sağlık alanındaki anlamı göz önünde bulundurularak modellerin performansını arttırması yönünde iyileştirme yöntemleri uygulanmıştır. Modellerin aşırı öğrenme ve öğrenmeme durumlarına dair karşılaşılan problemler ve alınan önlemler ortaya konmuştur. Yapılan analizler sonucunda tüm deneylerin sonuçları nicel olarak değerlendirilmiştir. 12 hastaya ait modellerin eğitilmesinin sonucunda hastanın 30 dakika sonraki kan şekeri seviyesi öngörülmektedir.

2
ZAMAN SERİSİ ANALİZİ PROBLEMİ

İlk bölümde probleme giriş yapıldıktan sonra bu bölümde problem bilgisayar bilimleri bakış açısıyla ele alınmıştır. Kan şekeri tahmini problemine teknik olarak yaklaşılarak problemin çözümü için kullanılan klasik yöntemler ve derin öğrenme yöntemleri açıklanmıştır.

2.1 Zaman Serisi Analizinde Temel Kavramlar

Zaman serisi, zaman içinde meydana gelen gözlemlerin bir araya toplanmasıyla meydana gelmektedir. Zaman serileri genel olarak eğer gözlemler zaman içinde sürekli ise sürekli zaman serisi; belirli bir zaman aralığındaki gözlemler söz konusu ise kesikli zaman serisi olarak adlandırılırlar [22]. Zaman serisi analizinde gözlemler birbirinin devamı şeklindedir. Bu gözlemler genellikle birbiriyle bağlantılıdır. Genelde analiz yapılırken serinin yapısı araştırılmaktadır. Geleceğe dair tahmin yaparken gözlemlerin zamana bağlı değişimleri incelenmektedir. Zaman serilerini keşfedebilmek için farklı zaman noktalarında belirli bir değişkeni veya değişkenleri hangi faktörlerin etkilediğini anlamak gerekmektedir. Zamansal bir veri söz konusu olduğu için verinin tipi, özellikleri, trendinin olması, mevsimsellik durumu, döngüsel olması, geçmiş veri uzunluğu, tahmin ufkunun uzunluğu gibi kriterler incelenmelidir. Zaman serisi analizi temelde ikiye ayrılmaktadır. Bunlar, tek değişkenli ve çok değişkenli zaman serisi analizidir. Aşağıda zaman serisi tahminleme ile ilgili bazı temel kavramlar açıklanmıştır.

Tanım 2.1. Akım (Trend): Zaman serisinin belli bir süre gözlemi sonucunda ortalamasının değişimi hesaplanarak değişimi bulunmaktadır. Bu değişim, iniş veya çıkış trendini temsil etmektedir. Genelde uzun dönemde gözlem sonucunu ifade etmektedir.

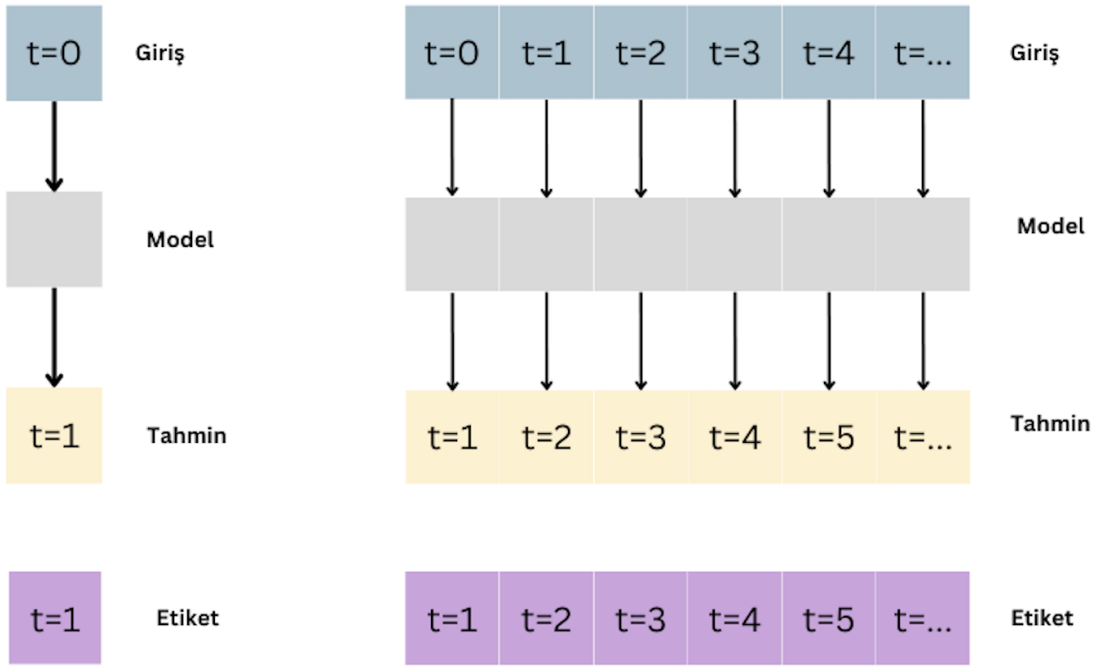
Tanım 2.2. Mevsimsellik (Seasonality): Veride gözlenen takvime bağlı düzenli değişimleri belirtmektedir. Bu değişimlerin görülme sıklığı sabittir. Kısa dönem aralıklarında gerçekleşen dalgalanmaları temsil etmektedir.

Tanım 2.3. Gürültü (Noise): Düzensizlik (Irregular) olarak da bilinmektedir. Verideki mevsimsellik ve trend etkilerini kaldırdıktan sonra geriye kalan düzensizliklere denir. Gürültüler, verideki kısa süreli dalgalanmalardır. Zaman serisinde tahmini zorlaştıran bir etmendir.

Tanım 2.4. Durağanlık (Stationary): Verinin zaman içerisinde, geçmişte ve gelecekte benzer olma durumunu ifade etmektedir. Bu durum varyans, ortalama ve otokovaryansının sabit olması ile açıklanmaktadır.

2.1.1 Zaman Serisinde Tahmin Adımları

Zaman serisi analizinde tahmin yapılırken modelin ne kadar geleceği tahmin edeceği ayarlanmalıdır. Aynı şekilde ne kadar önceki bilgiden çıkarımda bulunacağı ayarlanmalıdır. Şekil 2.1 'de tek adımlı (single-step) ve çok adımlı (multi-step) tahmin modelleri temsil edilmiştir [23].

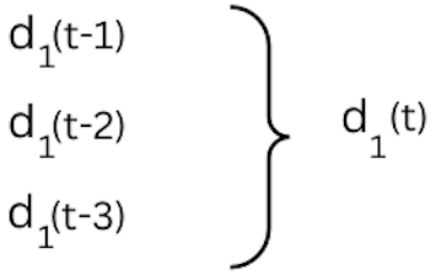


Şekil 2.1 Zaman Serilerinde Tahmin Adımları

2.1.2 Tek Değişkenli Zaman Serisi Analizi

Tek değişkenli (univariate) zaman serilerinde, incelenecek olan tek bir değişken vardır. Zaman serisi analizi bu değişkenin geçmiş gözlemlerine dayanmaktadır. Örneğin; glikoz seviyesinin tahmin edilmesi sadece geçmiş glikoz seviyelerine bağlı olarak yapılıyorsa bu durum tek değişkenli olarak ele alınmaktadır. Girdi olarak geçmiş kan şekeri değeri gözlemleri alınırken çıktı olarak glikoz seviyesinin gelecekteki

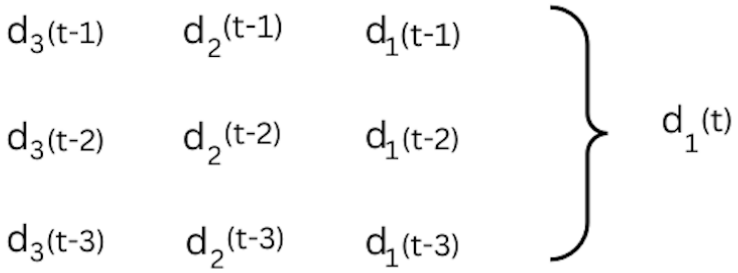
değerlerinin tahmini gözlemlenir. Şekil 2.2 'de tek değişkenli (univariate) tahmin temsil edilmiştir. Burada d_1 değişkenin t anındaki değeri tahmin edilirken 3 adım önceki değerleri alınarak aynı değişkene ait tahmin yapılmaktadır.



Şekil 2.2 Tek Değişkenli Tahmin

2.1.3 Çok Değişkenli Zaman Serisi Analizi

Çok değişkenli (multivariate) zaman serilerinde, incelenecek olan bir veya birden çok ana değişken ve bu değişkeni destekleyen yardımcı değişkenler bulunmaktadır. Birden çok değişkenin aynı zaman adımlarındaki geçmiş gözlemlerinin değerlendirilmesi ile analiz gerçekleştirilmektedir. Örneğin glikoz seviyesi; öğünlerin zamanı, türü ve miktarı, fiziksel aktivite süresi ve yoğunluğu, stres durumu, uyku kalitesi gibi birçok çevresel faktörden etkilenmektedir. Bu durumda glikoz seviyesi geçmişteki birçok değişkene bağlı olarak tahminlendiğinde çok değişkenli bir analiz gerçekleştirilmektedir. Girdi olarak birden çok yardımcı değişken ve glikoz seviyelerinin geçmiş değerleri alınırken çıktı olarak glikoz seviyesinin gelecekteki tahminleri gözlemlenir.



Şekil 2.3 Çok Değişkenli Tahmin

Şekil 2.3 'te çok değişkenli (multivariate) tahmin temsil edilmiştir. Burada d_1 değişkenin t anındaki değeri tahmin edilirken d_1, d_2 ve d_3 olmak üzere 3 değişkenin geçmiş değer bilgisi kullanılmaktadır. Girdi olarak 3 değişkenin 3 adım önceki değerleri alınarak çıktı elde edilmektedir.

2.2 Zaman Serisi Analizinde Klasik Tahmin Yöntemleri

Zaman serisi analizinde tahmin yöntemleri araştırılırken ilk olarak istatistiksel yöntemlere dayalı modeller kullanılmaktadır. Çalışma kapsamında klasik yöntemlerden probleme en uygun olarak Otoregresif Entegre Hareketli Model seçilmiştir. Modelin açıklanması için aşağıda temel bileşenleri açıklanmıştır.

2.2.1 Otoregresif Model

1920'lerdeki bir çalışmada zaman serilerini periyodik modellerle modelleme ve analiz etmek için Otoregresif Model (AR) önerilmiştir [24]. Geçmiş değerleri girdi olarak kullanarak çıktıların doğrusal olarak önceki değere bağlı olduğunu kabul etmektedir. Serideki değerler ile önceki ve sonraki değerler arasında bir korelasyon olduğu durumlarda kullanılmaktadır. AR(p) modeli aşağıda verilmiştir. Denklemde p gecikme (lag) gözlemlerinin sayısını, θ_t gecikme gözlemi ile o andaki değerin korelasyona bağlı değerini belirtmektedir.

$$Y_t = \beta_1 + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + ... + \theta_p Y_{t-p} \tag{2.1}$$

2.2.2 Hareketli Ortalama Modeli

Hareketli Ortalama Modeli (MA), gelecekteki değerin geçmiş hata terimlerine bağlı olduğunu belirtmektedir. Model çıktı değerini elde ederken serinin ortalamasını, o anki hata terimini ve geçmiş hata terimlerini doğrusal olarak ifade etmektedir. MA(q) modeli aşağıda verilmiştir. Eşitlikte q hareketli pencere sayısını, ε_t gecikme gözlemlerinin hatasını ve ω_t hatalardaki korelasyona bağlı değerini belirtmektedir.

$$Y_t = \beta_2 + \omega_1 \varepsilon_{t-1} + \omega_2 \varepsilon_{t-2} + ... + \omega_q \varepsilon_{t-q} + \varepsilon_t \tag{2.2}$$

2.2.3 Otoregresif Entegre Hareketli Model

AR ve MA modellerinin birleşimi ilk olarak Otoregresif Hareketli Model (ARMA) ile 1950 'lerdeki bir çalışmada keşfedilmiştir [25]. Otoregresif Entegre Hareketli Model (ARIMA) ise ARMA modelinin genişletilmiş halidir. Yalnızca zaman serileri durağan olduğunda çalışmaktadır. Bu yüzden durağan olmayan serileri durağan hale getirebilmek için verideki değerlerin önceki değerleri ile farkı alınmaktadır. Bu yönüyle ARMA 'dan ayrılmaktadır. Durağanlık durumunda ortalama sabit ve standart sapma sabit olmalıdır. Ek olarak mevsimsellik durumu olmamalıdır.

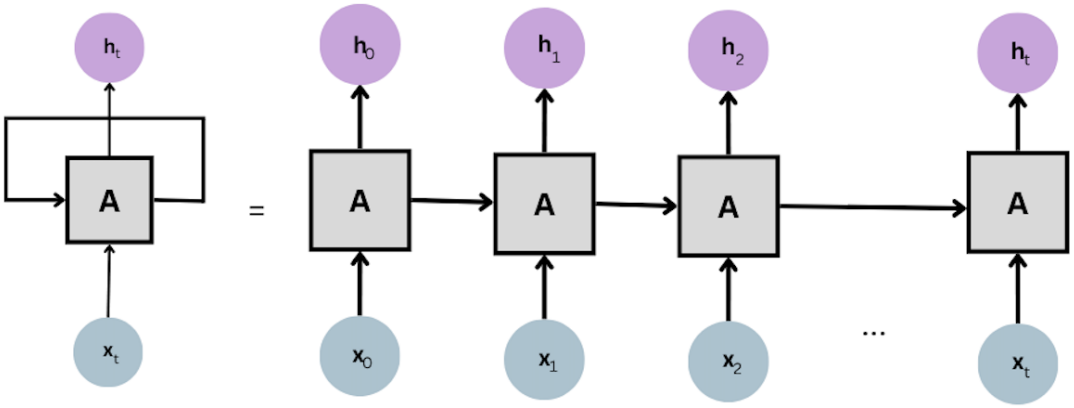
ARIMA modeli genelde negatif olmayan üç parametre olan p, d ve q ile tanımlanır. "p" değeri otoregresif modelin sırası, "q" değeri hareketli ortalamanın sırası ve "d" değeri ise fark alma derecesidir. Aşağıdaki eşitliklerde ARIMA(p,d,q) modeli verilmiştir. Burada AR(p) ve MA(q) modellerinin kombinasyonuna 1. dereceden fark uygulanarak Z_t elde edilmektedir.

$$Y_t = (\beta_1+\beta_2)+(\theta_1 Y_{t-1}+\theta_2 Y_{t-2}+...+\theta_p Y_{t-p})+(\omega_1\varepsilon_{t-1}+\omega_2\varepsilon_{t-2}+...+\omega_q\varepsilon_{t-q}+\varepsilon_t) \quad (2.3)$$

$$Z_t = Y_{t+1} + Y_t \quad \text{d=1} \quad (2.4)$$

2.3 Zaman Serisi Analizinde Derin Öğrenme Yöntemleri

Derin öğrenme yaklaşımları sağlık alanındaki problemlere güvenilir çözümler getirmektedir. Zaman serisi problemlerini çözmek için veriden öğrenen derin öğrenme modelleri kullanılmaktadır. Çalışma kapsamında üç farklı sinir ağı; Tekrarlayan Sinir Ağları (RNN), Geçitli Tekrarlayan Birimler (GRU) ve Uzun Kısa Süreli Bellek Ağları (LSTM) tabanlı modeller açıklanmıştır.



Şekil 2.4 Tekrarlayan Sinir Ağları Açılımı

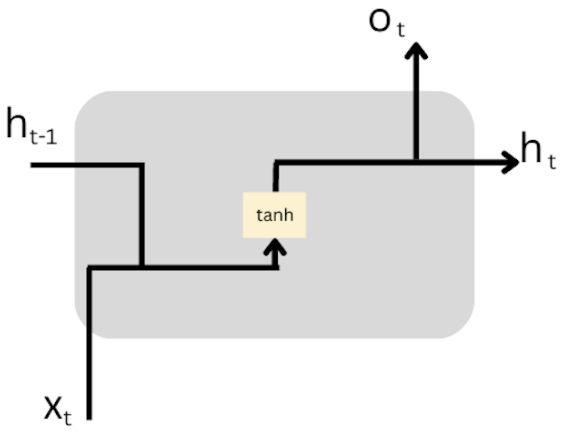
2.3.1 Yinelemeli Sinir Ağları

Yinelemeli Sinir Ağları (RNN), bir veri noktasının bir öncekine bağlı olduğu sıralı verilerle çalışmak üzere 1980 'lerdeki bir çalışmada icat edilmiştir [26]. Geleneksel sinir ağlarının sahip olmadığı, önceki bilgilerin saklanabileceği bir belleğe sahiptir. RNN 'ler sahip oldukları kısa süreli bellek sayesinde, bir sonraki çıktıyı oluşturmak

için önceki girdilerden gelen bilgileri saklamakta ve sonraki değerlerde bu bilgileri kullanmaktadırlar.

Şekil 2.4 'de, Yinelemeli Sinir Ağları 'nın mimarisinin genel gösterimi verilmiştir. Burada her adımda girdi olarak bir önceki adımın çıktısı beslenmektedir. Tekrarlı yapının genel formülasyonu aşağıda verilmiştir. Burada t-1 anındaki gizli durum h_{t-1} ve girdi vektörü x_t bir f aktivasyon fonksiyonundan geçirilerek t anındaki gizli durum (hidden state) h_t hesaplanmaktadır. Gizli durum ağa kısa süreli bellek özelliğini kazandırmaktadır. Öğrenilen bilgiyi sonraki hücreye aktarmaktadır.

$$h_t = f(h_{t-1}, x_t) \tag{2.5}$$



Şekil 2.5 Yinelemeli Sinir Ağlarının Bir Hücresi

- h_{t-1} : t-1 anındaki gizli durum (kısa süreli bellek)
- x_t : t anındaki giriş vektörü
- h_t : t anındaki gizli durum
- o_t : ağın t anındaki çıktısı

Şekil 2.5 'te bir tekrarlama birimi verilmiştir. Burada t-1 zamanındaki gizli durum h_{t-1} ile t zamanındaki giriş vektörü x_t birleştirilmektedir. Ardından tanh aktivasyon fonksiyonundan geçerek hücrenin t anındaki çıktısı olan o_t 'yi ve sonraki hücreye aktırılan gizli durum h_t 'yi oluşturmaktadır.

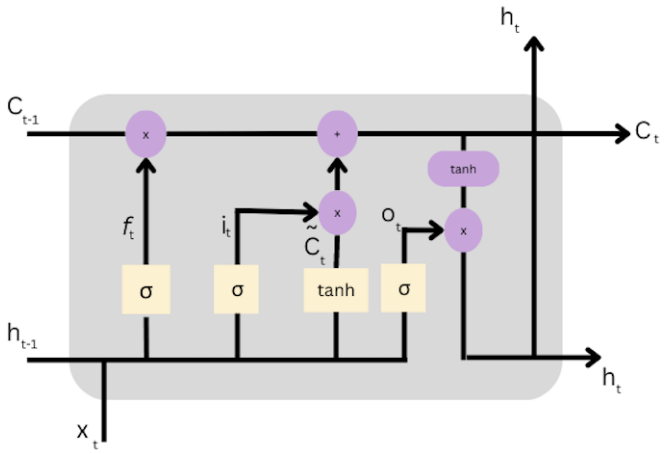
$$h_t = tanh(W_x x_t + W_h h_{t-1} + b_h) \tag{2.6}$$

$$o_t = tanh(W_o h_t + b_o) \tag{2.7}$$

Tekrarlayan Sinir Ağları'nda geri yayılımda ortaya çıkabilen iki problem vardır; kaybolan gradyan (vanishing gradient) ve patlayan gradyan (exploding gradient). Geri yayılımda gradyanlar, matris çarpımlarına tabidirler. Gradyan küçük değerlere sahipse gitgide küçülebilir ve sonunda kaybolabilir. Kaybolan gradyan problemi bilgi kaybına sebep olmaktadır. Eğer gradyan, büyük değerlere sahipse gitgide büyüyerek en sonunda patlayabilir. Bu durumda ise modelin çalışmasını durdurabilmektedir.

2.3.2 Uzun Kısa Süreli Bellek Ağları

Uzun Kısa Süreli Bellek Ağları [27], RNN mimarisinin özelleştirilmiş bir versiyonu olarak 1990'larda geliştirilmiştir. Uzun Kısa Süreli Bellek Ağları, hücreler (cell) ve kapılardan (gate) oluşmaktadır. Kapıların temel mantığı, hücrelerdeki bilgi akışını kontrol etmektir. Kısa süreli bellek ve gradyan problemlerine çözümler sunmaktadır. Ağ yapısında hücre durumu (cell state), giriş kapısı (input gate), çıkış kapısı (output gate) ve unutma kapısı (forget gate) bulunmaktadır. Hücre durumu, ağa uzun süreli bellek özelliğini kazandırmaktadır. Kapılar, hücre durumunu kontrol ederek bilgilerin hatırlanması, unutulması gibi işlevleri yönetmektedir.



Şekil 2.6 Uzun Kısa Süreli Bellek Ağlarının Bir Hücresi

- h_{t-1} : t-1 anındaki gizli durum (kısa süreli bellek)
- C_{t-1} : t-1 anındaki hücre durumu (uzun süreli bellek)
- x_t : t anındaki giriş vektörü
- h_t : t anındaki gizli durum

- C_t : t anındaki hücre durumu
- $\tilde{C_t}$: t anındaki yeni hücre bilgisi
- f_t : unutma kapısı
- i_t : giriş kapısı
- o_t : çıkış kapısı

Şekil 2.6 'da LSTM ağının bir tekrarlama birimi verilmiştir. Bu birime girişte 3 bileşen; hücre durumu, gizli durum ve giriş vektörü bulunmaktadır. Burada t-1 zamanındaki gizli durum h_{t-1} ile t zamanındaki giriş vektörü x_t birleştirilmektedir. Ardından kapı işlemlerine geçmektedir.

Unutma kapısında, sigmoid (σ) aktivasyon fonksiyonundan geçmektedir ve hücre durumu c_{t-1} ile noktasal çarpım işlemi gerçekleşmektedir. Burada, hangi bilgilerin silineceğine ve hangilerinin tutulacağına karar verilmektedir.

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{2.8}$$

Giriş kapısı, hangi bilgilerin güncelleneceğine karar vermektedir. Girdi vektörü ile önceki gizli durum sigmoid fonksiyonundan geçmektedir. Giriş kapısından gelen yeni bilgiler $\tilde{C}$, giriş kapısı tarafından ağırlıklandırılmaktadır.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{2.9}$$

$$\tilde{C_t} = tanh(W_C.[h_{t-1}, x_t] + b_C) \tag{2.10}$$

Hücre durumu, hücreler arasında taşınması gereken bilgileri hareket ettirerek veri akışını sağlamaktadır, bu nedenle ağın uzun süreli hafızası işlevini yerine getirmektedir. Önceki bilgiler ile işlemin çıktısı yeni hücre durumunu oluşturmaktadır.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C_t} \tag{2.11}$$

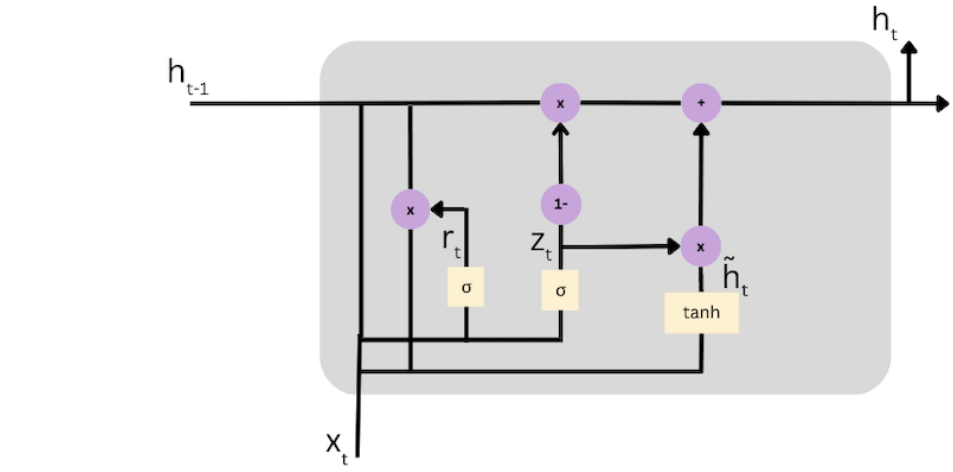
Çıkış kapısı, hücrenin çıkışına yani bir sonraki hücrenin girdisini oluşturmaktadır. Son gizli durumda ise hücrenin çıktısını vermektedir. Eğer ağdaki son birim ise ağın tahmini temsil etmektedir.

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{2.12}$$

$$h_t = o_t * tanh(C_t) \tag{2.13}$$

2.3.3 Kapılı Yinelemeli Geçitler

2010'larda ortaya çıkan Kapılı Yinelemeli Geçitler [28], RNN 'in geliştirilmiş bir versiyonu ve LSTM 'in basitleştirilmiş bir versiyonudur. Uzun Kısa Süreli Bellek Ağları gibi kısa süreli bellek ve gradyan problemlerine çözümler sunmaktadır. Kapılı Yinelemeli Geçitler, güncelleme kapısı (update gate) ve sıfırlama kapısından (reset gate) oluşmaktadır. LSTM Ağları' na kıyasla daha az bellek kullanılmaktadır bu sebeple daha hızlı çalışan bir modeldir. Kapılar, bilginin hatırlanması, unutulması işlevlerini yönetmektedir.



Şekil 2.7 Kapılı Yinelemeli Geçitlerin Bir Hücresi

- x_t : t anındaki giriş vektörü
- h_{t-1} : t-1 anındaki gizli durum
- h_t : t anındaki gizli durum
- $\tilde{h_t}$: t anındaki aday gizli durum
- C_t : t anındaki hücre durumu
- r_t : sıfırlama kapısı

- z_t : güncelleme kapısı

Şekil 2.7 'da Kapılı Yinelemeli Geçitler 'in bir tekrarlama birimi verilmiştir. Bu birime girişte 2 bileşen; gizli durum ve giriş vektörü bulunmaktadır. Burada t-1 zamanındaki gizli durum h_{t-1} ile t zamanındaki giriş vektörü x_t birleştirilmektedir. Ardından kapı işlemlerine geçilmektedir. Birimin son hücresinde olması durumunda h_t ağın çıktısını vermektedir.

Sıfırlama kapısı, bilgilerin hangilerinin unutulacağına hangilerinin tutulacağına karar vermektedir. Ağın kısa süreli bellek işlevini yerine getirmektedir. Giriş vektörü ve önceki gizli durum Sigmoid fonksiyonundan geçirilir. Böylece unutulan (0), hatırlanan (1) ve kısmen hatırlanan (0-1) değerlere karar verilmektedir.

$$r_t = \sigma(W_r.[h_{t-1}, x_t] + b_r) \tag{2.14}$$

Güncelleme kapısı, modelin geçmiş bilgilerinden ne kadarının geleceğe aktarılacağına karar vermektedir. Bu karara Sigmoid fonksiyonun sonucunda ulaşmaktadır. Ağın uzun süreli hafızası görevini sağlamaktadır.

$$z_t = \sigma(W_z.[h_{t-1}, x_t] + b_z) \tag{2.15}$$

Gizli durumun güncel değeri yeni gizli durum adayıdır ve $\tilde{h_t}$ olarak ifade edilmektedir. Burada sıfırlama kapısının etkisinin önemi görülmektedir. Buradaki "*" işlemi eleman bazında çarpımı ifade etmektedir.

$$\tilde{h_t} = tanh(W_r.[r_t * h_{t-1}, x_t] + b_r) \tag{2.16}$$

Gizli durumun son değeri hesaplanırken güncelleme kapısının kararı uygulanmaktadır. Burada önceki bilgi h_{t-1} ve yeni aday bilgi $\tilde{h_t}$ 'den hangi bilgilerin yeni durum h_t 'yi oluşturucağına karar verilmektedir. Bu değer birimin çıktı değerini belirtmektedir.

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h_t} \tag{2.17}$$

3
OHIOT1DM VERİ SETİ VE ANALİZİ

Çalışmada ilk etapta önemli bir adım olan veri analizi gerçekleştirilmiştir. Bu keşif sırasında her hasta için hangi bilgilere sahip olunduğu, sahip olunan bilgilerin boşluk durumu, frekansları, tipleri, uzunlukları ve dağılımları analiz edilmiştir. Zaman serisi verisi olduğu için kaç dakika geçmişe dönük bir eğitim yapılcağı ve kaç dakika geleceğe yönelik tahmin yapılacağı ayarlanmıştır. Veri en optimal şekilde kullanıma hazırlanmıştır. Kan şekeri tahmini çalışmalarında kişiselleştirilmiş modellere odaklanıldığı için hangi hasta için hangi bilgilerin kullanılacağı konusu önemlidir. Her hasta bazında en uygun bilgilerin kullanılması için öznitelikler oluşturulmuş ve değerlendirilmiştir.

3.1 Verinin Elde Edilmesi

Çalışmada OhioT1DM[29] veri seti kullanılmıştır. Veri seti Ohio Üniversitesi ve Yıldız Teknik Üniversitesi arasında yapılan D201804 numaralı Veri Kullanım Sözleşmesi ile sağlanmıştır. Veri kümesi ilk kez 2018 yılında yapılan Blood Glucose Level Prediction (BGLP) Yarışması 'nda 6 hastanın verisi ile ortaya çıkmıştır. 2020 yılında yapılan ikinci yarışmada ise 6 hasta verisi daha eklenerek veri kümesi 12 hasta verisi ile güncellenmiştir. Veri setinde tip-1 diyabete sahip hastaların 8 haftalık verisi bulunmaktadır. Veriler hastanın kullandığı sensör, vücut bandı ve yaşamsal aktivitelerinin manuel kaydedilmesi ile elde edilmiştir.

Tablo 3.1 'de her hasta için cinsiyet, sensör modeli ve çalışmada sıklıkla bahsedilecek olan hasta numarası, kohort bilgisi verilmiştir. Kohortlardaki hastaların kullandıkları vücut bantlarının farklı olması sebebiyle 2018 ve 2020 kohortundaki hastalarda bazı bilgilerde frekans farklılıkları ve bilgi eksiklikleri mevcuttur. Örneğin, 2018 kohortundaki hastalarda Basis sensöründen elde edilen kalp atışı, galvanik deri tepkisi, vücut sıcaklığı mevcutken 2020 kohortunda bulunmamaktadır. Aynı zamanda hastaların manuel giriş yapmasından kaynaklı bilgi eksiklikleri de mevcuttur.

Tablo 3.1 Hasta Bilgisine Genel Bakış

Hasta No.	Cinsiyet	Sensör Bandı	Kohort
559	Kadın	Basis	2018
563	Erkek	Basis	2018
570	Erkek	Basis	2018
575	Kadın	Basis	2018
588	Kadın	Basis	2018
591	Kadın	Basis	2018
540	Erkek	Empatica	2020
544	Erkek	Empatica	2020
552	Erkek	Empatica	2020
567	Kadın	Empatica	2020
584	Erkek	Empatica	2020
596	Erkek	Empatica	2020

3.1.1 Veri Dosyalarının Analizi

Paylaşılan veride 12 hasta için eğitim ve test dosyaları olmak üzere 24 adet XML formatında dosya mevcuttur. Dosyalarda her hasta için elde edilen bilgiler etiketlerde tutulmaktadır. Bu etiketler incelendiğinde 19 adet etiket bilgisi bulunmaktadır. Bu inceleme sonucunda hastaların hangi etiket bilgilerine sahip olduğu ortaya konmuştur. Veri dosyalarının hasta bazında incelenmesi sonucunda elde edilen sonuçlar ortaya konmuştur. Tablo 3.2 'de her hasta için eğitim verisinde hangi etiket bilgisine sahip olduğu verilmiştir. Eğitim dosyasının varlığı veri sayısı açısından önemlidir. Çünkü oransal olarak baktığımızda eğitim veri sayısı teste göre daha fazladır.

Tablo 3.2 Eğitim Verisinde Etiket Bilgisi Analizi

Etiket/Hasta	540	544	552	567	584	596	559	563	570	575	588	591
glucose_level	+	+	+	+	+	+	+	+	+	+	+	+
finger_stick	+	+	+	+	+	+	+	+	+	+	+	+
basal	+	+	+	+	+	+	+	+	+	+	+	+
temp_basal	+	+	+	+	+	+	+	+	+	+	+	+
bolus	+	+	+	+	+	+	+	+	+	+	+	+
meal	+	+	+	+	+	+	+	+	+	+	+	+
sleep		+	+	+	+	+	+	+	+	+	+	+
work		+	+		+		+	+	+	+	+	
stressors		+				+	+				+	
hypo_event		+	+			+	+	+	+	+	+	+
illness			+	+		+	+	+		+	+	+
exercise		+	+		+	+	+	+	+	+	+	+
basis_heart_rate							+	+	+	+	+	+
basis_gsr	+	+	+	+	+	+	+	+	+	+	+	+
basis_skin_temperature	+	+	+	+	+	+	+	+	+	+	+	+
basis_air_temperature							+	+	+	+	+	+
basis_steps							+	+	+	+	+	+
basis_sleep							+	+	+	+	+	+
acceleration	+	+	+	+	+	+						

Tablo 3.3 'de her hasta için test verisinde hangi etiket bilgisine sahip olduğu verilmiştir. Test dosyalarının yok olma durumu eğitim dosyalarına göre fazladır. Veri sayısı açısından kaybı önlemek için eğitim ve test verileri art arda eklenmiştir.

Tablo 3.3 Test Verisinde Etiket Bilgi Analizi

Etiket/Hasta	540	544	552	567	584	596	559	563	570	575	588	591
glucose_level	+	+	+	+	+	+	+	+	+	+	+	+
finger_stick	+	+	+	+	+	+	+	+	+	+	+	+
basal	+	+	+	+	+	+	+	+	+	+	+	+
temp_basal	+	+	+	+		+	+			+	+	+
bolus	+	+	+	+	+	+	+	+	+	+	+	+
meal	+	+	+		+	+	+	+	+	+	+	+
sleep		+	+		+	+	+	+	+	+	+	+
work		+	+		+		+	+	+	+	+	
stressors											+	
hypo_event		+	+		+	+						
illness							+					+
exercise		+	+		+	+	+	+	+		+	+
basis_heart_rate							+	+	+	+	+	+
basis_gsr	+	+	+	+	+	+	+	+	+	+	+	+
basis_skin_temperature	+	+	+	+	+	+	+	+	+	+	+	+
basis_air_temperature							+	+	+	+	+	+
basis_steps							+	+	+	+	+	+
basis_sleep							+	+	+	+	+	+
acceleration	+	+	+	+	+	+						

3.1.2 Verideki Etiket Bilgisinin Analizi

Veri incelendiğinde toplamda 19 etiket bilgisi bulunmaktadır. Bazı etiketler kendi içerisinde ek etiketler de barındırabilmektedir. Bu da bize yeni öznitelikler üretme olanağı sağlamaktadır. Etiketlerin bir kısmı öznitelik olarak kullanılırken bir kısmı yeni özniteliklerin çıkarılması için kullanılmıştır. Tablo 3.4 'te ham veri üzerinde eğitim ve test verisi birlikte olmak üzere, etiketlerden elde edilen veri noktalarının sayısal analizi verilmiştir. Burada her etiket için 12 hastaya bakılmıştır. Söz konusu bilgi dosyasına sahip hastalara ait minimum ve maximum veri noktası sayısı verilmiştir. Veriler incelendiğinde her hasta için ortak bir ölçekte örnekleme yaptığımızı düşünürsek minimum veri noktasına sahip hastalarda . Etiket bilgisine genel olarak baktığımızda verinin ortak bir ölçekte hazırlanmasının ve ön işleme adımlarının kritik bir aşama olduğu anlaşılmaktadır. Aşağıda etiket bilgileri tanımlanmıştır.

<glucose_level> : Kan şekeri değeri ve ölçülen zaman bilgisine sahiptir. Ölçümler 5 dakikada bir olmak üzere kaydedilmiştir.

<finger_stick> : Hastanın parmak ölçümü ile ettiği kan şekeri değerini ve zamanını belirtmektedir. Hastanın ölçümlerine bağlı olduğu için düzensiz sıklıkta bir bilgidir.

<basal> : Sürekli salgılanan bazal insülin oranının değerini ve zamanını belirtir.

Tablo 3.4 Etiket Veri Noktası Sayı Analizi

Bilgi	Min. Veri Noktası	Maks. Veri Noktası
glucose_level	9080	12640
finger_stick	137	592
basal	23	163
temp_basal	2	232
bolus	134	347
meal	32	265
sleep	12	46
work	6	35
stressors	1	3
hypo_event	1	3
illness	1	4
exercise	2	46
basis_heart_rate	11966	12980
basis_gsr	11769	53219
basis_skin_temperature	11822	53219
basis_air_temperature	11822	12948
basis_steps	11971	13297
basis_sleep	876	1087
acceleration	25604	53219

Belirtilen zamanda başlar ve yeni bir bazal insülin oranı ayarlanana kadar aynı oran devam eder. Düzensiz sıklıkta bir bilgidir.

<temp_basal> : Hastanın normal aktivitelerinin dışında takviye gereken durumlarda aldığı bazal insülindir. Geçici bazal insülin oranı değeri, başlangıç ve bitiş zamanı bilgilerine sahiptir. Eğer hasta hiç ihtiyaç duymadıysa tüm değerleri sıfırdır. Bazal değer ile birlikte anlamlı bir bilgi bütünü oluşturmaktadır.

<bolus> : Hastaya tipik olarak öğün öncesi veya hiperglisemi durumunda verilen insülin türüdür. Etiket içerisinde bolus dozu değeri, bolus tipi, başlangıç ve bitiş zamanı bilgilerini tutmaktadır. 2018 kohortunda ek olarak alınan karbonhidrat tahmini bilgisi vardır. Bolus tipi etiketinde; normal, normal dual, dual, square dual olmak üzere 4 farklı tip bulunmaktadır. Düzensiz sıklıktadır.

<meal> : Hastanın öğünün karbonhidrat alımını, öğün tipini ve öğün zamanını belirtir. Veri sıklığı hastanın yaşamına bağlı olduğu için düzensizdir. Öğün tipi etiketinde breakfast, lunch, dinner, snack, hypocorrection olmak üzere 5 farklı tür bulunmaktadır.

<sleep> : Hastanın uyku saat aralığını ve uyku kalitesini belirtir. Uyku kalitesi zayıftan iyiye olmak üzere 1-2-3 değerleri ile ifade edilmiştir. Genel olarak günde

bir kayıtlar mevcut olsa da veri aralıkları düzenli değildir.

<work> : Hastanın işe gidip gelme saatlerini ve yoğunluğunu belirtir. Yoğunluk değeri, hastanın gün içerisinde harcadığı fiziksel eforu 1-10 arasında puanlanması ile oluşturulmuştur. Manuel bilgi girişi yapıldığı için düzensiz sıklıktadır.

<stressors> : Hastanın stresli olduğu zamanlar kaydettiği zaman bilgisinden oluşmaktadır. Ek olarak tür ve açıklama etiketlerini barındırır ancak bilgiler boş olduğu için kullanılamamıştır. Düzensiz bir bilgidir.

<hypo_event> : Hastanın hipoglisemik olay yaşadığı zamanları kaydetmesiyle oluşmuştur. Bazı hastalarda bu bilginin bulunmaması hipoglisemik olay yaşamadığını belirtmektedir.

<illness> : Hastanın hasta olduğu zamanları kaydetmesiyle oluşmuştur. Hasta veriye sahip değilse 8 haftada hiç hasta olmadığı yönünde değerlendirilmektedir. GEnel olarak bu bilgi eksik veya 8 haftada bir kez bulunmaktadır.

<exercise> : Hastanın egzersiz yaptığı zamanı, egzersiz süresini ve yoğunluğunu ifade etmektedir. Egzersiz yoğunluğu değeri azdan çoğa olmak üzere 1 ile 10 arasındaki puanlandırmayı belirtmektedir. Düzensiz bir bilgidir.

<basis_heart_rate> : Kalp atışı ölçümü ve ölçüm zamanını ifade etmektedir. Ölçümler 5 dakikalık periyotta kaydedilmiştir. Sensör farklılığı sebebiyle sadece 2018 kohortunda mevcuttur. Düzenli bir veridir.

<basis_gsr> : Galvanik cilt tepkisi, elektrodermal aktivite veya cilt iletkenliği olarak da bilinmektedir. Hastanın galvanik cilt tepkisi değerini ve kaydedildiği zamanı belirtmektedir. Sensör farklılığı sebebiyle 2018 kohortunda 5 dakikalık, 2020 kohortunda 1 dakikalık sıklıkta kaydedilmiştir. Düzenli bir veridir.

<basis_skin_temperature> : Hastanın vücut sıcaklığını ve kaydedildiği zamanı belirtmektedir. Sensör farklılığı sebebiyle 2018 kohortunda 5 dakikalık, 2020 kohortunda 1 dakikalık periyota sahiptir. Düzenli bir veridir.

<basis_air_temperature> : Hava sıcaklık değerini ve ölçüm zamanını belirtmektedir. Sensör farklılığı sebebiyle sadece 2018 kohortundaki hastalarda mevcuttur. Ölçümler 5 dakikalık periyotta kaydedildiği için düzenli bir veridir.

<basis_steps> : Hastanın adım sayısını ve ölçüm zamanını belirtmektedir. Sensör farklılığı sebebiyle sadece 2018 kohortunda mevcuttur. Ölçümler 5 dakikada bir kaydedilmektedir. Düzenli bir veridir.

<basis_sleep> : Hastanın uykudaki zaman aralıklarını ve bu aralıklardaki uyku kalitesi değerini belirtmektedir. Veri aralıkları düzenli değildir. Sensör farklılığından ötürü sadece 2018 kohortunda mevcuttur.

<acceleration> : Hastanın ivme büyüklüğü değerini ve ölçülen zamanı belirtmektedir. Sensör farklılığı sebebiyle sadece 2020 kohortundaki hastalarda mevcuttur. Ölçümler 1 dakikalık periyotta olduğu için düzenli bir veridir.

Tanımlanan etiketler veri setinde bulunduğu isimleriyle listelenmiştir. Etiketler özniteliklere dönüştürüldükten sonra çalışma dahilinde örneklenirken hangi bilginin ne şekilde örneklenirse daha anlamlı olacağı araştırılmıştır. Sağlık alanındaki anlamı düşünülerek sonraki kısımlarda anlatılacak olan veri ön işleme işlemleri gerçekleştirilmiştir.

3.1.3 Etiketlerin Kullanılabilirlik Analizi

Tablo 3.5 'te ham veri üzerinde etiketlerden elde edilen bulgular verilmiştir. Burada her etiket için veri sıklığı, zaman bilgisi, veri tipi ve ek etiket bilgisi ortaya konmuştur. Veri setinde bulunan etiket bilgilerinin kullanılabilirlik analizi sonucunda ilk aşamada temel öznitelikler belirlenmiştir. Zaman serisinde öznitelik oluşturulurken verinin ölçüm sıklığı önemli bir rol oynamaktadır. Buna dayanarak ilk aşamada sensörden elde edilen kan şekeri ve düzenli bilgiler çalışmaya dahil edilmiştir. Kohortlardaki hastaların sahip olduğu bilgiler farklı olduğundan öznitelikler kohort bazlı değişebilmektedir.

3.2 Veriye Yaklaşım Stratejisi

Veri analiz edildikten sonra hangi etiket bilgisinin sağlık verisinde ne ifade ettiğine odaklanılmıştır. Çünkü verilerin doldurulması ve yeniden örneklenmesi aşamasında öznitelikler birleştirilirken anlamlı veya anlamsız boşluklar oluşmaktadır. Mesela kan şekeri ölçümleri saat 09:55'te başlarken kalp atışı sensör ölçümleri 10:25'te başlamaktadır. Bu durumda iki özelliğin birleştirilmesi kalp atışı bilgisinde yarım saatlik bir boşluğa sebep olmaktadır. Diğer bir örnekte ise ham veride ardışık görünse de 5 dakikalık zaman aralıklarına örneklendikten sonraki durumda boşluklar oluşmaktadır. Bu durumlarda veriye yaklaşım stratejisi önemlidir. Verinin daha iyi anlaşılması için sağlık alanında uzman kişilerin görüşleri öznitelikler bazında alınmıştır.

Tablo 3.5 Etiketlerin Kullanılabilirlik Analizi

Etiket	Veri Sıklığı	Zaman	Veri Tipi	Ek Etiket Bilgisi
glucose_level	5 dk	ts	nümerik	-
finger_stick	düzensiz	ts	nümerik	-
basal	düzensiz	ts	nümerik	-
temp_basal	düzensiz	ts-ts	nümerik	-
bolus	düzensiz	ts-ts	nümerik	type(nominal)
meal	düzensiz	ts	nümerik	type(nominal)
sleep	düzensiz	ts-ts	oransal	-
work	düzensiz	ts-ts	oransal	-
stressors	düzensiz	ts	nominal	
hypo_event	düzensiz	ts	nominal	-
illness	düzensiz	ts	nominal	-
exercise	düzensiz	ts	oransal	duration(nümerik)
basis_heart_rate	5 dk	ts	nümerik	-
basis_gsr	1 dk / 5 dk	ts	nümerik	-
basis_skin_temperature	1 dk / 5 dk	ts	nümerik	-
basis_air_temperature	5 dk	ts	nümerik	-
basis_steps	5 dk	ts	nümerik	-
basis_sleep	düzensiz	ts-ts	oransal	-
acceleration	1 dk	ts	nümerik	-

3.3 Özniteliklerin Oluşturulması

Veri setinde bulunan etiket bilgilerinin kullanılabilirlik analizi sonucunda temel öznitelikler belirlenmiştir. Zaman serisinde öznitelikler oluşturulurken verinin ölçüm sıklığı önemli bir rol oynamaktadır. Buna dayanarak ilk aşamada sensörden elde edilen düzenli bilgiler çalışmaya dahil edilmiştir. Devamında ise sonuçları iyileştirmeye yönelik deneyler yapılmış olup sonucu iyileştiren öznitelikler eklenmiştir. Kohortlardaki hastaların sahip olduğu bilgiler farklı olduğundan öznitelikler kohort bazlı değişmektedir. Aşağıda özniteliklerin kullanılabilir olanların kohorta göre ayrılmış genel bir listesi verilmiştir.

- 2020 kohortu; glucose_level, basis_gsr, basis_skin_temperature, acceleration, basal, bolus, temp_basal, meal
- 2018 kohortu; glucose_level, basis_gsr, basis_heart_rate, basis_steps, basis_sleep, basis_skin_temperature, basis_air_temperature, basal, bolus, temp_basal, meal

3.3.1 Düzenli Özniteliklerin Kullanımı

Sensör verisinden elde edilen bilgiler düzenli frekansa sahiptir. Tablo 3.5 'te düzenli etiketler araştırılmıştır. Düzenli veri sıklığına sahip bilgiler incelendiğinde sıklığın 1 dakika veya 5 dakika olduğu görülmektedir. Çalışmada kan şekeri tahmini yapıldığı için diğer tüm özelliklerin sıklığı 5 dakika olacak şekilde hizalanmıştır.

3.3.2 Düzensiz Özniteliklerin Kullanımı

Kullanıcı tarafından manuel olarak girilen veya eşit örnekleme frekansına sahip olmayan bilgiler düzensiz olarak ele alınmıştır. Tablo 3.4 'teki veri bilgisinde düzensiz etiketler araştırılmıştır. Düzensiz veri sıklığına sahip bilgilerin veri sayısına baktığımızda ve hasta bazında incelendiğinde ne kadar düzensiz olduğu hakkında bilgi sahibi olunmaktadır. Mesela "meal" bilgisini günde ortalama 3 olarak düşünürsek 8 haftalık veride 168 civarını normal olarak kabul edebiliriz. Bu etikette 32 veri noktasına sahip hasta için düzensizlik söz konusudur.

Düzensiz bilgilerin kullanılabilmesi için iki durum araştırılmıştır;
1) Düzensiz sıklıktaki verilerin düzenli hale getirilmesi
2) Düzenli hale gelen veriden yeni özniteliklerin çıkarımı

3.4 Verinin Zaman Serisine Dönüştürülmesi

Veri, özniteliklerin seçimi, birleştirme ve ön işleme adımları gerçekleştirildikten sonra zaman serisi verisine dönüştürülmüştür. Şekil 3.1 'de bir hasta için hazırlanan örnek veri seti gösterilmiştir. Veri seti 6 adım önceki değerleri modele girdi olarak almakta olup, 6 adım sonraki kan şekeri değerini tahmin etmektedir.

ts	var1(t-1)	var2(t-1)	var3(t-1)	var4(t-1)	var5(t-1)	var6(t-1)	var1(t+5)
2021-12-08 07:05:00	0.5722222222222222	1.4814859470894955e-06	0.5114502384266494	0.2727272727272727	0.48369551070522365	0.13333339956071377	0.49722222222222223
2021-12-08 07:10:00	0.5638888888888889	2.4444520821784217e-05	0.5038166382910158	0.2727272727272727	0.48913030207746644	0.13333339956071377	0.4777777777777778
2021-12-08 07:15:00	0.5444444444444444	4.180568643271719e-05	0.4809161614377169	0.1958041958041958	0.5326086330554087	0.13333339956071377	0.44722222222222224
2021-12-08 07:20:00	0.5444444444444444	0.0008909287103809098	0.5648854393760852	0.23776223776223776	0.5489130071721371	0.13333339956071377	0.4583333333333333
2021-12-08 07:25:00	0.5305555555555556	0.0010212995330105515	0.5343510388335504	0.22377622377622378	0.47826071933298087	0.13333339956071377	0.4305555555555556
2021-12-08 07:30:00	0.5027777777777778	3.2546397747578597e-05	0.5114502384266494	0.20279720279720279	0.4728259279607381	0.13333339956071377	0.41388888888888886
2021-12-08 07:35:00	0.49722222222222223	1.347226424190842e-05	0.5725190395117188	0.20279720279720279	0.5163042589386804	0.13333339956071377	0.4222222222222222
2021-12-08 07:40:00	0.4777777777777778	9.828734094452447e-05	0.5648854393760852	0.26573426573426573	0.47826071933298087	0.13333339956071377	0.42777777777777776
2021-12-08 07:45:00	0.44722222222222224	5.76390686291081e-05	0.5419846389691841	0.22377622377622378	0.48369551070522365	0.13333339956071377	0.4527777777777778
2021-12-08 07:50:00	0.4583333333333333	3.5648258338754905e-06	0.5114502384266494	0.16083916083916083	0.47826071933298087	0.13333339956071377	0.4222222222222222
2021-12-08 07:55:00	0.4305555555555556	1.4351894901897642e-06	0.5343510388335504	0.16083916083916083	0.47826071933298087	0.13333339956071377	0.39166666666666666
2021-12-08 08:00:00	0.41388888888888886	1.157411085642324e-06	0.5496182391048178	0.20279720279720279	0.499999884821952	0.13333339956071377	0.37777777777777777
2021-12-08 08:05:00	0.4222222222222222	2.5000076620326267e-06	0.5801526396473524	0.2727272727272727	0.5108694675664376	0.8333333002196431	0.37222222222222223
2021-12-08 08:10:00	0.42777777777777776	4.0740865229215826e-06	0.5572518392404514	0.27972027972027974	0.5217390503109232	0.8333333002196431	0.4027777777777778
2021-12-08 08:15:00	0.4527777777777778	0.0004078716508831009 7	0.5572518392404514	0.24475524475524477	0.5163042589386804	0.8333333002196431	0.41388888888888886
2021-12-08 08:20:00	0.4222222222222222	0.0001575004982031989	0.5419846389691841	0.16783216783216784	0.499999884821952	0.8333333002196431	0.4111111111111111

Şekil 3.1 Eğitimden Önce Veri Seti Görünümü

4
METODOLOJİ

Bu bölümde kan şekeri tahmini yapmak için veriye uygulanan teknikler ve uygulanan modeller açıklanmıştır. Veri ön işleme adımında veriye yaklaşım şekli detaylandırılmıştır. Özniteliklerin seçilmesi sebepleriyle birlikte açıklanarak modellere olan etkisi araştırılmıştır. Kan şekeri tahmini problemi için çözüm algoritması aşağıdaki adımlarda açıklanmıştır:

1) Veri seti araştırılması ve verinin elde edilmesi
2) Veriye uygulanacak ön işleme adımlarının araştırılması
3) Yeni özniteliklerin oluşturulması
4) Çok adımlı zaman serisi tahmini için modellerin oluşturulması ve seçimi
5) Modeler üzerinde hiperparametre ayarının ve denemelerin yapılması
6) Nihai modelin seçilerek 30 dakika sonraki kan şekeri seviyesinin tahmin edilmesi

4.1 Veri Ön İşleme

Veri setindeki bilgilere zaman açısından bakıldığında gerçeklenme anına veya başlangıç-bitiş zamanlarına sahiptir. Dolayısıyla birden çok özniteliği birleştirdiğimizde veride boşluklar oluşmaktadır. Bu boşlukların anlamlı bir şekilde doldurulabilmesi veya en az veri kaybı ile silinebilmesi araştırılmıştır. Verinin eğitimden önce hazırlanma aşamaları açıklanmıştır.

4.1.1 Veri Birleştirme

Her öznitelik için değerlerin ölçüldüğü zaman aralıkları farklı olduğu için farklı tablolarda tutulmaktadır. Bu yüzden kullanılacak özniteliklerin ortak bir tabloda ortak bir zaman aralığı için ayarlanması gerekmektedir. Mesela 540 numaralı hastada, kan şekeri tablosu "2027-05-19 11:36:29" tarihinde başlarken, cilt sıcaklığı tablosu "2027-05-19 09:55:00" tarihinde başlamaktadır. Bu şekilde her öznitelik için birleştirme yapıldığı düşünüldüğünde ortaya boş satırlar çıkmaktadır.

İlk adımda kan şekeri verisi ve diğer sensör verileri en erken başlayan tarihe göre birleştirilmiştir. Çünkü sensör verilerinde saat farkı 12 saati geçmemektedir. İkinci adımda düzenli verilerin üstüne düzensiz veriler eklenirken ise en geç başlayan tarihe göre birleştirilmiştir. Çünkü tablolar arası saat farkının gün bazında olması büyük boşluklar oluşturmaktadır.

4.1.2 Veri Temizliği

Boşlukların doldurulması modelin performansını olumsuz etkilediği koşullarda verinin temizlenmesi gerekmektedir. Verinin en az veri kaybı ile silinebilmesi modelin performansını etkilemektedir. Veri birleştirildikten sonra eğer kan şekeri değeri boş ise tablodaki ilk 60 satır yani 5 saat silinmiştir.

4.1.3 Veri Doldurma Yaklaşımları

Veri 5 dakikaya örneklenerek aynı zaman aralığına hizalandıktan sonra veri noktaları arasındaki eksikliklerden dolayı oluşan boşluklar bulunmaktadır. Bu boşluklarda özniteliğin ifade ettiği anlama göre doldurma yöntemleri araştırılmıştır. Veri doldurma detayları aşağıda verilmiştir.

4.1.3.1 İnterpolasyon

İlk etapta düzenli verilerdeki boşluklar ve "0.0" olma durumları analiz edilmiştir. Ham veride normalde olmaması gereken "0.0" değerleri cihaz ölçüm veya kalibre hatasından kaynaklanmaktadır. Çünkü normal koşullarda sensör verilerinin sıfır olma durumu yoktur. İlk olarak tüm "0.0" değerleri ileri doğru (forward filling) doldurulmuştur. İlk değerlerdeki boşlukları önlemek amacıyla geriye doğru(backward filling) uygulanmıştır. 1 dakikadan 5 dakikaya örneklenen basis_skin_temperature, basis_skin ve acceleration özniteliklerinde 5 örneğin ortalaması alınarak alt örnekleme (downsampling) yapılmıştır.

4.1.3.2 Düzensiz Özniteliklere Özgü Doldurma

Düzensiz özniteliklerden oluşturulan yeni özniteliklerde bazı özel hesaplamalar uygulanarak eksik veri giderilmiştir. Öznitelik oluşturmanın detayları sonraki bölümde verilmiştir. Aşağıda düzensiz özniteliklere uygulanan veri doldurma yöntemleri verilmiştir.

- new_basal : Geçici bazal tablosundaki başlangıç ve bitiş saatlerindeki değerler, bazal tablosundaki zaman ile karşılaştırılmıştır. İlk adımda geçici bazaldaki

bilgiler belirtilen zaman aralığına örneklenmiştir. İkinci adımda, örneklenen veriler ile bazal tablosundaki boşluklar doldurulmuştur.

- basal : İnsülin pompasında bazal insülin kullanımı yeni oran ayarlanana kadar sabit kalmaktadır. Bu sebeple bazal tablosunda yeni bir değer ayarlanana kadar son geçerli değer ile boşluklar doldurulmuştur.
- bolus : Bolus tablosunda 4 farklı bolus tipi bulunmaktadır. "normal" ve "normal dual" tipinde anlık zaman bilgisi vardır. Bu veriler tek değer olarak korunmuştur. "square" ve "square dual" tipinde ise başlangıç ve bitiş zamanına sahip aralıklar bulunmaktadır. Veri bu zaman aralığına değer eşit olarak bölünerek "bolus = bolus/6" işlemi ile örneklenmiştir. Bu öznitelikte "0" değerleri bolus insülin verilmediğini belirttiği için korunmuştur.
- meal_diff : Öğün tablosu düzenli verilerin zaman aralığına göre hizalanmıştır. Hastanın bir önceki (zaman-yediği öğünün zamanı) dakika cinsinden alınarak örneklenmiştir.
- meal_count : Günlük öğün sayısı hesaplanarak zaman aralıklarına örneklenmiştir.

4.1.4 Normalizasyon

Öznitelikler arasında farklı ölçekte veriler mevcut olduğu için verilerin aynı değerlerde ölçeklenmesi gerekmektedir. Veriyi modele beslemeden önce modelin amacımız doğrultusunda çalışabilmesi için bu adım önemlidir. Mesela; kan şekeri değerlerinde 40-400 aralığında ölçümler mevcutken, galvanik cilt tepkimesi değeri 0-1 aralığındadır. Bu durumda verilerin arasındaki oran bozulmadan aynı ölçeğe çevrilmelidir. Çalışma kapsamında normalizasyon olmadan, Minimum-Maksimum Normalizasyonu ve Standardizasyon ile modellerin performansı değerlendirilmiştir.

4.1.4.1 Minimum-Maksimum Normalizasyonu

Minimum-Maksimum Normalizasyon yöntemi verileri normalleştirmek için literatürde yaygın olarak kullanılmaktadır. Veri setindeki değerleri en düşük ve en yüksek değerler arasında bir aralığa dönüştürülmektedir. Böylece değerler ölçeklenerek daha karşılaştırılabilir hale gelmektedir. Aşağıda her özellik için sütun bazında veriye uygulanan formül verilmiştir.

$$x^{'} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4.1}$$

4.1.4.2 Standardizasyon

Standardizasyon, literatürde Z-Skor Normalizasyonu olarak da adlandırılmaktadır. Veri setindeki değerlerin ortalama ve standart sapmaya göre ölçeklendirilmesini sağlamaktadır. Aşağıda veriye sütun bazında uygulanan formül verilmiştir.

$$z = \frac{x - \mu}{\sigma} \tag{4.2}$$

4.2 Özniteliklerin Oluşturulması

Veri seti analizinden edinilen bilgilerin doğrultusunda modellerde denenmek üzere öznitelikler seçilmiştir ve yeni öznitelikler oluşturulmuştur.

4.2.1 Öznitelik Seçimi

Öznitelik seçiminde referans noktası olarak sadece kan seviyesi özniteliğinin kullanımı ile elde edilen sonuçlar alınmıştır. Ardından düzenli ve düzensiz öznitelikler eklenerek modellerin eğitim süreçleri tekrarlanmıştır. Modelin validasyon ve eğitim kayıp fonksiyonundaki ve test kümesindeki hata oranları gözlenmiştir. Modelin hatasını azaltan öznitelikler çalışmaya dahil edilmiştir.

4.2.2 Yeni Öznitelik Çıkarımı

Yeni özniteliklerin oluşturulabilmesi için düzensiz bilgiler üzerinde araştırma yapılmıştır. Aşağıda yeni oluşturulan öznitelikler açıklanmıştır.

- new_basal : "basal" ve "temp_basal" bilgilerinden elde edilmiştir. İki bilgi de bazal insülin kullanımı ile ilgili bilgiler içermektedir ancak ayrı etiketlerde verilmiştir. Geçici bazal tablosunda, bazal insülin akışının askıya alındığı veya ilave insülin kullanımının zaman aralıkları belirtilmiştir. Bazal insülinde ise yaklaşık 6 saatte bir ayarlanan bazal insülin oranı verilmiştir. Bu zaman aralıklarındaki değerler, bazal tablosundaki değerlerle birleştirilerek öznitelik oluşturulmuştur.
- meal_diff : Öğün tablosundan üretilen yeni bir özniteliktir. Düzenli özniteliklerin oluşturduğu ortak zaman çizelgesine göre zamanlar alınarak öğün tablosunda hastanın en son ne zaman yemek yediğiyle eşleştirilmiştir. Mesela zaman 19:50 ise ve hasta en son 18:30'da yemek yediyse meal_diff değeri 80 olmaktadır. Öğün farkını dakika cinsinden tutmaktadır.

- meal_count : Öğün tablosunda veriler güne göre filtrelenmiştir. Örneklenen zaman hangi güne aitse o değer yazılmıştır. Mesela 12 Temmuz gününde hasta 4 öğün yemek yediyse o günkü tüm zaman örneklemelerinde meal_count değeri 4 olmaktadır. Burada hastanın öğün girişi yapmadığı günler "0" olarak girilmiştir.

4.3 Verinin Ayrılması

Veri ön işleme adımları bittikten sonra verinin eğitime hazırlanması gerekmektedir. Veri tüm modellerde %70 eğitim kümesi, %20 validasyon kümesi, %10 test kümesi olarak ayrılmıştır. Düzenli ve düzensiz özniteliklerin birlikte kullanıldığı model için, eğitim ve test için ayrılan verilerin hasta bazında analizi Tablo 4.1 'de verilmiştir. Derin öğrenme modellerinin eğitiminde ne kadar çok veri görürse modelin başarısı doğru orantılı olarak artmaktadır. Burada en az veri noktasına sahip hastanın 552 ve en çok veri noktasına sahip hastanın 596 olduğunu görmekteyiz.

Tablo 4.1 Hasta Bazında Veri Noktası Analizi

Hasta No.	Eğitim	Validasyon	Test
540	8491	2426	1213
544	8682	2481	1241
552	6953	1987	994
559	8216	2348	1174
563	8952	2558	1279
567	8777	2508	1254
570	8012	2289	1145
575	9033	2581	1291
584	8732	2495	1248
588	9079	2594	1298
591	8799	2514	1257
596	9852	2815	1408

4.4 Değerlendirme Metrikleri

Çalışma kapsamında oluşturulan modellerin performanslarının ölçülmesi için üç değerlendirme metriğine odaklanılmıştır. Ortalama Hatanın Karekökü ve Ortalama Mutlak Hata metrikleri kan şekeri seviyesi tahmininde sonuçları değerlendirmede yaygın olarak kullanılmaktadır.

4.4.1 Ortalama Hatanın Karekökü (RMSE)

Gerçek değer ile tahmin edilen değer arasındaki farkın karesini alarak ortalama hatayı hesaplamaktadır.

$$\sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2} \tag{4.3}$$

4.4.2 Ortalama Mutlak Hata (MAE)

Gerçek değer ile tahmin edilen değer arasındaki farkın mutlak değerine bakarak ortalama hatayı hesaplamaktadır.

$$\frac{1}{n}\sum_{j=1}^{n}\left|y_j - \hat{y}_j\right| \tag{4.4}$$

4.4.3 Metriklerin Puanlandırılması

Genel durumu ölçebilmek için hasta bazında en iyi RMSE ve en iyi MAE sonuçları kendi içinde puanlandırılmıştır. Her hasta için en iyi modele 1 puan eklenerek sonuçlar ortaya konmuştur. Bu sistem yöntem bazında karşılaştırma için kullanılmıştır.

———————— DEVAMI HAZIRLANIYOR ————————————

————————

————————-

————————-

————————

4.5 Kullanılan Yöntemler

Çalışma kapsamında klasik yöntemlerden ARIMA ve derin öğrenme yöntemlerinden RNN, LSTM, GRU yöntemleriyle her hasta için kişiselleştirilmiş modeller oluşturulmuştur. Modellerin detayları aşağıda açıklanmıştır.

4.5.1 Klasik Yöntemler

ARIMA-VARIMA ???
Random Forest ??

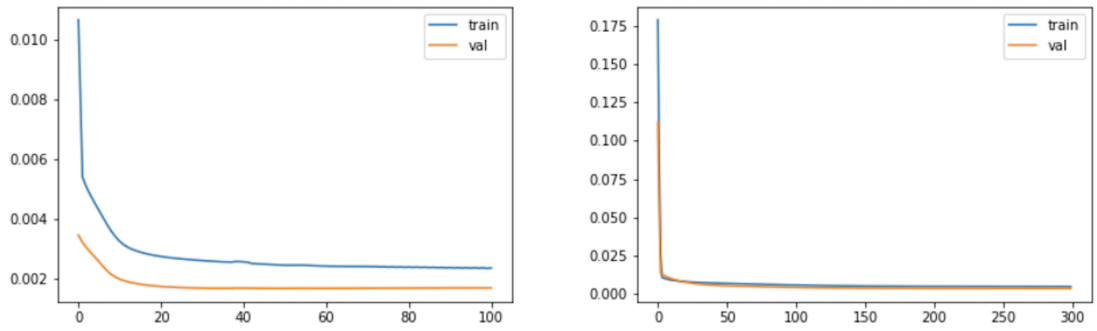
4.5.2 Derin Öğrenme Yöntemleri

Derin Öğrenme yöntemlerinden RNN, LSTM ve GRU ağları kullanılmıştır. Yapılan denemelerde veri ön işleme adımları, model parametreleri ortak kullanılmıştır. Parametrelere karar verilirken ağ üzerinde tekrarlı denemeler yapılmıştır. Mesela Unit

RMSPro [30]
ADAM optimizer [31]
Relu acitivation function [32]

4.5.2.1 Modellerde Parametre Seçimi

RNN, GRU ve LSTM ağlarında ağdaki birim (nöron) sayısına karar vermek için farklı sayıda denemler yapılmıştır. Şekil 4.1 'de 50 nöronlu bir ağın eğitim süreci verilmiştir. Nihai modelde öznitelik sayısı kadar nörondan oluşturulmuştur.



Şekil 4.1 Birim Sayısının Modele Etkisi

Tablo 4.2 Model Parametreleri

Parametreler	**Değerler**
Birim Sayısı	Öznitelik sayısı
Tur	300
Optimizasyon Algoritması	ADAM
Normalizasyon	Min-Maks
Batch Size	128
Geçmiş Zaman Adımı	6
Tekrarlama Birimi Aktivasyon Fonksiyonu	ReLU

5
DENEYSEL SONUÇLAR

Bu bölümde hasta bazında oluşturulan kişiselleştirilmiş modellerin performansları ölçülmüştür. Performanslar hasta, kohort ve yöntem bazında ölçülmüştür. En iyi ve en kötü performansa sahip hasta ve model ortaya konmuştur. Modellerin performansları yöntemler, öznitelikler ve hastaların etkileri araştırılarak çalışmanın sonuçları elde edilmiştir. Oluşturulan modeller ile yapılan testlerde tüm modeller 10 kere çalıştırılarak sonuçlar elde edilmiştir. Sonuçların ortalaması ve standart sapmaları değerlendirme metrikleri ile birlikte verilmiştir.

5.1 Temel Modelin Değerlendirilmesi

Oluşturulan temel modelde hastaya ait kan şekeri değerleri kullanılarak 30 dakika sonraki kan şekeri seviyesi tahmin edilmiştir. Bu model ile tek değişkenli zaman serisinde tahmin yapılmıştır. Aşağıda temel modele ait sonuçlar değerlendirilmiştir.

5.1.1 Hasta Bazında Sonuçlar

Tablo 5.1 'de 12 hastaya ait 4 yöntemin sonuçları 10 deneyin sonucundaki ortalama ve standart sapması ile birlikte verilmiştir.

5.1.2 Kohort Bazında Sonuçlar

Tablo 5.2 'de 4 yöntemin kohort bazında sonuçları verilmiştir. Kohort bazında sonuçlar

5.1.3 Yöntem Bazında Sonuçlar

Tablo 5.3 'de 4 yöntemin puanlandırılmasının sonuçları verilmiştir. Bu sonuçlara baktığımız zaman

Tablo 5.1 Hasta Bazında Sonuçlar

Hasta No.		ARIMA		RNN		LSTM		GRU	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
540	AVG			28.98	37.91	32.66	42.33	38.27	49.01
	SD			14.87	17.89	14.21	17.09	12.63	15.35
544	AVG			36.56	43.62	36.91	43.91	39.95	47.09
	SD			12.65	13.04	11.95	12.46	8.96	9.27
552	AVG								
	SD								
559	AVG								
	SD								
563	AVG								
	SD								
567	AVG								
	SD								
570	AVG								
	SD								
575	AVG								
	SD								
584	AVG								
	SD								
588	AVG								
	SD								
591	AVG								
	SD								
596	AVG								
	SD								

Tablo 5.2 Kohort Bazında Sonuçlar

Kohort		ARIMA		RNN		LSTM		GRU	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
2018	AVG								
	SD								
2020	AVG								
	SD								

Tablo 5.3 Yöntem Bazında Puanlandırma

Yöntem	MAE Puanı	RMSE Puanı
ARIMA		
RNN	4	4
LSTM	4	4
GRU	4	4

5.2 Düzenli Özelliklerden Oluşan Modelin Değerlendirilmesi

Bu modelde hastaya ait düzenli verilerden oluşturulan öznitelikler kullanılarak 30 dakika sonraki kan şekeri tahmin edilmiştir.

5.2.1 Hasta Bazında Sonuçlar

Tablo 5.4 'de 12 hastaya ait 4 yöntemin sonuçları verilmiştir.

Tablo 5.4 Hasta Bazında Sonuçlar

Hasta No.		**ARIMA**		**RNN**		**LSTM**		**GRU**	
		MAE	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**
540	**AVG**			28.98	37.91	32.66	42.33	38.27	49.01
	SD			14.87	17.89	14.21	17.09	12.63	15.35
544	**AVG**			36.56	43.62	36.91	43.91	39.95	47.09
	SD			12.65	13.04	11.95	12.46	8.96	9.27
552	**AVG**								
	SD								
559	**AVG**								
	SD								
563	**AVG**								
	SD								
567	**AVG**								
	SD								
570	**AVG**								
	SD								
575	**AVG**								
	SD								
584	**AVG**								
	SD								
588	**AVG**								
	SD								
591	**AVG**								
	SD								
596	**AVG**								
	SD								

5.2.2 Kohort Bazında Sonuçlar

Tablo 5.5 'de 4 yöntemin kohort bazında sonuçları verilmiştir. Bu sonuçlara baktığımız zaman

Tablo 5.5 Kohort Bazında Sonuçlar

Kohort		ARIMA		RNN		LSTM		GRU	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
2018	AVG								
	SD								
2020	AVG								
	SD								

5.2.3 Yöntem Bazında Sonuçlar

Tablo 5.6 'de 4 yöntemin puanlandırılmasının sonuçları verilmiştir. Bu sonuçlara baktığımız zaman

Tablo 5.6 Yöntem Bazında Puanlandırma

Yöntem	MAE Puanı	RMSE Puanı
ARIMA		
RNN	4	4
LSTM	4	4
GRU	4	4

5.3 Düzensiz ve Düzenli Özelliklerden Oluşan Final Modelinin Değerlendirilmesi

Oluşturulan modelde hastaya ait sensör veri değerleri kullanılarak 30 dakika sonraki kan şekeri tahmin edilmiştir.

5.3.1 Hasta Bazında Sonuçlar

12 hasta ve 3 model için rmse

5.3.2 Yöntem Bazında Sonuçlar

3 model için puanlandırma

5.3.3 Hasta Bazında Sonuç Grafikleri

12 hasta ve 3 model için mae

5.4 Düzensiz ve Düzenli Özelliklerden Oluşan Modelin Değerlendirilmesi

Oluşturulan modelde hastaya ait sensör veri değerleri kullanılarak 30 dakika sonraki kan şekeri tahmin edilmiştir.

5.4.1 Hasta Bazında RMSE Sonuçları

12 hasta ve 3 model için rmse

5.4.2 Hasta Bazında MAE Sonuçları

12 hasta ve 3 model için mae

5.4.3 Yöntem Bazında Sonuçlar

3 model için puanlandırma

5.5 Hasta Bazında Performansların Değerlendirilmesi

En iyi ve en kötü sonuçlara sahip hastalar için yorum yap

5.6 Yöntemlerin Performanslarının Karşılaştırılması

her model için 3 yöntemde mae rmse doğruluk değerlerini karşılaştır

6
SONUÇ VE ÖNERİLER

Once you have opened this page you will either have to log in if you are already registered, or register if this is your first visit to this page. When you first register, an e-mail will be automatically sent to you confirming your registration and giving you your password. Once you have logged in you will be able to access the registration form, and you should enter all their details.

Please note that when you have entered your university code and the name of your department, only information which is relevant to your university and department, and which is covered by the relevant bilateral agreement will be displayed.

When you have completed the form, you will find at the bottom of the page a button called 'Attach Files'. This provides a link to where you can attach your completed university application form and Learning Agreement. If you have any special conditions or modifications to the Learning Agreement, you have to contact with your relevant departmental coordinator.

When you are finished, click 'Save' and then 'Send', in order the forms to be sent to this office. Once we have your completed forms we can pass them on to your coordinator here in Yıldız Technical University. Please note that The EU Office cannot make any changes to their form once it has been submitted. Once an understanding has been reached by the coordinators about their application, a Letter of Invitation will be prepared and sent to you.

KAYNAKÇA

[1] *World health organization*. [Online]. Available: https://www.who.int/health-topics/diabetes#tab=tab_1 (visited on 06/02/2022).

[2] A. D. Association, "2. classification and diagnosis of diabetes," *Diabetes care*, vol. 40, no. Supplement_1, S11–S24, 2017.

[3] *Türkiye sağlık enstitüleri başkanlığı*. [Online]. Available: https://www.tuseb.gov.tr/tuhke/uploads/genel/files/dunya_diyabet_gunu_2021_son.pdf.

[4] I. Federation, "Idf diabetes atlas, tenth," *International Diabetes*, 2021.

[5] T. Bremer, D. A. Gough, "Is blood glucose predictable from previous values? a solicitation for data.," *Diabetes*, vol. 48, no. 3, pp. 445–451, 1999.

[6] A. Zaitcev, M. R. Eissa, Z. Hui, T. Good, J. Elliott, M. Benaissa, "A deep neural network application for improved prediction of HbA_{1C} in type 1 diabetes," *IEEE journal of biomedical and health informatics*, vol. 24, no. 10, pp. 2932–2941, 2020.

[7] K. Li, C. Liu, T. Zhu, P. Herrero, P. Georgiou, "Glunet: A deep learning framework for accurate glucose forecasting," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 414–423, 2019.

[8] A. Aliberti *et al.*, "A multi-patient data-driven approach to blood glucose prediction," *IEEE Access*, vol. 7, pp. 69 311–69 325, 2019.

[9] J. Daniels, P. Herrero, P. Georgiou, "A multitask learning approach to personalized blood glucose prediction," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 436–445, 2021.

[10] X. Yu *et al.*, "Deep transfer learning: A novel glucose prediction framework for new subjects with type 2 diabetes," *Complex & Intelligent Systems*, vol. 8, no. 3, pp. 1875–1887, 2022.

[11] K. Li, J. Daniels, C. Liu, P. Herrero, P. Georgiou, "Convolutional recurrent neural networks for glucose prediction," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 603–613, 2019.

[12] *Preface the 5th ınternational workshop on knowledge discovery in healthcare data*. [Online]. Available: https://sites.google.com/view/kdh-2020/bglp-challenge?authuser=0 (visited on 09/10/2022).

[13] M. Mayo, T. Koutny, "Neural multi-class classification approach to blood glucose level forecasting with prediction uncertainty visualisation," in *KDH 2020*, CEUR Workshop Proceedings, vol. 2675, 2020, pp. 80–84.

[14] R. Bevan, F. Coenen, "Experiments in non-personalized future blood glucose level prediction," in *CEUR Workshop Proceedings*, vol. 2675, 2020, pp. 100–104.

[15] H. Hameed, S. Kleinberg, "Investigating potentials and pitfalls of knowledge distillation across datasets for blood glucose forecasting," in *Proceedings of the 5th Annual Workshop on Knowledge Discovery in Healthcare Data*, 2020.

[16] G. Cappon *et al.*, "A personalized and ınterpretable deep learning based approach to predict blood glucose concentration in type 1 diabetes.," in *KDH@ ECAI*, 2020, pp. 75–79.

[17] A. Bhimireddy, P. Sinha, B. Oluwalade, J. W. Gichoya, S. Purkayastha, "Blood glucose level prediction as time-series modeling using sequence-to-sequence neural networks," CEUR Workshop Proceedings, 2020.

[18] J. Daniels, P. Herrero, P. Georgiou, "Personalised glucose prediction via deep multitask networks.," in *KDH@ ECAI*, 2020, pp. 110–114.

[19] J. Freiburghaus, A. Rizzotti, F. Albertetti, "A deep learning approach for blood glucose prediction of type 1 diabetes," in *Proceedings of the Proceedings of the 5th International Workshop on Knowledge Discovery in Healthcare Data co-located with 24th European Conference on Artificial Intelligence (ECAI 2020), 29-30 August 2020, Santiago de Compostela, Spain*, 29-30 August 2020, 2020.

[20] H. Nemat, H. Khadem, J. Elliott, M. Benaissa, "Data fusion of activity and cgm for predicting blood glucose levels," in *Knowledge Discovery in Healthcare Data 2020*, CEUR Workshop Proceedings, vol. 2675, 2020, pp. 120–124.

[21] J. Pavan *et al.*, "Personalized machine learning algorithm based on shallow network and error ımputation module for an ımproved blood glucose prediction.," in *KDH@ ECAI*, 2020, pp. 95–99.

[22] I. AKGÜL, "Zaman serİsİ analİzİ ve öngörü modellerİ," *Öneri Dergisi*, vol. 1, no. 1, pp. 52–69, 1994.

[23] *Tensorflow core*. [Online]. Available: `https://www.tensorflow.org/tutorials/structured_data/time_series` (visited on 10/07/2022).

[24] G. Udny Yule, "On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers," *Philosophical Transactions of the Royal Society of London Series A*, vol. 226, pp. 267–298, 1927.

[25] P. Whittle, *Hypothesis testing in time series analysis*. Almqvist & Wiksells boktr., 1951, vol. 4.

[26] D. E. Rumelhart, G. E. Hinton, R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[27] S. Hochreiter, J. Schmidhuber, *Long short-term memory. j neural comput. 1997; 9 (8): 1735–80*.

[28] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[29] C. Marling, R. Bunescu, "The ohiot1dm dataset for blood glucose level prediction: Update 2020," in *CEUR workshop proceedings*, NIH Public Access, vol. 2675, 2020, p. 71.

[30] T. Tieleman, G. Hinton, *et al.*, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.

[31] D. P. Kingma, J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[32] V. Nair, G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Icml*, 2010.

TEZDEN ÜRETİLMİŞ YAYINLAR

İletişim Bilgileri: gizem.celik@yildiz.edu.tr

Konferans Bildirisi

1. M. G. Celik and S. Varli, "Deep Learning Approaches for Type-1 Diabetes: Blood Glucose Prediction," 2022 7th International Conference on Computer Science and Engineering (UBMK), 2022, pp. 1-5, doi: 10.1109/UBMK55850.2022.9919446.