

T.C.
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI



YAPAY ZEKA TABANLI NON-INVAZIV KAN ŞEKERİ
ÖLÇÜMÜ

YÜKSEK LİSANS TEZİ

GÖKHAN ADIGÜZEL

TEZ DANIŞMANI
Dr. Öğr. Üyesi Ümit Şentürk
İKİNCİ TEZ DANIŞMANI
Prof. Dr. Kemal Polat

BOLU, TEMMUZ - 2024

KABUL VE ONAY SAYFASI

Gökhan ADIGÜZEL tarafından hazırlanan “**YAPAY ZEKA TABANLI NON-INVAZIV KAN ŞEKERİ ÖLÇÜMÜ**” adlı tez çalışması jürimiz tarafından Bilgisayar Mühendisliği Anabilim Dalı’nda Yüksek Lisans Tezi olarak oy birliği ile kabul edilmiştir. 29/07/2024

Jüri Üyeleri

İmza

Danışman
Dr. Öğr. Üyesi Ümit Şentürk
Bolu Abant İzzet Baysal Üniversitesi

.....

Üye
Doç Dr. Uğurhan KUTBAY
Gazi Üniversitesi

.....

Üye
Doç Dr. Ali YILMAZ
Bolu Abant İzzet Baysal Üniversitesi

.....

Lisansüstü Eğitim Enstitüsü Onayı

Prof. Dr. İbrahim KÜRTÜL
Lisansüstü Eğitim Enstitüsü Müdürü

ETİK BEYAN

Bolu Abant İzzet Baysal Üniversitesi, Lisansüstü Eğitim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu bildirir,

aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Teze ilişkin Turnitin adlı programında enstitü müdürlüğünce belirlenen filtrelemeler uygulanarak alınmış olan benzerlik raporuna göre, tezin benzerlik oranı %30'u geçmemektedir.

Bu çalışma için Bolu Abant İzzet Baysal Üniversitesi Klinik Araştırmalar Etik Kurulundan 448 sayısı ile etik izin alınmıştır.

.....

GÖKHAN ADIGÜZEL

ÖZET

**YAPAY ZEKA TABANLI NON-INVAZIV KAN ŞEKERİ ÖLÇÜMÜ
YÜKSEK LİSANS TEZİ
GÖKHAN ADIGÜZEL
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
(TEZ DANIŞMANI: DR. OGR. ÜYESİ ÜMİT ŞENTÜRK)
(İKİNCİ DANIŞMAN: PROF. DR. KEMAL POLAT)**

**BOLU, TEMMUZ - 2024
XI + 44**

Kan şekeri seviyelerinin tahmini, diyabetin etkili yönetiminde kritik bir görevdir. Çalışma, Fotopletismografi (PPG) sinyallerini kullanarak kan şekeri seviyelerini tahmin etmek için CatBoost, XGBoost ve ekstra ağaç regresör gibi makine öğrenimi modellerinin gücünden, SHAP değerleri ve karışıklık matrisi gibi açıklanabilir yapay zeka teknikleriyle birlikte yararlanmaya odaklanıyor. Bu araştırmada oluşturulan veri seti, PPG sinyallerinden glikoz tahmini için dikkatlice seçilmiştir. Etik kurul onayı alınarak çalışma süresi boyunca Abant İzzet Baysal Üniversitesi Dahiliye bölümüne diyabet şikâyeti için gelen 50 kişiden alınan verilerden oluşmaktadır. Her bireyin bilgileri, laboratuvar glikoz ölçümlerini ve yaklaşık bir dakikalık kaydedilen parmak PPG sinyallerini içerir. Test edilen çeşitli makine öğrenimi modelleri arasında CatBoost, kan şekeri seviyelerini tahmin etmede en iyi performansı gösteren model olarak ortaya çıktı. CatBoost modeli, 0.7191'lik etkileyici bir determinasyon katsayısı (R^2) metriğine ve 25.21'lik ortalama mutlak hataya (MAE) ulaşarak glikoz seviyesi tahminlerindeki verimliliğini ve doğruluğunu ortaya koydu. Özellik önemi analizi, CatBoost ile oluşturulan tahmin modelinde medyan fark ve basıklık gibi belirli özelliklerin önemini vurgulayarak bunların kan şekeri seviyelerinin belirlenmesindeki önemli rolünün altını çizdi. Açıklanabilir yapay zeka tekniklerinin dahil edilmesi, tahmine dayalı modellerin yorumlanabilirliğini ve şeffaflığını arttırdı. SHAP değerlerini ve karışıklık matrisini kullanarak çalışma, tahminleri etkileyen faktörlere ilişkin değerli bilgiler sunarak modellerin karar verme sürecinin daha derinlemesine anlaşılmasını kolaylaştırdı. Sonuç olarak bu araştırma, PPG sinyallerinden kan şekeri seviyelerinin tahmin edilmesinde makine öğrenimine dayalı yaklaşımların potansiyelini vurgulamaktadır. CatBoost gibi gelişmiş modellerden yararlanan ve açıklanabilir yapay zeka yöntemlerini kullanan bu çalışma, doğru, invaziv olmayan ve veriye dayalı tahmine dayalı metodolojiler yoluyla gelişmiş diyabet yönetiminin yolunu açıyor.

ANAHTAR KELİMELER: Kan Şekeri Tahmini, Fotopletismografi, Makine Öğrenmesi, SHAP, Açıklanabilir Yapay Zeka

ABSTRACT

ARTİFİCİAL INTELLIGENCE-BASED NON-INVASİVE BLOOD GLUCOSE MEASUREMENT

MSC THESIS

GÖKHAN ADIGÜZEL

BOLU ABANT İZZET BAYSAL UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF COMPUTER ENGINEERING

SUPERVISOR: DR. OGR. ÜYESİ ÜMİT ŞENTÜRK)

CO-SUPERVISOR: PROF. DR. KEMAL POLAT)

BOLU, JULY 2024

XI + 44

Prediction of blood sugar levels is a critical task in the effective management of diabetes. The study focuses on leveraging the power of machine learning models such as CatBoost, the dataset created in this research was carefully selected for glucose estimation from PPG signals. It consists of data obtained from 50 people who came to Abant İzzet Baysal University Internal Medicine Department with diabetes complaints during the study period, with ethics committee approval. Information for each individual includes laboratory glucose measurements and approximately one minute of recorded finger PPG signals. Among the various machine learning models tested, CatBoost emerged as the best-performing model for predicting blood sugar levels. The CatBoost model demonstrated its efficiency and accuracy in glucose level predictions by achieving an impressive coefficient of determination (R^2) metric of 0.7191 and a mean absolute error (MAE) of 25.21. Feature importance analysis highlighted the importance of certain features, such as median difference and kurtosis, in the prediction model built with CatBoost, underlining their important role in determining blood glucose levels. The inclusion of explainable AI techniques increased the interpretability and transparency of predictive models. Using SHAP values and the confusion matrix, the study facilitated a deeper understanding of the models' decision-making process by providing valuable insight into the factors influencing the predictions. In conclusion, this research highlights the potential of machine learning-based approaches in predicting blood glucose levels from PPG signals. Leveraging advanced models such as CatBoost and employing explainable artificial intelligence methods, this study paves the way for improved diabetes management through accurate, non-invasive, and data-driven predictive methodologies.

KEYWORDS: Blood Sugar Prediction, Photoplethysmography, Machine Learning, SHAP, XAI

İÇİNDEKİLER

Sayfa

KABUL VE ONAY SAYFASI	iii
ETİK BEYAN	iv
ÖZET	v
ABSTRACT	vi
İÇİNDEKİLER	vii
ŞEKİL LİSTESİ	viii
TABLO LİSTESİ	ix
KISALTMA VE SEMBOLLER LİSTESİ	x
TEŞEKKÜR	xi
1. GİRİŞ	1
1.1 Araştırmanın Amacı	9
1.2 Araştırmanın Sınırlılıkları	10
2. LİTERATÜR ÖZETLERİ	11
3. MATERYAL VE YÖNTEM	14
3.1 Veri Setinin Oluşturulması	15
3.2 Veri Seti Kümesi	17
3.3 Veri Ön İşleme	17
3.4 Segmentasyon (Bölütleme)	18
3.5 Öznitelik Çıkarma ve Seçme İşlemi	19
3.6 Makine Öğrenmesi Modeli	21
3.7 Açıklanabilir Yapay Zeka (XAI-SHAP)	24
3.8 Kan Glukoz Seviyesi Tahmini	28
4. BULGULAR	29
5. TARTIŞMA	34
6. SONUÇ VE ÖNERİLER	36
7. KAYNAKLAR	39
8. EKLER	44

ŞEKİL LİSTESİ

Sayfa

Şekil 1.1. İnvaziv kan şekeri ölçüm cihazı ve ekipmanları.....	4
Şekil 1.2. Evde kan şekeri ölçüm aşamaları.....	4
Şekil 1.3. Veri okuma sisteminin parmak üzerinde gösterimi	7
Şekil 1.4. PPG sensörü ölçüm sinyali	7
Şekil 1.5. Veri okuma sisteminin parmak üzerinde gösterimi	8
Şekil 3.1. Çalışma yöntem adımları	14
Şekil 3.2. MAX305 sensörü iç yapısı.....	15
Şekil 3.3. Sensör ve mikrodenetleyici bağlantı şeması	16
Şekil 3.4. Temizlenen sinyal ve ham sinyal	17
Şekil 3.5. Bölütleme işlemi adımları.....	18
Şekil 3.6. Üç numaralı hastanın kalp atışı ortalaması	19
Şekil 3.7. CatBoostRegresör algoritması feature importance plot grafiği	25
Şekil 3.8. CatBoostRegresör-LGBM feature importance plot karşılaştırması..	26
Şekil 3.9. Kolerasyon saptanan öznelilikler	28
Şekil 4.1. CatBoostRegresör Modeli R^2 Prediction Error Grafiği	31
Şekil 4.2. LGBM Regresör Modeli R^2 Prediction Error Grafiği.....	33

TABLO LİSTESİ

Sayfa

Tablo 1.1. Kan şekeri belirleme ölçütleri.....	1
Tablo 1.2. Kan şekeri ölçüm yöntemleri.....	6
Tablo 3.1. Kullanılacak özneliklerden bazıları.....	19
Tablo 3.2. İlk yedi regresyon modeli performans değerleri.....	22
Tablo 4.1. Regresyon modellerine göre performans metrikleri	29
Tablo 5.1. Literatürdeki çalışmalar	35



KISALTMA VE SEMBOLLER LİSTESİ

WHO	: Dünya Sağlık Örgütü
RNN	: Tekrarlayan Sinir Ağı
AR	: Otoregresif
T2D	: Tip 2 Diyabet
SVR	: Destek Vektör Regresyonu
CGM	: Sürekli Glukoz İzleme
LTSM	: Uzun Kısa Süreli Bellek
BLSTM	: İki Yönlü Uzun Kısa Süreli Bellek
DNN	: Derin Sinir Ağı
FNN	: İleri Beslemeli Sinir Ağı
XGBoost	: Aşırı Gradyan Artırma
CNN	: Evrişimli Sinir Ağları
ARX	: Otoregresyon eksojen
PPG	: Fotopletismografi
RF	: Rastgele Orman
ADC	: Analog Dijital Çevirici
LightGBM	: Hafif Gradyan Artırma Makinesi
MAE	: Ortalama Mutlak Hata
MSE	: Ortalama Kare Hatası
RSME	: Kök Ortalama Kare Hatası
R²	: R Kare
RMSLE	: Kök Ortalama Kare Logaritmik Hata
MAPE	: Ortalama Mutlak Yüzde Hata
NN	: Sinir Ağları
BN	: Bayes Ağları
SVM	: Destek Vektör Makineleri
IOT	: Nesnelerin İnterneti

TEŞEKKÜR

Bu tez çalışması boyunca, ilgi ve yardımlarını esirgemeyen danışmanım Dr. Öğr. Üyesi Ümit ŞENTÜRK'e, çalışmam boyunca desteklerini esirgemeyen Prof. Dr. Kemal POLAT'a çalışmam boyunca beni özveri ve sabırla destekleyen eşim Nurşen ADIGÜZEL'e, teşekkürlerimi ve şükranlarımı sunarım.



1. GİRİŞ

Diyabet, vücudun kan şekeri (glikoz) seviyesini düzenleyemediği kronik bir hastalıktır. Glikoz, hücrelerin enerji kaynağıdır ve kandaki seviyeleri insülin hormonu tarafından düzenlenir. Diyabet hastalığında, vücut ya yeterli insülin üretmez ya da üretilen insülini etkin bir şekilde kullanamaz. Bu durum, kan şekeri seviyelerinin sürekli yüksek kalmasına neden olur ve zamanla ciddi sağlık komplikasyonlarına yol açabilir.

Diyabet, birey ve toplumu ilgilendiren ciddi bir sağlık sorunudur. Ayrıca hızla artış göstermesinden dolayı küresel bir hastalık olarak kabul edilmektedir [1]. Diyabet, Dünya Sağlık Örgütü (WHO) tarafından “tüm dünyayı etkisi altına alan majör bir sağlık problemi, yeni bir salgın hastalık” olarak ilan edilmiştir [2]. Dünyada bütün yaş gruplarında diyabet prevalansı 2000 yılında %2,8 iken (yaklaşık 171 milyon) 2030 yılında %4,4'e (yaklaşık 366 milyon) ulaşması beklenmektedir [3]. Diyabet hastalığında pankreas yeteri kadar insülin üretemez veya hücre ve dokular üretilen insülini etkin şekilde kullanamaz[4]. Sağlıklı bireylerde kandaki glikoz miktarı 70 ile 99 mg/dL düzeyindedir, 100- 125 mg/dL glikoz konsantrasyonuna sahip bireyler ise prediyabetik olarak değerlendirilir.[5] Aşağıda Tablo 1.1. 'de kandaki şeker seviyelerinin diyabet hastalığını belirlemekte kullanılan ölçütleri paylaşılmıştır [5].

Tablo 1.1. Kan şekeri belirleme ölçütleri

Durum	Açlık Kan Şekeri (mg/dL)	Oral Glukoz Tolerans Testi (OGTT) 2 saat sonra (mg/dL)	Hemoglobin A1c (HbA1c) (%)
Normal	70-99	<140	< 5.7
Prediyabet	100-125	140-199	5.7-6.4
Diyabet	≥ 126	≥ 200	≥ 6.5

Açlık Kan Şekeri(FBS): En az 8 saat açlıktan sonra ölçülen kan şekeri düzeyi.[68]

Oral Glikoz Tolerans Testi (OGTT): 75 gram glikoz tüketildikten 2 saat sonra ölçülen kan şekeri seviyesi.[69]

Hemoglobin A1c (HbA1c): Son 2-3 aydaki ortalama kan şekeri seviyesini gösteren bir kan testi.[69]

İnsülin, pankreas tarafından üretilir ve glikozun kandan hücrelere geçişini sağlar. Bu süreç, enerji üretimi için önemlidir. İnsülin eksikliği veya etkisizliği, hücrelerin yeterince glikoz almasını engeller ve bu da kanda aşırı glikoz birikimine yol açar. Kandaki glikoz seviyesinin normal aralığın üzerinde kalması durumuna hiperglisemi denir. Diyabetin gelişimiyle ilgili iki ana mekanizma vardır. Bunlardan birincisi pankreasın yeterli insülin üretmemesidir. Bu durum genellikle otoimmün reaksiyonlar sonucunda ortaya çıkar. İkincisi ise hücrelerin insüline karşı direnç göstermesi, yani insülini etkin bir şekilde kullanamamasıdır. Bu durum, genellikle obezite ve sedanter yaşam tarzı ile ilişkilidir[70].

Diyabetin etkin yönetimi, kan şekeri seviyelerini kontrol altında tutmayı içerir. Bu, aşağıdaki önlemleri gerektirir:

Diyet ve Beslenme: Sağlıklı ve dengeli bir beslenme, kan şekeri kontrolünü destekler.

Egzersiz: Düzenli fiziksel aktivite, insülin duyarlılığını artırarak kan şekeri seviyelerini düşürür.

İlaç Tedavisi: Oral ilaçlar veya insülin tedavisi, diyabet yönetiminde önemli bir rol oynar.

Kan Şekeri İzleme: Düzenli kan şekeri ölçümleri, hastaların seviyelerini kontrol etmelerine yardımcı olur.

Diyabet, yaşam boyu süren bir hastalık olmasına rağmen, uygun yönetim ve yaşam tarzı değişiklikleri ile kontrol altında tutulabilir. Hastalar, düzenli tıbbi kontroller ve destekleyici eğitimle komplikasyon riskini azaltabilir ve yaşam kalitesini artırabilirler. Bu çalışma diyabet kontrolünde kan şekeri seviyelerinin kontrol altında tutulmasına odaklanmaktadır. Hem hastalığın teşhisi hem de takibi kan şekeri seviyelerinin bilinmesi ile mümkündür.

Günümüzde, kan şekeri seviyesinin takibi için invaziv (girişimsel) yöntemler kullanılmaktadır. Ancak, bu yöntemler hasta için rahatsız edici olabilir ve bazen de zaman alabilir. Bu nedenle, non-invaziv kan şekeri ölçüm sistemlerinin geliştirilmesi önem arz etmektedir. Hastanın en kısa sürede ve en az ağrı ile kan

şekerinin ölçülmesi vücuttan alınacak sinyallerle mümkündür. Ancak ölçülen sinyallerin anlamlandırılması gerekmektedir. Bu süreç için sinyallerin karmaşık bir dizi matematiksel işlemlerden geçmesi ile mümkün hale gelmektedir. İnsan aklı ve gücüyle bu süreci yürütmek zor ve zahmetlidir. Bu aşamada yapay zekanın gücünden faydalanılacaktır. Yapay zeka algoritmalarının gelişmesi ile non-invaziv (girişimsel olmayan) yöntemlerle alınan kan şekeri ölçümlerinin doğruluğunu arttırmak mümkün hale gelmiştir. Yapılacak araştırma ile hastalara herhangi bir girişimsel müdahalede bulunmadan (kan alma, damar yolu açma vb.) kan şekeri değerlerinin ölçülmesi amaçlanmaktadır. Non-invaziv sistemler ile tarafsız veriler elde ederek diyabet tedavisini daha etkili hale getirmek istenmektedir.

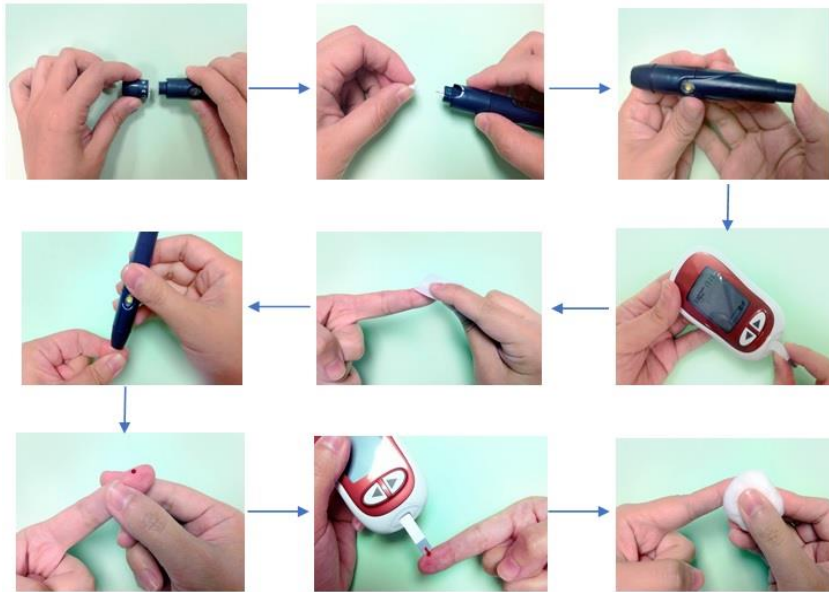
İnvaziv Ölçüm Yöntemi: İnvaziv ölçüm yöntemleri, vücuda girerek veya cilde nüfuz ederek bilgi toplamayı içerir. Bu yöntemler tipik olarak hastanın iç organlarına doğrudan erişim sağlar ve çeşitli tıbbi prosedürlerde kullanılır. Örneğin kalp kateterizasyonu, kalp hastalıklarının teşhisinde ve cerrahi prosedürlerin uygulanmasında yaygın olarak kullanılırken, laboratuvar testleri için kan numunesi alınması standart bir uygulamadır. Genel olarak invaziv ölçüm yöntemleri, vücut içindeki hedef organ veya dokuya doğrudan ulaşarak daha doğru sonuçlar verir. Bu nedenle kritik hastalıkların tanı ve tedavisinde tercih edilmektedir.

Bu ölçüm sistemi genellikle insuline dayalı başlangıç tedavisi sırasında kullanılmaktadır. Burada kullanılan cihazlar aktif ilaç tedavisi ile beraber kullanılan glukoz monitörleridir ve bunlar güçlü optik sensör teknolojisine dayalı durumdadırlar. Sadece glukozun tespiti için uygulanan optik sensör teknolojisi sayesinde doğrudan veriler elde edilebilmektedir. Glukoz Monitörünün temel amacı vücutta bulunabilecek maksimum glukozu tespit etmektir. Ancak, bu tip monitörler gerçekleştirebilecek enfeksiyon riskini arttırdığından, diğer alternatif yöntemlere ihtiyaç duyulmaktadır. Hastaların kendi imkanları ile kan şekerlerini ölçebilmeleri için aşağıda Şekil 1.1. de gösterilen cihaz ve ekipmanlara ihtiyaçları vardır.



Şekil 1.1. İnvaziv kan şekeri ölçüm cihazı ve ekipmanları[71]

Cihaz ve ekipmanlar temin edildikten sonra da hasta için zor ve ağrılı bir süreç başlamaktadır. Geleneksel yöntemlerle yapılan kan şekeri ölçüm işlemi genellikle hastanın tek başına yapabileceği bir işlem olmamaktadır. Hem ölçümden önce hem de ölçümden sonra alınması gereken önlemler vardır. Vücudun enfeksiyon kapmaması için ölçüm alınacak parmağın sürekli steril tutulması gerekmektedir. Aşağıda şekil 1.2’de evde kan şekeri ölçüm aşamaları gösterilmiştir.



Şekil 1.2. Evde kan şekeri ölçüm aşamaları[72]

Geleneksel süreçlerin hastanede gerçekleştiği düşünüldüğünde evde kan şekeri ölçümüne nazaran hastane ölçümün daha zahmetli ağırlı ve zaman alan bir süreç olduğu gözükmemektedir. Hastanelerde enfeksiyon riski daha fazladır bu sebeple hastanın ölçüm süresi boyunca azami önem göstermesi gerekmektedir. Yaşlılar ve farklı rahatsızlıkları olan insanlar için bütün bu işlemleri tek başına tamamlamak mümkün değildir. Ayrıca sıra bekleme gibi durumlar hastanın hastanede kaldığı süreyi uzattığı için bu durum hastanın psikolojisi üzerinde de olumsuz bir etki bırakmaktadır.

İnvaziv Olmayan Ölçüm Yöntemi: İnvaziv olmayan ölçümler, vücuda girmeden veya minimum temasla bilgi toplar. Özellikle kronik hastalıkların takibinde invaziv yöntemlerin olası riskleri ve rahatsızlıkları nedeniyle hasta tercihleri sıklıkla daha az invaziv, invaziv olmayan yöntemlere yönelmektedir. Non-invaziv yöntemler, kronik hastalıkların takibinde daha sık ve daha az invaziv olması nedeniyle genellikle tercih edilmektedir. İnvaziv olmayan ölçüm yöntemleri, vücudun dışından bilgi topladıklarından genellikle invaziv yöntemlere göre daha az doğru sonuçlar verir. Ancak teknolojiadaki gelişmeler invaziv olmayan yöntemlerin doğruluğunu arttırmaktadır. Makine öğrenmesi, derin öğrenme ve yapay sinir ağı modellerinin gelişmesi ile birlikte invaziv olmayan yöntemlerin kullanım alanı genişlemektedir.

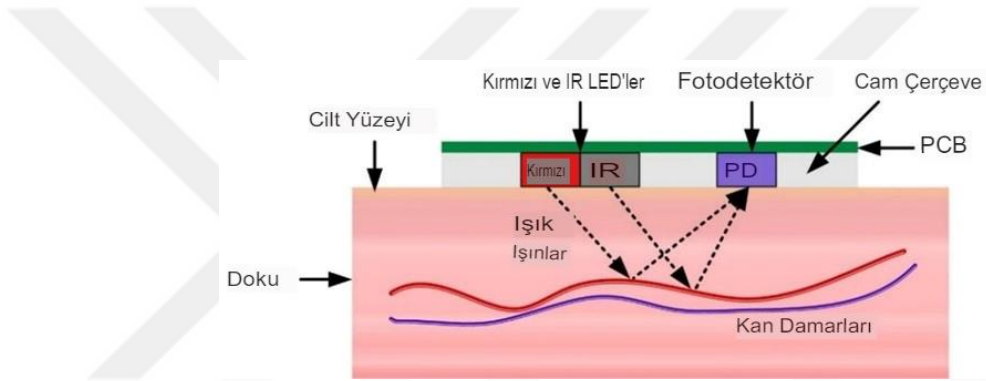
İnvaziv ve yarı invaziv yöntemlere kıyasla, sürekli glikoz takibi için invaziv olmayan yöntemler tercih edilir[6]. Glikoz ölçümünde optik yöntemlerin daha güvenilir ve doğru olduğu bulunmuştur[7]. İnvaziv olmayan ölçümler için yaygın optik yöntemler arasında Raman spektroskopisi, yakın kızılötesi spektroskopi, polarimetri, saçılma spektroskopisi[8], fotoakustik spektroskopi[9] ve diğerleri yer alır. İnvaziv olmayan ölçümler önceki sorunları ele alabilir ve ağrısız ve doğru çözümler sağlayabilir[10]. Non-invaziv bir yöntem kullanılarak kan örnekleme ortadan kaldırılarak bakteriyel ve viral enfeksiyon riski azaltılır ve hastanın rahatsızlığı hafifletilir.

Hem invaziv hem de invaziv olmayan ölçüm yöntemleri fiziksel temas gerektirir. Ancak invaziv olmayan yöntemler cilt yüzeyiyle sınırlı olduğundan invaziv yöntemlere göre daha az rahatsızlık verir. Hastalar genellikle daha az invaziv olması, daha az risk ve daha fazla konfor nedeniyle invaziv olmayan yöntemleri tercih etmektedir. Hasta konforu, düşük maliyetler ve tekrarlanabilirlik gibi faktörler invaziv olmayan yöntemlerin popülaritesine katkıda bulunmaktadır. Ancak hastalığın şiddeti ve spesifik takip parametreleri bazen invaziv yöntemlerin kullanımını gerektirebilmektedir. Tablo 1.2.'de halihazırda kullanımda olan invaziv ve invaziv olmayan kan şekeri ölçüm yöntemlerinin özellikleri gösterilmektedir.

Tablo 1.2. Kan şekeri ölçüm yöntemleri

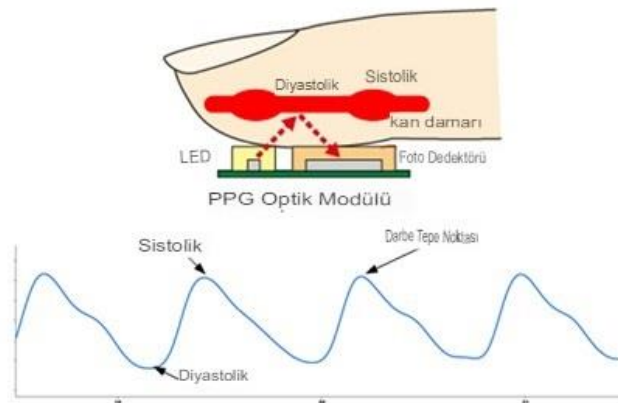
Metot	Açıklama	İnvaziv/Non- invaziv
Parmak Ucu Testi (Kan Glukoz Testi) [6]	Parmak ucundan kan örneği toplamak için iğne veya lanset kullanılır, bu örnek glukoz ölçer ile ölçülür.	İnvaziv
Sürekli Glukoz İzleme (CGM) [11]	Deri altına yerleştirilen bir sensör, kan glukoz seviyelerini sürekli olarak ölçer ve verileri bir monitör veya akıllı telefona iletir.	Non-invaziv
İntravenöz Kan Alma [12]	Laboratuvar testleri ve analizler için intravenöz olarak kan örnekleri alınır.	İnvaziv
Elektrokimyasal Metotlar [13]	Kan örneğinde glukoz seviyelerini ölçmek için elektrotlar kullanılır. Glukoz bir enzim ile reaksiyona girer, glukoz miktarına orantılı bir elektrik sinyali üretir.	İnvaziv
Transdermal Glukoz İzleme [14]	Deriye yerleştirilen bir cihaz, elektromanyetik dalgalar veya ultrason gibi teknolojileri kullanarak deriden glukoz seviyelerini ölçer.	Non-invaziv
Tükürük veya Gözyaşı Analizi [15]	Tükürük veya gözyaşı gibi vücut sıvılarında glukoz seviyelerini ölçer, hala araştırma ve geliştirme aşamasındadır.	Non-invaziv
Optik Sensörler [16]	Glukoz seviyelerini belirlemek için lazer ışığı veya kızılötesi ışık gibi optik sensörler kullanılır.	Non-invaziv
Biyoelektrik Empedans Analizi [17]	Deri yüzeyine yerleştirilen elektrotlarla vücudun elektriksel özelliklerini ölçerek glukoz seviyelerini belirler.	Non-invaziv

Noninvaziv kan şekeri ölçüm sistemlerinin geliştirilmesi diyabet hastalığının tanı ve tedavisinde büyük önem taşımaktadır. Vücuttan alınan sinyaller kullanılarak noninvaziv kan şekeri ölçümünün en az acı verici ve en kısa sürede gerçekleştirilmesine yönelik önlemler alınabilmektedir. Bu sinyalleri ölçmek için fotopletismografi (PPG) sensörleri kullanılır. Bu sensörler kan akışını ve hacmini ölçmek için ışık kullanarak çalışır ve sıklıkla kalp atış hızını, kan oksijen doyumunu (SpO2) ve diğer kardiyovasküler parametreleri izlemek için kullanılır [18]. PPG sensörünün sinyal ölçüm sistemi Şekil 1.3'te gösterilmektedir.[19]



Şekil 1.3. Veri okuma sisteminin parmak üzerinde gösterimi[73]

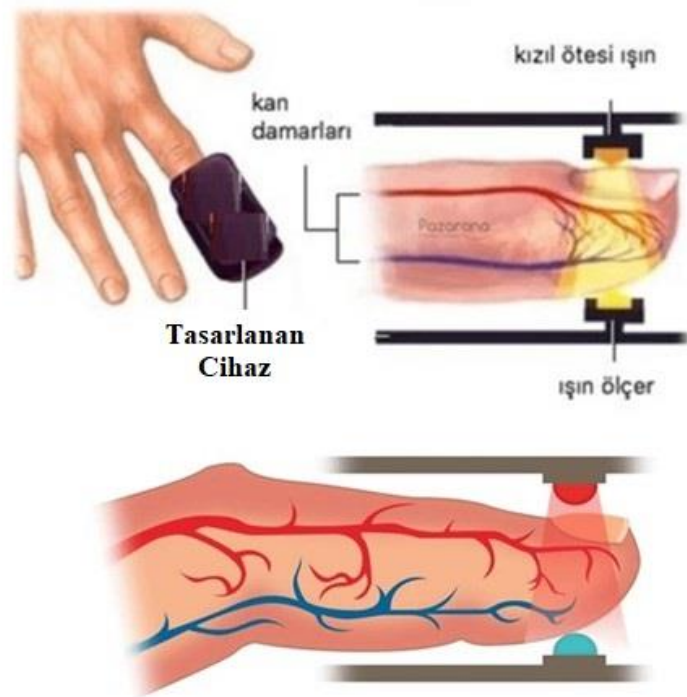
PPG sensörüyle ölçülen sinyaller, karmaşık bir dizi matematiksel işlemi içeren yorumlama gerektirir. Bu sürecin, yalnızca insan zekası ve çabasıyla gerçekleştirilmesi zorlu ve hantaldır. Şekil 1.4' te PPG sensörü ve ölçüm sinyali gösterilmektedir.[20]



Şekil 1.4. PPG sensörü ölçüm sinyali[74]

PPG sinyalleri analiz edilirken yapay zekanın potansiyeli ortaya çıkıyor. Yapay zeka algoritmalarının gelişmesiyle birlikte, invaziv olmayan kan şekeri ölçümlerinin doğruluğunu artırmak artık mümkün. Bu çalışma, kan alma gibi invaziv prosedürler olmadan kan şekeri seviyelerini ölçmeyi amaçlamaktadır. İnvazif olmayan sistemlerin kullanımının tarafsız veriler sağlaması ve potansiyel olarak diyabet tedavisinin etkinliğini artırması bekleniyor.

Yapay Zeka Tabanlı Non-İnvaziv Kan Şekeri Ölçüm Sistemi: Yapay zeka tabanlı non-invaziv kan şekeri ölçme sistemi klasik invaziv sistemin eksikliklerini ortadan kaldırmak için geliştirilmiştir. Bu sistem parmağa takılacak veri okuma sistemi ile okunan verinin işlenerek kaydedileceği mikrodenetleyici sistemden oluşmaktadır. Parmağa takılan veri okuma sistemi kızılötesi ışığın yansıması ve geri gelmesi prensibine dayalı olarak çalışmaktadır. Veri okuma sistemi parmağa herhangi bir cerrahi müdahale yapmamaktadır. Şekil 1.5'te tasarlanan veri okuma sisteminin parmak üzerindeki gösterimi yer almaktadır. Ölçümler sırasında ölçüm cihazı hastaların işaret parmağına takılmıştır ancak tüm parmaklardan ölçüm alınabilmektedir.



Şekil 1.5. Veri okuma sisteminin parmak üzerinde gösterimi[75]

1.1 Araştırmanın Amacı

Geliştirilecek olan yapay zeka tabanlı non-invaziv kan şekeri ölçümü sistemindeki temel prensip, günlük hayatımızda kullanılabilecek nesnelerden gelen bilgilerden hareketle vücudumuzdaki glukoz seviyesi hakkında bilgi toplamaya çalışmaktır.

Bazı ara çalışmalarda, bu tip teknolojiye en önemli unsurlardan birisinin sinirsel ağ başlangıcında meydana gelen işaretlemeler olduğu görülmüştür. Bu işaretlemeler sayesinde, vücutta bulunabilecek baskılar hakkında bilgi toplanabilmektedir. Bu bilgiler yapay zeka ile anlamlandırılabilir. Sonuç olarak yapay zeka tabanlı non-invaziv kan şekeri ölçümü sistemindeki işaretlemeler vücudumuzdaki glukoz seviyesi hakkında bilgi vermektedir, bu işaretleme sayısı geçmiş döneme ait incelendiğinde yapay zeka eğitimi ile birlikte anlamlı sonuçlar üretecektir.

Diyabet hastalarının yaşam kalitesini artırmak, sürekli ve acısız glukoz izlemeye olanak tanıyan yenilikçi çözümler geliştirmek amacıyla, farklı yapay zeka tekniklerinin uygulanabilirliğini araştırmak hedeflenmektedir. Bu kapsamda, çeşitli veri analizi ve makine öğrenme modellerini kullanarak glukoz seviyelerinin tahmin doğruluğunu değerlendirecek, en iyi performans gösteren modelleri belirlenecektir.

Araştırmanın amaçlarından bir tanesi hastaların girişimsel bir müdahaleye maruz kalmadan kan şekerlerinin ölçülmesidir.

Hastanın hastanede geçirdiği süreyi azaltarak hastalığın verdiği psikolojik etkiyi azaltmak hedeflenmektedir.

Hastaların hastane ortamında ve ev ortamında kullanımı süresinde geleneksel ölçüm sisteminde olan enfeksiyon riskini ortadan kaldırmak amaçlanmaktadır.

Geleneksel yöntemin eksikliklerini kapatarak hastalar ve sağlık personelleri için konforlu bir ölçüm amaçlanmaktadır.

Yapay zeka eğitimi ile girişimsel olmayan non-invaziv kan şekeri ölçüm sisteminin doğruluğu arttırılacaktır.

Gerçek hastalardan veri tabanı oluşturulmuştur, araştırma daha sonraki çalışmalar için dayanak teşkil edecektir. Bu yönüyle de araştırma süreklilik arz etmektedir.

1.2 Araştırmanın Sınırlılıkları

Bu tez çalışması,

2023 – 2024 Yılları arasındaki hastaların bilgileri ile sınırlıdır.

Abant İzzet Baysal Hastanesi Dahiliye Polikliniğine diyabet şikâyeti ile gelen hastaların ölçümelerini kapsamaktadır.

İlgili hastaların sadece kan şekeri düzeylerini içeren laboratuvar sonuçları çalışmaya dahil edilmiştir.

Tez çalışmasında makine öğrenmesi yöntemlerinden Kategori Hızlandırma Regresörü (CatBosst), Hafif Gradyan Arttırma Makinesi (Light Gradient Boosting Machine-Lightgbm), Extra Ağaç Regresörü (Extra Trees Regresör-Et), Gradyan Arttırma Regresörü (Gradient Boosting Regresör-Gbr), Ada Arttırma Regresörü (AdaBoost Regresör-Ada), Rastgele Orman Regresörü (Random Forest Regresör-Rf), Aşırı Gradyan Arttırma (Extreme Gradient Boosting-Xgboost) yöntemleri kullanılmıştır.

İlgili bütün yazılımlar Anaconda Navigator içinde bulunan Jupiter Notebook ortamı ile sınırlıdır.

İlgili sinyal ölçümlerinde kalp atış hızı (HR), oksijen satürasyonu (SpO2) ve diğer biyometrik ölçümleri yapmak için tasarlanmış Maxim Integrated firması tarafından üretilen MAX30105 sensörü kullanılmıştır.

MAX30105 sensörün kırmızı led için 660 nm, kızılötesi (IR) led için 880 nm, yeşil led için 537 nm dalga boyları kullanılmaktadır.

Sensör bilgilerinin alınması ve kaydedilmesi aşamasında Arduino ekibi tarafından üretilen 16Mhz iletişim hızına sahip Arduino Nano mikrodenetleyicisi ve Arduino arayüz yazılımı kullanılmıştır.

İlgili sensör ölçümleri Windows ortamında bilgisayarda saklanmaktadır.

2. LİTERATÜR ÖZETLERİ

Kan şekeri değerlerini tahmin etmeye dayalı invaziv ve non- invaziv birçok çalışma yapılmıştır. Non-invaziv sistemler, tarafsız veriler elde etmeyi amaçlayarak diyabet tedavisini daha etkili hale getirmektedir. İnvaziv olmayan glikoz izleme teknolojisi, glikoz tespitini kolaylaştırma potansiyeline sahiptir ve daha özel tedavi seçeneklerine yol açabilir [21]. Optik non-invaziv sensörler kullanarak kan şekeri seviyelerini hesaplayan bir cihaz tasarlamak ve uygulamak için bir çalışma başlatılmıştır [21]. Karar bakım yöntemlerinin de otomatikleştirdiği, özelleştirdiği ve akıllı glikoz yönetimini sağladığı düşünülmektedir. Doğru teşhis, izleme ve diyabetik ilaç yönetimi için büyük çabalar harcandı. Asıl zorluk, günlük kişiselleştirilmiş kullanım için biyosensörlerin oluşturulmasıdır.

Kan şekeri düzeylerini belirlemeye yönelik geleneksel yöntemler arasında deri altı ölçümler ve invaziv müdahaleler yer alır. Sürekli izlemenin gerekli olduğu durumlarda, hastanın derisinin altına, glikoz seviyelerini sürekli olarak takip eden bir sensör implante edilir. Bu sensörler sürekli olarak ölçüm alır ve kaydeder. Bu prosedür genellikle hastane ortamında bir doktorun gözetiminde gerçekleştirilir. Elektrokimyasal ölçüm sistemleri üzerine yapılan araştırmalar, sensör boyutlarını azaltmayı ve ölçüm doğruluğunu artırmayı amaçlamaktadır. Kubilay A., Hakan B. ve meslektaşlarının yaptığı çalışmada tekrar kullanılabilen ve enzim içermeyen bir sensör geliştirildi [22]. Bu sensörlerde enzim bulunmaması vücuttaki olası komplikasyonları azaltmaya yardımcı olur.

Fatemeh Karimi ve meslektaşlarının yaptığı bir çalışmada, enzim içermeyen sensörler oluşturmak için ekonomik üretim yöntemleri kullanıldı [23]. Geliştirilen sensörün son derece kararlı, ayırt edici ve hassas olduğu görüldü; bu da onun pazar için bir glikoz sensörü olma potansiyelini gösteriyor. Bu sensörler, geleneksel glikoz sensörlerinin ve invaziv yöntemlerin dezavantajlarının üstesinden gelmek üzere tasarlanmış olup gelecekte glikoz tespit sistemlerini iyileştirme potansiyeli sunmaktadır.

Araştırma grupları, onlarca yıldır güvenilir ve uzun vadeli doğru ticari, invaziv olmayan bir glikoz cihazı piyasaya sürmeye çalışıyorlar [24][25]. Bununla

birlikte, mevcut non-invaziv cihazların hiçbirisi, parmak delme yöntemlerinin yerini almak için gereken doğruluğa ve uzun vadeli tutarlılığa sahip değildir. Bu nedenle, normal ev koşullarında kamu kullanımına uygun, invaziv olmayan bir glikoz sensörü üretmek için ek çalışma gereklidir. İnvaziv olmayan ölçüm, önceki tüm sorunları hafifletecek ve ağrısız ve doğru çözümler sağlayacaktır [6][10]. İnvaziv olmayan bir yöntem kullanıldığında, kan örnekleme, bakteri ve patojenik virüslerin vücuda girmesi hariç tutulur ve hasta ağrıdan kurtulur.

Diyabet hastalarının kan şekeri tahminlerinin genelleştirilebilir özelliklerine sahip olduğu bir çalışmada başarı oranının yeterince yüksek olmadığı, hastaya özel tahminlerde başarı oranının daha yüksek olduğu rapor edilmiştir [26]. Diğer bir çalışmada, kan şekeri tahmini için makine öğrenme yöntemleri ile hem glikoz hem de insülin bilgisini besleyen gerçek zamanlı öğrenen tekrarlayan sinir ağı kullanılmıştır. Model, hem glikoz bilgisi hem de harici insülin girdisi olan bir otoregresif (Autoregressive-AR) modelini makine öğrenmesi yöntemleriyle birleştirerek klasik bir otoregresif modelden daha iyi performans göstermiştir [27].

1990 yılında tedavi sürecinde hekimlere yardımcı olacak bir diyabet tahmin modeli geliştirilmesi amacıyla Ulusal Diyabet ve Sindirim ve Böbrek Hastalıkları Enstitüsü tarafından, 768 tip 2 diyabet (T2D) hastasından önemli bilgiler içeren veri seti oluşturulmuştur. Çalışmada karar ağacı algoritması uygulanarak 27 kural çıkarılmıştır [28]. Başka bir çalışmada, sürekli glukoz izleme (CGM) verilerinden kan şekeri seviyelerini tahmin etmek için destek vektör regresyonu (Support Vector Regression-SVR) kullanılmıştır. Ancak çalışmada, düzenli bir kübik spline interpolasyonu kullanılarak düzeltilmiş veriler kullanılmıştır [29]. Kişiselleştirilmiş bir glikoz tahmini sağlamak için yemek, insülin ve egzersiz modellerini SVR yöntemini ile birleştiren başka bir çalışma önerilmiştir.

2018 yılında yapılan araştırmaya göre kan şekeri değerlerinin tahmininde uzun kısa süreli bellek (Long Short Term Memory-LSTM) ve iki yönlü uzun kısa süreli bellek (Bidirectional Long Short Term Memory) yöntemleri kullanılarak derin sinir ağı (Deep Neural Network-DNN) ile sonuçlar alınmıştır [30]. CGM ile elde edilen verileri girdi olarak alan ve 15, 30 ve 45 dakika seviyelerini geleceğe tahmin eden ileri beslemeli bir sinir ağı (Feedforward Neural Network-FNN)

kullanılmıştır [31]. Aşırı gradyan artırma (Extreme Gradient Boosting-XGBoost) özelliğini tahmin etme yöntemlerinin geliştirilmesinde kullanılmıştır [32]. Otoregresyon eksojen (AutoRegressive Exogenous-ARX) girdiler, RNN'ler ve evrişimli sinir ağları (Convolutional Neural Network-CNN) ile karşılaştırılmış ve daha basit ARX modellerinin Ohio üniversitesinde kan şekeri tahmini yapılmıştır [33].

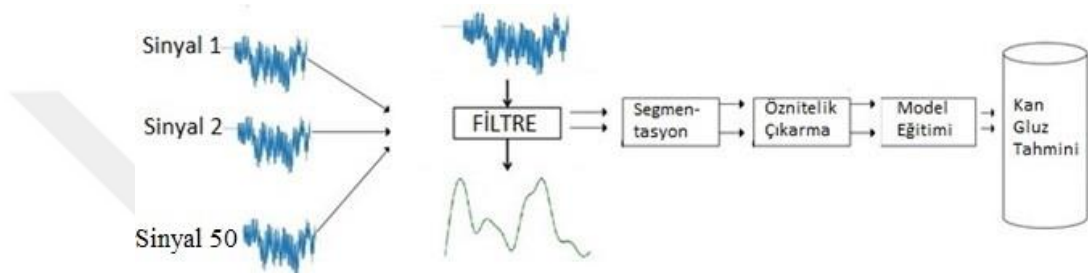
2021 yılında ise sinyal ölçüm cihazlarının gelişimi ile birlikte girişimsel olmayan non-invaziv ölçüm sistemleri önerilmeye başlanmıştır. Sentanu S. G. Ve arkadaşları yaptıkları çalışmada donanım sistemi geliştirdikten sonra 24 gönüllünün Fotopletismografi (PPG) verilerini toplamış, her gönüllü için toplam dört dakikalık yansıtıcı ve aktarıcı PPG sinyali almışlardır. Aynı zamanda referans olması için Caresens II Plus adı verilen klinik invaziv tipte bir cihaz kullanılarak glikoz seviyesi ölçümü yapmışlardır. Çalışmada sinyal ön işleme ve öznitelik çıkarma işlemi yapıldıktan sonra, Rastgele Orman (RF) ve Ekstra Gradyen Arttırma (XGBoost) makine öğrenmesi modelleri ile PPG sinyallerinden kan şekeri tahmini gerçekleştirilmeye çalışılmıştır [34].

2022 yılında Ernias Susana ve arkadaşlarının yaptığı başka bir çalışmada ise hastalardan hem parmak sensöründeki elektronik devre aracılığıyla ortaya çıkan PPG sinyalleri hem de invaziv yöntemle kan şekeri değerleri alınmış ve şeker ölçüm sonuçlarına göre hastalar normal ve diyabet olarak sınıflandırılmıştır. Ardından Rastgele Orman (RF) makine öğrenmesi modeli içerişindeki torbalama(bagging) fonksiyonu kullanılarak PPG sinyalleri ile klinik invaziv ölçüm ile alınan kan şekeri değerleri regresyona tabi tutulmuş ve hastaların PPG sinyallerine bakarak diyabet sınıflandırılması gerçekleştirilmeye çalışılmıştır [35].

Luis A.C.P. ve arkadaşlarını 2023 yılında herhangi bir patoloji veya herhangi bir ilacın etkisi bildirilmemiş, yaşları 17 ile 60 arasında değişen 187 kadın ve erkek denek kullanılarak geliştirdikleri sistemde, verileri rastgele kan şekeri testi modunun protokollerini uygulamışlardır. Bu çalışmada, üç dalga boyuna sahip PPG sinyalleri kullanılarak kan şekerinin invaziv olmayan tahmini için bir sistem önermişlerdir. Bu sistemde gözetimli öğrenme yöntemlerinden biri olan SVM yöntemi ile PPG sinyalleri kan şekeri tahmini gerçekleştirilmeye çalışılmıştır[36].

3. MATERYAL VE YÖNTEM

Çalışmada yöntem adımları veri setinin okunması ve verilerin ön işlemlerden geçirilmesi ile başlayacaktır. Ön işlemlerden geçirilen veriler segmentasyon işlemine tabi tutulacak ardından öznitelikler çıkarılacak ve hedef değişken ile en ilgili öznitelikler seçilecektir. Açıklanabilir yapay zeka ile eğitilen modelin kan glukoz seviyesi tahminini gerçekleştirmesi sağlanacaktır. Bu adımlar Şekil 3.1’de gösterilmiştir



Şekil 3.1. Çalışma yöntem adımları

Sinyallerin Toplanması: İlk adım, bütünleşik devreden gelen çeşitli biyomedikal sinyallerin toplanmasıdır. Bu sinyaller, kan glukoz seviyesini tahmin etmek için kullanılacak veriyi oluşturur.

Filtreleme: Toplanan sinyaller, gürültü ve istenmeyen bileşenlerin çıkarılması için filtrenir. Bu adım, sinyallerin daha doğru ve anlamlı hale getirilmesi için gereklidir.

Segmentasyon: Filtrelenmiş sinyaller, belirli zaman dilimlerine veya olaylara göre segmentlere ayrılır. Bu, sinyalin farklı bölümlerinin ayrı ayrı analiz edilmesini sağlar.

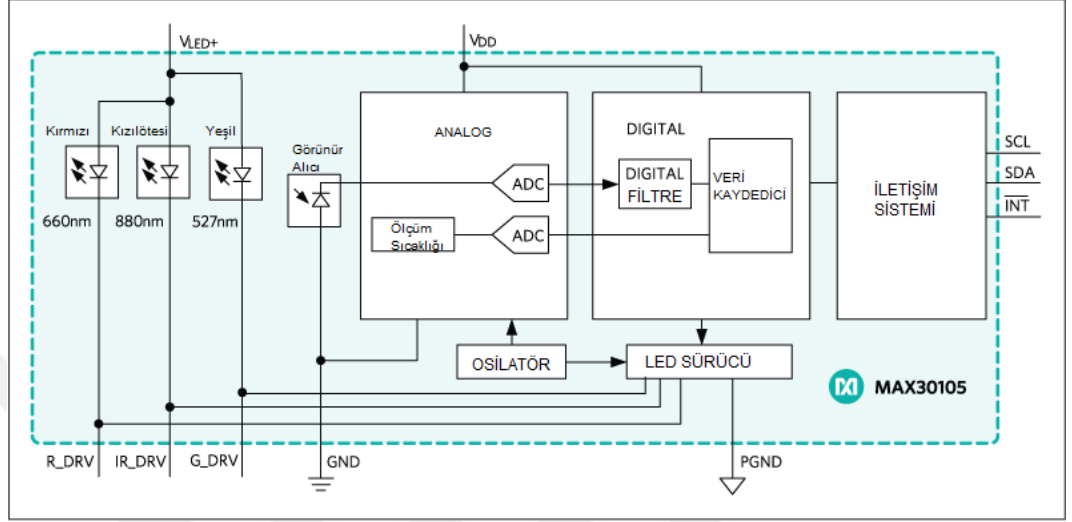
Öznitelik Çıkarma: Segmentlere ayrılmış sinyallerden anlamlı öznitelikler çıkarılır. Bu öznitelikler, sinyalin özelliklerini temsil eden matematiksel veya istatistiksel değerlerdir.

Model Eğitimi: Çıkarılan öznitelikler kullanılarak bir makine öğrenimi modeli eğitilir. Bu model, verilen özniteliklere dayanarak kan glukoz seviyesini tahmin etmeyi öğrenir.

Kan Glukoz Tahmini: Eğitilmiş model, yeni gelen sinyallere uygulandığında kan glukoz seviyesini tahmin eder. Bu sonuç, modelin doğruluğunu ve güvenilirliğini yansıtır.

3.1 Veri Setinin Oluşturulması

Parmağa takılan ve PPG sinyalleri kullanarak kan şekeri seviyelerini tahmin etmek amacıyla bir elektronik devre oluşturulmuştur. Şekil 3.2’de bu elektronik devrede PPG sinyali ölçümleri için kullanılan MAX30105 sensörünün iç yapısı görülmektedir.



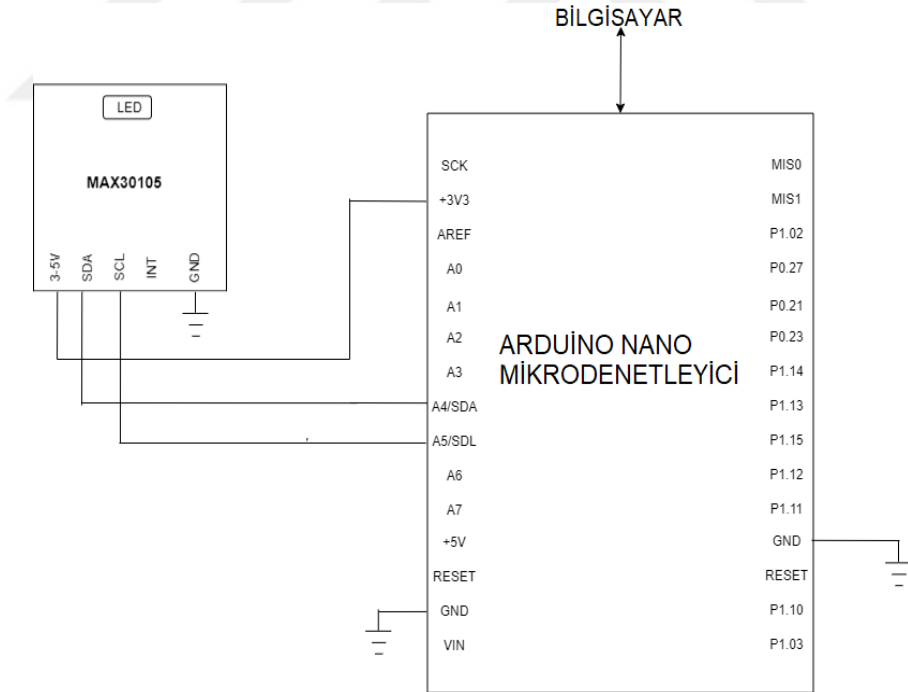
Şekil 3.2. MAX30105 sensörü iç yapısı[79]

- 1- Kızılötesi Algılama Sistemi LED ve Fotodiyot: Cihaz, kızılötesi ışık yayan bir LED ve bu ışığı algılayan bir fotodiyot içerir. LED, deri altındaki kan damarlarına kızılötesi ışık gönderir. Fotodiyot ise geri yansıyan ışığı algılar. Kan hacmindeki değişiklikler, ışığın emilme ve yansıma oranlarını etkiler ve bu sayede PPG sinyalleri elde edilir. Bu işlem için MAX30105 sensörü kullanılmıştır.
- 2- Analog Sinyal Kuvvetlendirme (Amplifikatör): Fotodiyot tarafından algılanan sinyaller çok düşük voltaj seviyelerindedir. Bu nedenle, bu sinyaller bir kuvvetlendirici devre kullanılarak güçlendirilir. Bu, sinyalin daha kolay işlenebilir hale gelmesini sağlar. Bu işlem için MAX30105 sensörü içerisindeki sinyal kuvvetlendirme sistemi kullanılmıştır.
- 3- Sinyal Dijitalleştirme Analog-Dijital Çevirici (ADC): Kuvvetlendirilmiş analog sinyaller, bir Analog-Dijital Çevirici (ADC) kullanılarak dijital formata dönüştürülür. Bu, sinyallerin dijital işlemciler veya mikrodenetleyiciler tarafından işlenmesini mümkün kılar. Bu işlem için MAX30105 sensörü içerisinde analog- dijital çevirici sistem kullanılmıştır.

- 4- Sinyal İşleme ve Analiz (Mikrodenetleyici): Dijital sinyaller, mikrodenetleyici tarafından işlenir. Mikrodenetleyici, sensörden gelen PPG sinyallerini analiz eder okuma hızına bağlı olarak ilgili sinyalleri sayısal verilere dönüştürür. Bu işlem için Arduino Nano mikrodenetleyicisi kullanılmıştır.
- 5- İletişim Modülü: Mikrodenetleyici ve sensörden oluşan tümleşik elektronik devre bilgisayara kablo aracılığıyla bağlıdır. Kablolu bağlantı seçilmesinin sebebi gecikmelerin ve sinyal kayıplarının önüne geçmek içindir.

Tümleşik Elektronik Devre Çalışma Prensibi; Cihaz, kullanıcının parmağına takılır ve sürekli olarak PPG sinyallerini toplar. Kızılötesi LED, deri altına ışık gönderir ve fotodiyot, geri yansıyan ışığı algılar. Kan hacmindeki değişiklikler, PPG sinyallerinde dalgalanmalara neden olur. Bu sinyaller, kuvvetlendirici ile güçlendirilir ve ADC ile dijital formata dönüştürülür.

Mikrodenetleyici, dijital PPG sinyallerini analiz eder ve sayısal verilere dönüştürür. Elde edilen veriler, kablolu iletişim modülü aracılığıyla mobil bilgisayarlara gönderilir ve burada kaydedilir. Sensör ve mikrodenetleyici bağlantı şeması Şekil 3.3. 'te gösterilmektedir.



Şekil 3.3. Sensör ve mikrodenetleyici bağlantı şeması

Kaydedilen veriler ile kan şekeri seviyesini tahmin etmek için bir veri seti oluşturulmuştur.

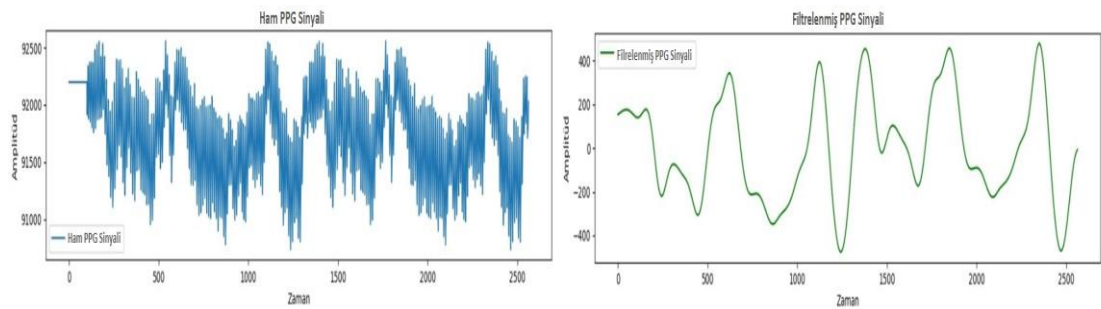
3.2 Veri Seti Kümesi

Önerilen sistemin geliştirilmesinde ilgili üniversiteden etik kurul onayı alınarak veri tabanı oluşturma işlemi gerçekleştirilmiştir. Hastalardan alınan imzalı bilgilendirilmiş olur formu ile ölçüm işlemi başlamıştır. Tüm ölçümler Abant İzzet Baysal Üniversitesi Dahiliye Polikliniğinde yapılmış ve ilgili birime diyabet şikâyeti ile gelen hastalardan ölçüm alınmıştır. Bu veri tabanı şunları içermektedir; 50 katılımcının klinik-kimyasal glikoz değerleri ve PPG sinyali değerleri. PPG sinyali ölçümünde çalışma için özel oluşturulmuş mikrodenetleyici ve sensör sisteminden oluşan bütünleşik devre kullanılmıştır.

Bütünleşik ölçüm sistemi ile hastanın parmağında ortalama 2 dakika süreyle, 64 Hz örnekleme frekansı ile alınmıştır. Bütünleşik devre, nabız dalgası algılamayı en üst düzeye çıkarmak için yeşil (550 nm) ve kırmızı (650 nm) ışık kullanır. Laboratuvar ortamında alınan glukoz değeri hedef değişken olarak kullanılacaktır.

3.3 Veri Ön İşleme

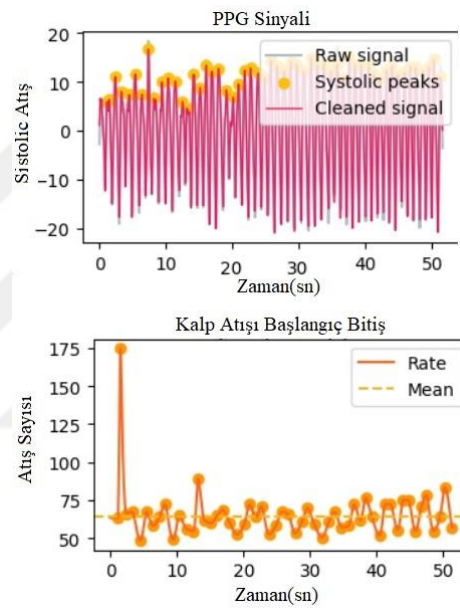
Cihazın başlangıç parametrelerindeki hatadan kaçınmak için ilk 10 saniyelik değerler sinyalden çıkarılmıştır. Sinyal ön işlemede NeuroKit2 [37] kütüphanesi kullanılmış ve Elgendi [38] metodu ile temizleme yapılarak PPG sinyalleri gürültü ve yapaylık özelliklerinden arındırılmıştır. Bu temizleme yöntemi, geniş bant genlik entegrasyonu (Elgendi) adını taşır ve PPG sinyalinin geniş bant özelliklerini değerlendirmek için kullanılır. Elgendi, özellikle PPG sinyalinin kalitesini arttırmak ve belirli frekans bantlarındaki genlikleri entegre etmek amacıyla kullanılır. Bu işlem Şekil 3.4 'de gösterilmiştir.



Şekil 3.4. Temizlenen sinyal ve ham sinyal

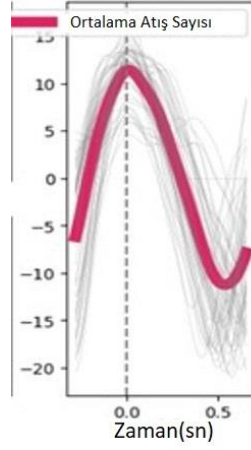
3.4 Segmentasyon (Bölütleme)

Fotopletişimografi (PPG) sinyali segmentasyonu, bir PPG sinyalini belirli olaylara veya dönemlere ayırmayı ifade eder. PPG sinyali genellikle kalp atış döngüsünü temsil eder ve segmentasyon, bu döngü içindeki belirli olayları veya özellikleri tanımlamak için kullanılır. PPG sinyali segmentasyonunda kullanılabilecek birkaç yöntem vardır. Bunlardan bazıları; r zirvesi algılama, tersine çevrilen zirve algılama, kalp atışı başlangıcı ve bitişi, frekans alanı analizi vb. Bu çalışmada PPG sinyalleri kalp atışı başlangıç ve bitiş aralığına göre bölütlenmiştir. Bölütleme işlemi Şekil. 3.5'te gösterilmiştir.



Şekil 3.5. Bölütleme işlemi adımları

Segmentasyon işleminden geçirilen bölütlenmiş PPG sinyalleri ortalama alma işlemine tabi tutulmuştur. Bunun sebebi segmentler üzerinde çalışırken, spesifik olayları veya durumları incelemek içindir. Özellikle PPG sinyali gibi tekrarlanan döngüler üzerinde çalışılıyorsa, ortalama alma işlemi sayesinde her bir döngünün özelliklerini öğrenmek kolaylaşır. Şekil 3.6.'da üç numaralı hastanın kalp atış başlangıç bitiş değerleri ve ortalama değeri görülmektedir. Burada kırmızı renkli koyu olarak belirtilen kısım ortalama kalp atış çizgisini gösterirken, soluk renkli ince çizgiler ise hastanın tüm kalp atışlarına göstermektedir.



Şekil 3.6. Üç numaralı hastanın kalp atışı ortalaması

3.5 Öznitelik Çıkarma ve Seçme İşlemi

Ön işlemlerden geçirilen PPG sinyalleri TSFEL [39] kütüphanesi kullanılarak özniteliklerine ayrılmıştır. TSFEL, zaman serisi verilerinden çeşitli özellikleri çıkarmak için tasarlanmış bir Python kütüphanesidir. Kütüphane, bir dizi özellik çıkarma işlevi ve bunları kullanarak özellik matrisi oluşturma yetenekleri sunar. TSFEL mevcut özellikleri üç alana ayırır: istatistiksel, zamansal ve spektral. Çalışmada zamansal ve istatistiksel özellikler kullanılmış ve PPG sinyali 230 adet özelliğe göre bölümlere ayrılmıştır. Bu özelliklerden kan şekeri tahmininde değişikliğe sebep olmayacak olan aynı değerlerdeki sinyalin maksimum değeri, ECDF (ampirik kümülatif dağılım fonksiyonu değerlerinin zaman eksenini boyunca hesaplanması) değeri, sinyalin entropi değeri gibi öznitelikler çalışmadan çıkarılmıştır. Tahmin hesaplanmasında 51 adet sinyal özneliği kullanılmış ve kullanılan özniteliklerden bazıları Tablo 3.1. 'de belirtilmiştir.

Tablo 3.1. Kullanılacak özniteliklerden bazıları

PPG Özellikleri	Açıklama
1	Tepe değeri
2	Zirveden zirveye
3	Ortalama değer
4	Standart sapma
5	Kök Ortalama Kare (RMS)
6	Eğiklik
7	Yapışkanlık

8	Tepe faktörü
9	Açılış faktörü
10	Dürtü faktörü
11	Şekil faktörü

1- Tepe Değeri (Peak Value): Bir zaman serisinin en yüksek değerini temsil eder. Bu değer, zaman serisinin zirvesini belirler ve belirli bir dönem içinde gerçekleşen maksimum değeri temsil eder. Zaman serisindeki en yüksek değer belirlenmesi, anomali tespiti, sinyal işleme ve veri madenciliği gibi alanlarda özellikle önemlidir [40].

2- Zirveden Zirveye (Peak-to-Peak): Bir zaman serisinin en yüksek (tepe) ve en düşük (çukur) noktaları arasındaki farkı ifade eder. Bu metrik, sinyalin dalgalanma genliğini ölçmek için kullanılır ve zaman serisinin değişkenliğini veya dalgalanma aralığını temsil eder [40].

3- Ortalama Değer (Mean Value): Bir zaman serisinin veya veri setinin elemanlarının toplamının eleman sayısına bölünmesiyle elde edilen değerdir. Ortalama, merkezi eğilimi bir ölçü olarak kullanılır ve veri setinin genel eğilimini veya tipik değerini temsil eder [40].

4- Standart Sapma (Standard Deviation): Veri setindeki elemanların ortalamadan ne kadar saptığını ölçen istatistiksel bir değerdir. Standart sapma, bir veri setinin yayılımını veya dağılımını anlamak için kullanılır ve verilerin ortalama etrafında nasıl dağıldığını gösterir [40].

5- Kök Ortalama Kare (Root Mean Square-RMS): Bir veri setinin ortalama kare değerinin karekökü olarak tanımlanan bir ölçüdür. RMS, sinyal işleme, mühendislik ve istatistik gibi alanlarda yaygın olarak kullanılır ve bir veri setinin enerji içeriğini veya genlik seviyesini temsil eder [40].

6- Çarpıklık (Skewness): Verilerin ortalama etrafındaki dağılımının asimetrisini tanımlayan istatistiksel bir ölçüdür. Verilerin ortalamasının bir tarafında diğerine göre daha yoğun olup olmadığını gösterir. Çarpıklık pozitif, negatif veya sıfır olabilir ve her durum, veri dağılımının şekli hakkında farklı bir şey söyler [41].

7- Basıklık (Kurtosis): Bir dağılımın kuyruklarının şeklinin genel şekline göre nasıl olduğunu tanımlayan istatistiksel bir ölçüdür. Dağılımın kuyruklarındaki veri noktalarının normal bir dağılıma göre ağır mı yoksa hafif mi olduğunu gösterir. Başka bir deyişle, basıklık, bir veri setindeki aşırı sapmaların derecesini ölçer [41].

8- Crest Faktörü (Crest Factor): Bir sinyalin veya veri setinin tepe değerinin ortalama değere oranını ifade eden bir ölçüdür. Elektrik mühendisliği, ses mühendisliği ve sinyal işleme gibi alanlarda sıkça kullanılır. Crest faktörü, bir sinyalin tepe değeri ile ortalama değeri arasındaki ilişkiyi belirler ve sinyalin genliği hakkında bilgi verir [42].

Belirlenen özniteliklere göre ayrılan sinyaller laboratuvar ortamında alınan glukoz değerleri ile hedef değişken olarak ilişkilendirilmiş ve makine öğrenmesi modeli için hazır hale getirilmiştir.

3.6 Makine Öğrenmesi Modeli

PPG sinyalleri ön işlemlerden ve öznitelik çıkarma işlemlerinden geçtikten sonra hedef değişken ile Regresyona tabi tutuldu. Regresyon, iki veya daha fazla değişken arasındaki ilişkiyi modelleme ve bu ilişkiyi anlama sürecini ifade eder. Genellikle, bağımlı değişken (outcome variable) ile bağımsız değişkenler (predictor variables) arasındaki ilişkiyi analiz etmek amacıyla kullanılır. Çalışmada bir bağımlı değişkenin (hedef değişken) bir veya daha fazla bağımsız değişkenle ilişkisini modellemeye çalışan Pycaret [43] kütüphanesine bağlı regresyon modelleri kullanılmıştır. Bu modeller bir değişkenin diğer değişkenlere nasıl bağlı olduğunu anlamak ve gelecekteki değerleri tahmin etmek için kullanılır. Çalışmada bir dizi regresyon modeli ile eğitim yapılmıştır.

Tablo 3.2’de eğitilen sistemin ilk yedi regresyon modeline göre performans değerleri gösterilmektedir. Tablo, farklı makine öğrenimi modellerinin performansını ölçen çeşitli metrikleri içermektedir. Her model için ortalama mutlak hata (MAE), ortalama kare hata (MSE), kök ortalama kare hata (RMSE), belirleme katsayısı (R^2), kök ortalama kare logaritmik hata (RMSLE), ortalama mutlak yüzde hata (MAPE) ve eğitim süresi (TT) gibi metriklerle değerlendirildiği görülmektedir.

Tablo 3.2. İlk yedi regresyon modeli performans deęerleri

Model	MAE	MSE	RMSE	R ²	RMSLE	MAPE	TT(Sn)
CatBoost Regresör	25.16	1591.78	39.13	0.71	0.24	0.18	31.15
Light Gradient Boosting Machine	28.64	1941.95	43.45	0.65	0.26	0.20	0.02
Extra Trees	26.98	2095.21	44.52	0.64	0.27	0.19	0.07
Gradient Boosting Regresör	29.94	2368.13	47.99	0.59	0.28	0.20	0.08
AdaBoost Regresör	36.78	2400.90	48.24	0.58	0.29	0.26	0.04
Random Forest Regresör	29.58	2493.49	48.99	0.56	0.29	0.20	0.10
Extreme Gradient Boosting	30.93	2833.09	52.19	0.52	0.30	0.21	0.21

MAE (Ortalama Mutlak Hata): Modelin gerek deęerler ile tahmin edilen deęerler arasındaki ortalama mutlak farkı gsterir. Daha dşük bir deęer, daha iyi tahmin performansını iřaret eder [44].

MSE (Ortalama Kare Hata): Modelin gerek deęerler ile tahmin edilen deęerler arasındaki ortalama kare farkı gsterir. Daha dşük bir deęer, daha iyi tahmin performansını iřaret eder [45].

RMSE (Kök Ortalama Kare Hata): MSE'nin karekökü alınarak elde edilir. Gerek ve tahmin edilen deęerler arasındaki ortalama kare farkın karekökünü temsil eder. Daha dşük bir deęer, daha iyi tahmin performansını iřaret eder [46].

R² (Belirleme Katsayısı): Bağımlı deęiřkendeki varyasyonun ne kadarının bağımsız deęiřkenler tarafından açıklandığını gsterir. Deęerler 0 ile 1 arasında deęiřir ve daha yüksek bir deęer, modelin daha iyi uyum saęladığını gsterir [47].

RMSLE (Kök Ortalama Kare Logaritmik Hata): Gerek ve tahmin edilen deęerlerin logaritmalarının ortalama kare farkının karekökünü temsil eder. Özellikle veri setinde büyük deęerler olduęunda kullanışlıdır [48].

MAPE (Ortalama Mutlak Yüzde Hata): Gerek ve tahmin edilen deęerler arasındaki ortalama mutlak yüzde farkı temsil eder. Daha dşük bir deęer, daha iyi tahmin performansını iřaret eder [49].

TT (Eęitim Süresi): Modelin eęitim süresini temsil eder. Daha kısa süre, daha hızlı eęitim süresini iřaret eder.

CatBoost (Categorical Boosting): Yandex tarafından geliştirilen bu model, özellikle kategorik verilerle iyi performans göstermesiyle bilinir. CatBoost, kategorik değişkenleri otomatik olarak işleyerek, kullanıcıların manuel olarak bu değişkenleri dönüştürme ihtiyacını azaltır. Diğer boosting algoritmalarına göre hız ve doğruluk açısından avantajlıdır. LightGBM: Microsoft tarafından geliştirilen bu model, büyük veri setleri ve yüksek boyutlu verilerle çalışmak için optimize edilmiştir. LightGBM, histogram tabanlı algoritma kullanarak daha hızlı ve daha düşük bellek kullanımı sağlar. Ayrıca, doğruluğu artırmak için yaprak odaklı büyüme stratejisi kullanır. Extra Trees (Extremely Randomized Trees) [63]: Random Forest ile benzer şekilde çalışan bu model, ancak ağaçları oluştururken daha fazla rastgelelik ekler. Split noktalarını rastgele seçer ve böylece modelin varyansını azaltır. Bu da genellikle aşırı uyumu azaltır ve daha genelleştirilebilir modeller oluşturur. Gradient Boosting Regresör [64]: Bu model, hataları azaltmak için ardışık karar ağaçları inşa eder. Her yeni ağaç, önceki ağaçların hatalarını öğrenir ve düzelterek toplam hata oranını düşürür. Gradient Boosting, güçlü performansı ve esnekliği ile bilinir, ancak doğru hiper parametre ayarları gerektirir. AdaBoost (Adaptive Boosting) [65]: Zayıf öğreniciler (genellikle karar ağaçları) kullanarak çalışan bu model, her bir zayıf öğreniciyi ardışık olarak eğitir. Her aşamada, önceki öğrenicinin hatalarına daha fazla ağırlık verir, böylece modelin doğruluğunu artırır. AdaBoost genellikle gürültüye karşı hassastır. Random Forest [66]: Birden fazla karar ağacı oluşturarak çalışan bu model, her bir ağacın öngörülerini birleştirir. Rastgele örnekler ve özellikler seçilerek ağaçlar oluşturulur, bu da aşırı uyumu azaltır ve modelin genelleme yeteneğini artırır. Random Forest, istikrarlı ve genellikle yüksek performanslıdır. XGBoost [67]: Gradient Boosting algoritmasının optimize edilmiş ve hızlandırılmış bir versiyonudur. XGBoost, hız ve performans açısından birçok optimizasyon içerir ve genellikle yüksek doğruluk elde etmek için kullanılır. Paralel hesaplama ve bellek verimliliği ile öne çıkar.

Çalışmada eğitilen sistemde regresyon modellerinden en yüksek performansa sahip iki model olan CatBoost Regresör [50] ve LightGBM [51] (Light Gradient Boosting Machine) algoritmaları incelenmiştir.

CatBoost, regresyon görevlerinde kategorik değişkenleri doğrudan destekler. Bu, kullanıcının kategorik değişkenleri özel olarak işlemeye veya

dönüştürmeye gerek kalmadan kullanabilmesini sağlar. CatBoost, gradient boosting algoritmasını temel alır. Gradient boosting, zayıf öğrencilerin (genellikle karar ağaçları) bir araya getirilmesiyle güçlü bir model oluşturmayı amaçlar.

LightGBM büyük veri setleri ve yüksek boyutlu özelliklere sahip veri setleri üzerinde paralel işleme ve özel histogram tabanlı öğrenme yöntemi kullanarak hızlı eğitim sağlar. Modelin oluşturulmasından sonra, LightGBM kullanıcılara her bir özelliğin tahmindeki katkısını değerlendirmek için özellik önemi bilgileri sunar.

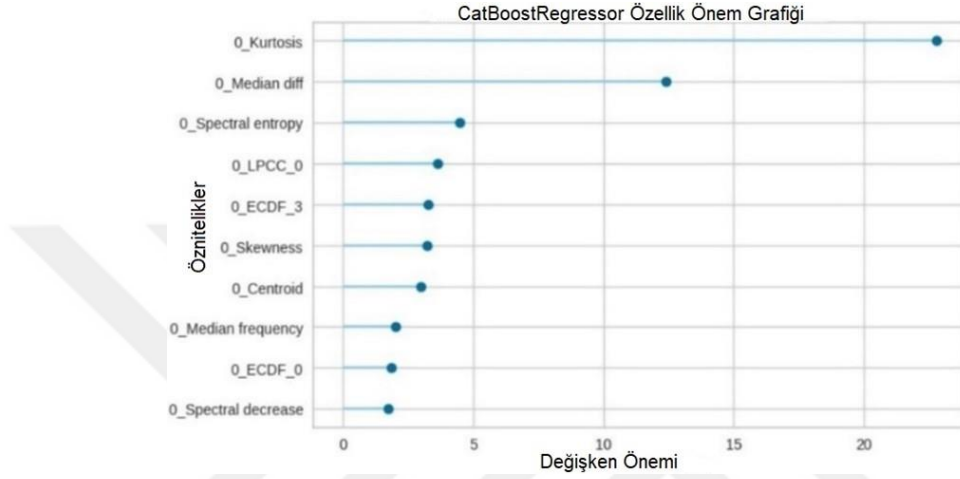
3.7 Açıklanabilir Yapay Zeka (XAI-SHAP)

Açıklanabilir yapay zeka (XAI), yapay zeka ve makine öğrenme modellerinin kararlarını ve tahminlerini daha anlaşılır ve şeffaf hale getirme sürecidir. XAI, kullanıcıların, geliştiricilerin ve denetleyicilerin, yapay zeka sistemlerinin nasıl çalıştığını ve belirli sonuçlara nasıl ulaştığını anlamalarını sağlar. SHAP (SHapley Additive exPlanations), bir modelin tahminlerini açıklamak için kullanılan bir yöntemdir. SHAP, kooperatif oyun teorisinden türetilmiş bir kavram olan Shapley değerlerine dayanmaktadır. Shapley değeri, her bir özelliğin (feature) modelin çıktısına ne kadar katkıda bulunduğunu adil bir şekilde hesaplamak için kullanılır. Shapley sistemi 3 ana unsurdan oluşmaktadır.

1. Shapley Değerleri: Shapley değeri, kooperatif oyun teorisinde her oyuncunun (özelliğin) oyunun (modelin) sonucuna olan katkısını ölçmek için kullanılan bir yöntemdir. SHAP, bu konsepti kullanarak her bir özelliğin bir tahmine olan katkısını hesaplar.
2. Katkı Hesaplama: SHAP, her bir özelliğin çıkarılması ve eklenmesi durumunda modelin çıktısındaki değişiklikleri göz önünde bulundurarak katkı değerlerini hesaplar. Bu hesaplamalar, tüm olası kombinasyonlar üzerinden ortalama alınarak gerçekleştirilir.
3. Açıklama ve Görselleştirme: SHAP değerleri kullanılarak, özelliklerin model tahminine ne kadar ve hangi yönde katkı sağladığı görselleştirilebilir. Bu görselleştirmeler, karar ağaçları, bar grafikler veya scatter plotlar gibi farklı şekillerde sunulabilir.

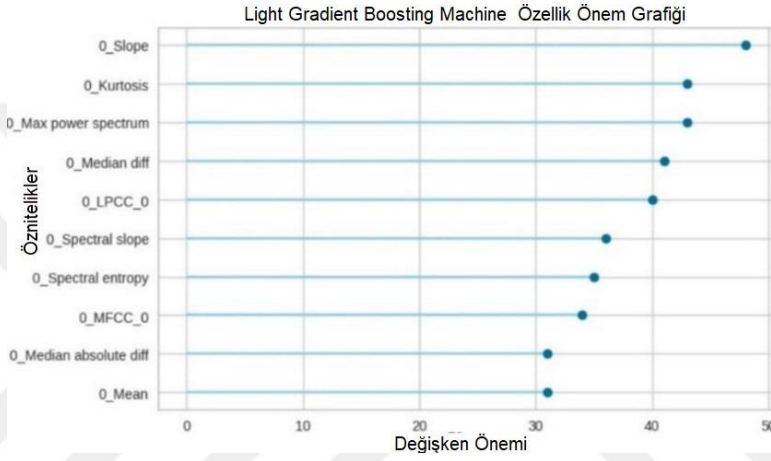
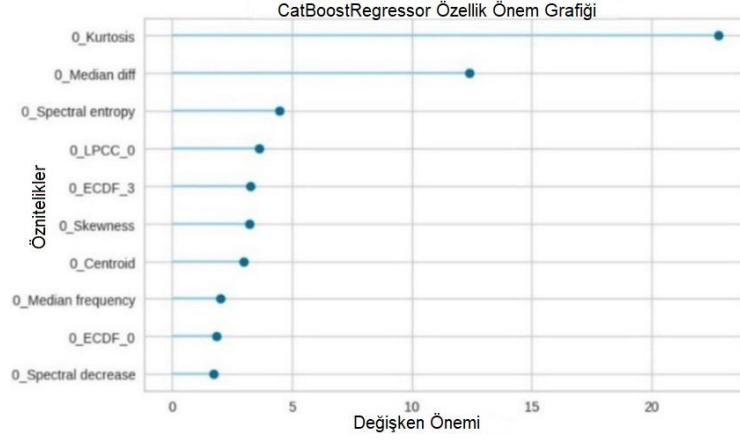
Bu çalışmada performans verilerinden farklı olarak Feature Importance Plot (Özellik Önem Sıralama Grafiği) kullanılmıştır. Özellik Önem Sıralama Grafiği bir

makine öğrenimi modelinde kullanılan özelliklerin, yani değişkenlerin veya özniteliklerin, modelin performansına olan katkısını görselleştirmek için kullanılan bir grafikdir. Bu grafik, her bir özelliğin modelin çıktısını oluştururken ne kadar etkili olduğunu anlamak için kullanılır. Grafik, her bir özelliğin önemini büyüklük sırasına göre sıralar. Şekil 3.7’te performans metriği en yüksek olan CatBoost Regresör algoritmasının Özellik Önem Sıralama Grafiği görülmektedir.



Şekil 3.7. CatBoost Regresör algoritması özellik önem sıralama grafiği

Çalışmada kullanılan performansı yüksek ikinci algoritma olan Light Gradient Boosting Machine algoritmasında da eğitim ve test verileri üzerinde inceleme yapıldı, benzer performans metrikleri sergiledikleri görüldü. Özellik Önem Sıralama Grafiği incelendiğinde ise farklılık olduğu görüldü. CatBoost Regresör algoritması ile Light Gradient Boosting Machine algoritması karşılaştırıldığında da iki regresyon algoritması arasında öznenitelik sıralamasında farklılık olduğu görülmüştür. Şekil. 3.8.’de öznenitelik önem sıralaması farklılıkları görülmektedir.



Şekil 3.8. CatBoost Regresör-LGBM Özellik Önem Sıralama Grafiği karşılaştırması.

İki grafik, CatBoost ve LightGBM regresyon algoritmalarının değişken önem derecelerini göstermektedir. Her iki algoritma da hedef değişkeni tahmin ederken hangi özelliklerin daha önemli olduğunu belirlemek için belirli bir veri seti üzerinde çalışmaktadır. Basıklık (Kurtosis [52]): Her iki modelde de çok önemli bir özellik olarak görülmektedir. Formül 1'de verilen basıklık formülleri bu özelliğin, veri setinde hedef değişkeni tahmin ederken kritik bir rol oynadığını önermektedir.

Medyan Farkı (Median Diff): CatBoost modelinde çok önemli iken, LightGBM modelinde dördüncü sırada yer almaktadır. Bu fark, iki modelin farklı algoritmik yapılarından kaynaklanıyor olabilir. Eğim (Slope [52]): LightGBM'de en önemli özellik olarak kabul edilirken, CatBoost modelinde ilk 10 özellik arasında bile yer almamaktadır. Bu durum, farklı modellerde özelliklerin nasıl değerlendirildiği konusunda bir ipucu verebilir.

Diğer Özellikler: Her iki modelde de farklı özellikler öne çıkmaktadır. Örneğin, 0_Spectral entropy her iki modelde de önemli bulunurken, 0_Spectral decrease sadece CatBoost modelinde önemli bulunmuştur. Bu analiz, her iki modelde de tutarlı bir şekilde önemli olan özellikleri ve modelden modele değişen özellikleri anlamak için yararlıdır. Bu bilgi, model seçimi ve iyileştirme sırasında dikkate alınabilir.

CatBoost Regresör algoritmasında en yüksek belirleyici öznelik sinyalin Kurtosis (Basıklık) değeridir. Basıklık değeri bir veri setinin, genelde normal dağılıma kıyasla daha fazla veya daha az sivriliğini ifade eden bir ölçüdür. Bir veri setinin basıklığı, veri noktalarının dağılımının ne kadar sivrildiğini veya yassılaştığını belirler. Aşağıda Formül 1 ' de basıklık formülü gösterilmektedir[76].

$$\text{Formül 1 : Kurtosis} = \frac{n * \sum_{i=1}^n i \sum (Y_i - \bar{Y})^4}{(\sum_{i=1}^n i (Y_i - \bar{Y})^2)^2}$$

Y_i: Dağılımın i Değişkeni

$\bar{Y}$: Dağılımın Ortalaması

n: Dağılımdaki Değişken Sayısı

Bu formül, dağılımın merkezine göre uç değerlerin etkisini belirlemek için kullanılır ve veri setinin uç olaylara ne kadar duyarlı olduğunu anlamamıza yardımcı olur.

Light Gradient Boosting Machine algoritmasında ise en yüksek belirleyici öznelik Slope (Eğim) fonksiyonudur. Eğim fonksiyonu, bir veri dizisinin eğimini hesaplayan bir fonksiyondur. Eğim, bir veri dizisinin değişim hızını ifade eder ve iki boyutlu uzayda bir çizgi veya eğrinin dikme yönünü belirtir. Aşağıda Formül 2 de eğim formülü gösterilmektedir.[77]

$$\text{Formül 2 : Slope} = ax + b$$

y= Bağımlı Değişken

a= Katsayı

x= Bağımsız Değişken

b= Sabit Değer

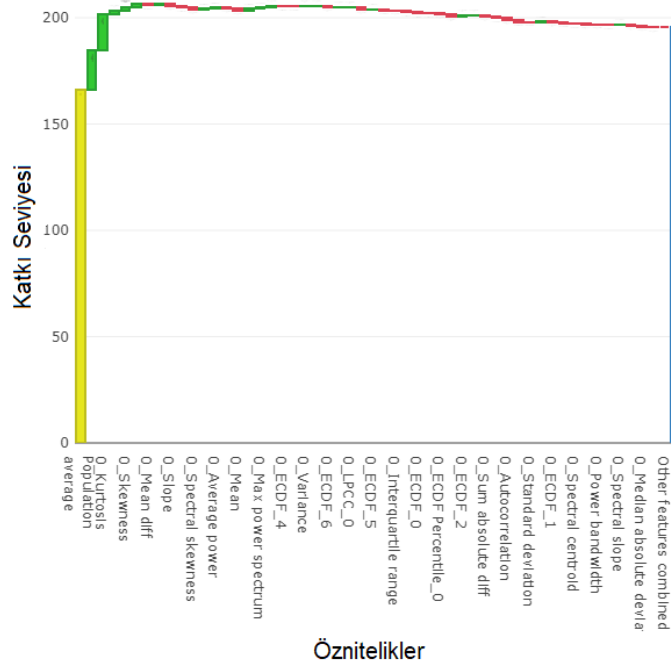
Bu formül, düz bir doğrunun grafik üzerinde nasıl yerleştiğini ve nasıl eğimli olduğunu anlamak için kullanılır.

3.8 Kan Glukoz Seviyesi Tahmini

Çalışmanın amacı, çeşitli regresyon modellerinin performansını değerlendirmek ve daha fazla analiz için en iyi performans gösteren modelleri seçmektir. Farklı metrikleri değerlendirdikten sonra, en yüksek performansa sahip regresyon modellerini belirledik. Bu modeller, en iyi performansı sağlamak için veri seti kullanılarak yeniden eğitildi. Seçilen modellerin performansı hakkında içgörü kazanmak ve hedef değişkeni ne kadar iyi tahmin ettiklerini anlamak için "tahmin hatası" grafiğini kullandık. Bu tür grafik, veri analizi ve makine öğreniminde, modelin ürettiği tahmin edilen değerler ile gerçek değerler arasındaki farklılıkları görselleştirmek için yaygın olarak kullanılır.

Genel olarak, tahmin hatası grafiği, regresyon modellerinin performansını değerlendirmek ve iyileştirme alanlarını belirlemek için görsel bir araç olarak hizmet eder. Araştırmacıların ve analistlerin, modellerin hedef değişkeni tahmin etmedeki güvenilirliği ve etkinliği hakkında bilinçli kararlar almalarına yardımcı olur.

Çalışmada kullanılan tüm öznitelikler hedef değişken ile ilişkilendirildiğinde bazı özniteliklerin sonuca etki etmediği görülmüştür. Bu durum bazı özniteliklerin sinyalin tamamı boyunca değişikliği ugramamasından kaynaklanmaktadır. Kan glukoz seviyesi tahmin edilirken sinyalin tamamında değişikliğe sebep olmayan ve hedef değişken ile korelasyon saptanamayan öznitelikler kullanılmamıştır. Hedef değişken ile korelasyon saptanan öznitelikler eğitimde kullanılmış ve kan glukoz seviyesi tahmin edilmeye çalışılmıştır. Şekil 3.9. hedef değişkenler korelasyon saptanan bazı öznitelikler gösterilmiştir.



Şekil 3.9. Korelasyon saptanan öznitelikler

4. BULGULAR

Çalışmada veriler ilgili ayarlar ve parametreler ile eğitilmiş regresyon modelleri karşılaştırılmıştır. Performans metrikleri olarak Ortalama Mutlak Hata (Mean Absolute Error-MAE), Ortalama Kare Hatası (Mean Squared Error -MSE), Kök Ortalama Kare Hatası (Root Mean Squared Error -RMSE), R-kare (R^2), Kök Ortalama Kare Logaritmik Hata (Root Mean Squared Logarithmic Error-RMSLE), Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error -MAPE) değerleri kullanılmıştır. Tüm regresyon modellerine göre performans metrikleri Tablo 4.1’de görülmektedir.

Tablo 4.1. Regresyon modellerine göre performans metrikleri

Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE	TT(Sn)
CatBoost Regresör	25.16	1591.78	39.13	0.71	0.24	0.18	31.15
Light Gradient Boosting Machine	28.64	1941.95	43.45	0.65	0.26	0.20	0.02
Extra Trees	26.98	2095.21	44.52	0.64	0.27	0.19	0.07
Gradient Boosting Regresör	29.94	2368.13	47.99	0.59	0.28	0.20	0.08
AdaBoost Regresör	36.78	2400.90	48.24	0.58	0.29	0.26	0.04
Random Forest Regresör	29.58	2493.49	48.99	0.56	0.29	0.20	0.10
Extreme Gradient Boosting	30.93	2833.09	52.19	0.52	0.30	0.21	0.21
Lasso Regression	38.30	2967.40	537.92	0.48	0.30	0.24	0.01
Ridge Regression	37.09	2795.83	521.44	0.47	0.30	0.24	0.08
Bayesian Ridge	39.49	3075.62	546.58	0.46	0.31	0.25	0.01
Lasso Least Angle Regression	39.93	3209.94	561.38	0.43	0.32	0.26	0.01
Decision Tree Regresör	31.58	3571.12	572.14	0.42	0.33	0.20	0.01
Elastic Net	43.97	3660.27	597.49	0.34	0.34	0.29	0.01
Linear Regression	41.37	3537.18	578.18	0.29	0.37	0.28	0.01
K Neighbors Regresör	43.58	4315.71	649.02	0.20	0.37	0.31	0.01
Orthogonal Matching Pursuit	60.01	5908.77	757.43	0.01	0.44	0.41	0.01

Huber Regresör	60.40	7068.74	820.82	-0.09	0.51	0.36	0.01
Dummy Regresör	70.61	7049.36	834.95	-0.18	0.48	0.50	0.0070

Daha düşük bir MSE, daha iyi bir tahmin performansını gösterir. MSE açısından en iyi performans gösteren model "CatBoost Regresör" modelidir (1591.78). Daha düşük bir RMSE, daha iyi bir tahmin performansını gösterir. Tablodaki modeller arasında RMSE açısından en iyi performans gösteren model "CatBoost Regresör" modelidir (39.13). Bağımsız değişkenler tarafından bağımlı değişkenin varyansının ne kadarının açıklandığını gösteren R^2 değeri, 1'e ne kadar yakınsa, uyum o kadar iyidir. Tablodaki modeller arasında R^2 açısından en iyi performans gösteren model "CatBoost Regresör" modelidir (0.71). Daha düşük bir RMSLE, daha iyi bir tahmin performansını gösterir. Tablodaki modeller arasında RMSLE açısından en iyi performans gösteren model "CatBoost Regresör" modelidir (0.24). Daha düşük bir MAPE değeri, gerçek değerler ile tahmin edilen değerler arasındaki ortalama mutlak yüzde farkını temsil eder. Tablodaki modeller arasında MAPE açısından en iyi performans gösteren model "CatBoost Regresör" modelidir (0.18). Eğitim Zamanı değerinde daha düşük bir eğitim süresi, model eğitiminin daha hızlı olduğunu gösterir. Tablodaki modeller arasında eğitim süresi açısından en iyi performans gösteren model "CatBoost Regresör" modelidir (31.150 saniye).

Çalışmada esas alınan performans metriği R-kare(R^2) metriğidir. R^2 , regresyon modellerinin performansını değerlendirmek için kullanılan bir istatistiksel ölçüdür. R^2 değeri, bağımlı değişkenin bağımsız değişkenler tarafından ne kadar iyi açıklandığını gösterir. Bu ölçü, 0 ile 1 arasında bir değer alır, 1'e daha yakın değerler daha iyi bir uyumu gösterirken, 0'a yaklaşan değerler daha zayıf bir uyumu ifade eder. R^2 , genellikle aşağıdaki Formül 3 ile ifade edilir[78]:

$$\text{Formül 3 : } R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

y_i = Gerçek Değerler

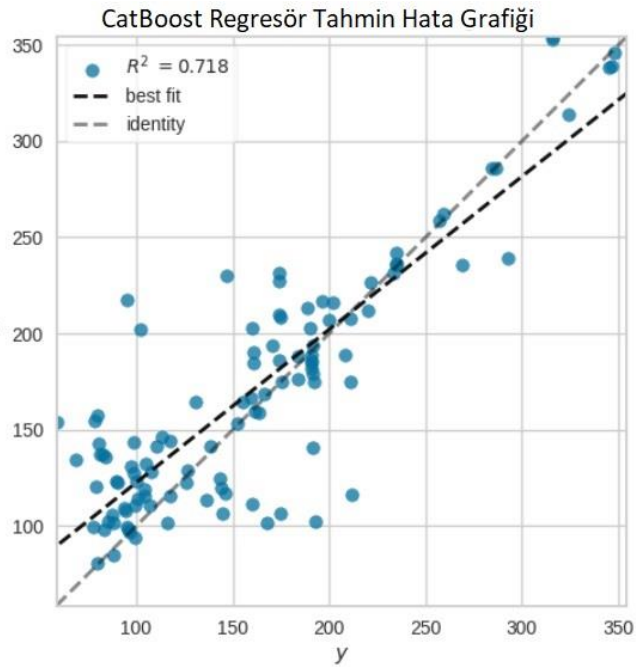
$\hat{y}_i$ = Tahmin Edilen Değerler

$\bar{y}$ = Bağımlı Değişkenin Ortalaması

$\sum_{i=1}^n$ = Toplam

n = Gözlem Sayısı

R^2 metriği için çizilen prediction error (tahmin hata) grafiğinde, x-ekseni gerçek değerleri temsil ederken, y-ekseni tahmin edilen değerleri temsil eder. Her nokta, bir gözlemin gerçek değerini ve modelin bu gözlem için yaptığı tahmini gösterir. İdeali, noktaların çizgi üzerinde olmasıdır. Bu tip bir grafikte, eğrisel bir dağılım veya düzensiz desenler, modelin belirli veri noktalarında hatalar yaptığını gösterebilir. Modelde kullanılan en yüksek performansa sahip CatBoost Regresör modelinin R^2 prediction error grafiği Şekil 4.1’de görülmektedir. R^2 performans metrik değeri 0.718 olarak bulunmuştur.



Şekil 4.1. CatBoost Regresör Modeli R^2 Tahmin Hata Grafiği

R^2 değeri grafiğin sol üst köşesinde 0.718 olarak belirtilmiştir. Bu değer, modelin açıklama gücünü gösterir ve 1'e ne kadar yakınsa modelin o kadar iyi olduğunu belirtir. Bu durumda R^2 değeri 0.718, yani model, veri setindeki varyansın %71,8'ini açıklayabiliyor. Bu, modelin oldukça iyi bir performansa sahip olduğunu gösterir, ancak mükemmel değildir.

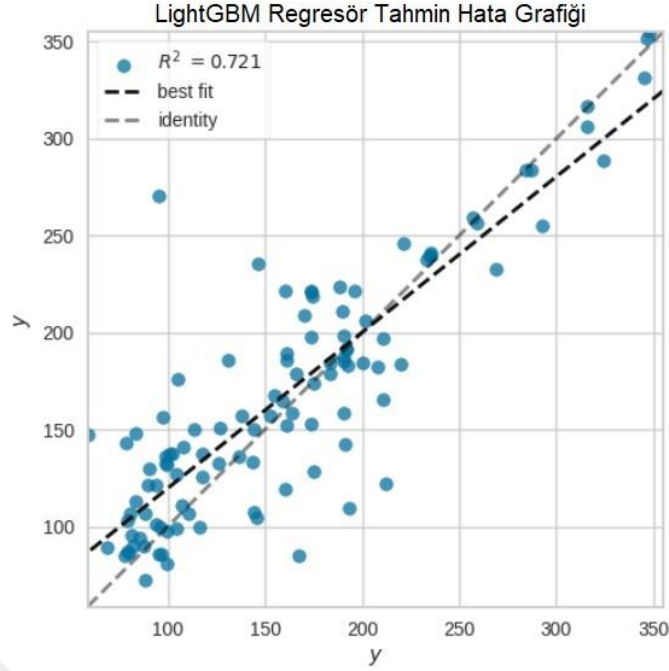
Kimlik Çizgisi (identity): Gri renkle gösterilen kesik çizgi, kimlik çizgisidir. Eğer tüm noktalar bu çizgi üzerinde olsaydı, model mükemmel tahminler yapıyor olurdu. Noktaların bu çizgi etrafında ne kadar yakın veya uzak oldukları, modelin hata miktarını gösterir.

En İyi Uyum Çizgisi (best fit): Siyah kesik çizgi, tahmin edilen ve gerçek değerler arasındaki en iyi uyumu gösterir. Bu çizgi, veri noktalarının genel eğilim yönünü gösterir. Kimlik çizgisine yakın olması, modelin genel olarak doğru yönde tahminler yaptığını gösterir, ancak bazı sapmalar olduğunu da belirtir.

Noktaların Dağılımı: Noktalar genellikle kimlik çizgisine yakın dağılmış, bu da modelin genel olarak iyi tahminler yaptığını gösterir. Ancak, bazı noktalar kimlik çizgisinden oldukça uzak. Bu noktalar modelin ciddi tahmin hataları yaptığı örneklerdir.

Sistemik Hata: Siyah kesik çizginin kimlik çizgisinden sapma derecesi, modelin sistemik hata yapıp yapmadığını gösterir. Çizgiler birbirine paralel değilse veya belirgin bir sapma varsa, bu modelin sistemik olarak bazı bölgelerde daha kötü tahminler yaptığını gösterir. Bu grafikte, en iyi uyum çizgisi kimlik çizgisine oldukça paralel görünüyor, bu da sistemik hatanın fazla olmadığını düşündürür. Genel olarak, bu grafik, CatBoost regresörünün performansının iyi olduğunu, ancak bazı verilerde hatalar yapabildiğini gösterir.

Aynı şekilde modeldeki en yüksek performansa sahip LGBM (Light Gradient Boosting Machine) Regresör algoritmasının prediction error grafiği çizildiğinde benzer sonuçlar verdiği görülmüştür. İki grafikte birbirine yakın sonuçlar vermiştir, aradaki farkın temel nedeni tahmin hata grafiğinde gösterildiği üzere performansa etki eden özniteliklerin iyi algoritmada ayrı olmasıdır. Özniteliklerin performansa etkisini anlamak eğitim algoritmasının seçilmesi ve geliştirilmesi konusunda büyük avantaj sağlamaktadır. Seçilen öznitelikler sonucunda eğitilen sistemin tahmin hata dağılımını çizgi üzerinde tutmak yapılan çalışma için önemlidir. Dağılmış saçılım gösteren grafiklerde özniteliklerin hedef değişkenle olan ilişkisinin zayıf olduğu söylenebilir. Aşağıda Şekil 4.2. 'de LGBM Regresör algoritmasına ait tahmin hata grafiği gösterilmiştir.



Şekil 4.2. LGBM Regresör Modeli R^2 Tahmin Hata Grafiği

LGBM algoritmasına ait tahmin hata grafiğinde mavi noktaların kesikli siyah çizgiye (identity) ne kadar yakın olduğuna bakarak modelin tahminlerinin doğruluğu değerlendirilebilir. Noktalar bu çizgiye ne kadar yakınsa, modelin performansı o kadar iyidir.

R^2 Değeri: 0.721 değeri, modelin tahmin performansının iyi olduğunu, ancak mükemmel olmadığını gösterir. Bu değer, modelin hedef değişkenin varyansının büyük bir kısmını açıklayabildiğini, ancak hala geliştirilmesi gereken bazı alanlar olduğunu gösterir. Mavi noktaların best fit çizgisine olan dağınıklığı, modelin tahminlerinde bazı hatalar olduğunu gösterir. Dağınıklık arttıkça, modelin tahmin hataları da artar.

Bu grafik, LightGBM Regresör modelinin tahminlerinin genel olarak doğru olduğunu, ancak bazı hataların mevcut olduğunu gösterir. Modelin performansı fena değil, ancak bazı iyileştirmeler yapılabilir. R^2 değeri, modelin bağımlı değişkenin varyansının büyük bir kısmını açıklayabildiğini, ancak hala açıklanamayan bazı varyanslar olduğunu belirtir. Modelin performansını artırmak için daha fazla veri, daha iyi özellik mühendisliği gerektiği düşünülmektedir.

5. TARTIŞMA

Makine öğrenmesi, PPG sinyalleri üzerinden kan şekeri seviyelerini tahmin etmek için kullanılan etkili bir yöntemdir. PPG sinyalleri, cilt üzerinden ölçülen ve kalp atışlarına bağlı olarak değişen ışık Emilimi veya yansımasıyla elde edilen bir sinyal türü olduğundan dolayı kan dolaşımındaki değişiklikleri yansıtarak kandaki glukoz seviyeleri hakkında bilgi sağlarlar. Makine öğrenmesi algoritmaları, PPG sinyallerinin karmaşıklığını analiz edebilir ve bu sinyallerden elde edilen verileri kullanarak bireylerin kan şekeri seviyelerini tahmin edebilir. Bu yöntem, geleneksel kan şekeri ölçümlerine alternatif olarak düşünülebilir, çünkü invaziv olmayan bir şekilde çalışır ve bireyin yaşam kalitesini artırabilir. Ancak, bu tür sistemlerin güvenilirliği ve doğruluğu üzerine daha fazla araştırma ve geliştirme gereklidir.

Çalışmanın başarısını benzer çalışmalarla karşılaştırıldığımızda başarı oranının yüksek olduğu görülmektedir. Yapılacak parametre ayarları ve sinyal ölçümü sırasındaki gürültü özelliklerinin bastırılması gibi düzenlemeler yapılarak başarı oranını arttırmak mümkündür.

Takoua HAMDI ve arkadaşlarının 2017 yılında yaptığı Kan Şekeri Seviyesi Tahmini için Yapay Sinir Ağı isimli çalışmada Freestyle Navigator Sürekli Glukoz İzleme Sistemi kullanılarak Fransa'daki Intercommunal Toulon La Seyne Hastane Merkezi'ndeki (CHITS) kan şekeri seviyeleri alınmış ve yapay sinir ağı modelleriyle sistemi eğitmişlerdir. Kan şekeri ölçüm sistemi her 15 dakikada bir tek bir kan şekeri ölçümü üretmiş (günde 96 numune), 12 hasta (insüline bağımlı diyabet hastası) 14 gün boyunca sürekli glukoz izleme sistemini kullanmıştır.[53] Takoua ve arkadaşlarına göre veri setinin bilgilendirici ve zengin olduğu düşünülmektedir. Hasta sayısının az olmasının eksiklerini sürekli glukoz izleme sistemi ile giderileceği tahmin edilmiştir. İlgili çalışmanın %20 sapma payı ile kan şekerini doğru tahmin ettiği belirtilmektedir.

2020 yılında Guillermo ve arkadaşlarının yaptığı Tip I Diyabetli Hastalarda Bazal İnsülin Tahmininde Yapay Zeka Tekniklerinin Uygulanması adlı çalışmada canlıların sinir sisteminde bulunan biyolojik sinir ağlarına benzeyen hesaplamalı yapay sinir ağı modelleri kullanılmıştır. İlgili çalışmada 24 saatinin her biri için bazal insülin dozunu tahmin etmek için dört yapay zeka tekniği araştırılmışlardır

[54]. Bunlar; sinir ağıları (NN), Bayes ağıları (BN), rastgele ormanlar (RF) ve destek vektör makineleridir (SVM). Seçilen yapay zeka teknikleri, makine öğrenme stratejileri olarak bilinir. Bir eğitim setinin parçası olan hastalardan alınan bilgilere dayanarak bir insülin dozu tahmin modeli oluşturuyorlar. Model, insülin doz değerlerini tahmin etme yeteneğine sahiptir. Yapılan çalışmada sinir ağı modelinin R^2 performans metriği 0.067 olarak ölçülürken Bayes ağı modelinin R^2 performans metriği 0.634 olarak ölçülmüştür. Ölçülen performans metrik değerleri makine öğrenmesi modellerinin yapay sinir ağı modellerine göre daha iyi performans verdiğini göstermektedir. Tablo 5.1 diyabet hastalarında glukoz tahmini ve izlenmesi üzerine yapılan çalışmaların karşılaştırmalı analizini ve önerilen bir modeli sunmaktadır.

Tablo 5.1. Literatürdeki çalışmalar

Referans	Çalışma Konusu	Sınıflandırıcı	Performans
Geshwaree Huzooree et al. – 2017[55]	Diyabetik Hastaların İzlenmesi için Glukoz Tahmini Veri Analitiği	Otoregresif (ARX)	$R^2 = 0.342$
Guillermo Edinson et all. – 2020 [56]	Tipi 1 Diyabetli Hastalarda Bazal İnsülin Tahmini için Yapay Zeka Tekniklerinin Uygulanması	YSA ve BN'ler	YSA $R^2 = 0.067$, BN'ler $R^2 = 0.634$
Nanayakkara et al. - 2018 [57]	Hibrit Teknik Kullanarak Non- invaziv Kan Glukozu Monitörizasyonu	BIS + NIRS	$R^2 = 0.58$
Rui Sun et al. – 2023 [58]	Tipi 2 Diyabet Hastalarında Zaman Aralığı Tahmini, Açlık ve Sonrası Glukoz Düzeylerinin Dahil Edilmesiyle İyileştirilir	Doğrusal Regresyon	$R^2 = 0.36$
Chowdhury Azimul Haque et al. – 2020 [59]	Noninvaziv İn Vivo Kan-Glukoz Konsantrasyonunun Monte Carlo Simülasyonu ile Tahmini	XGBoost	$R^2 = 0.68$
Pappada, S. M. et al. -2011 [60]	İnsüline Bağımlı Diyabet Hastalarında Gerçek Zamanlı Glukoz Tahmini İçin Sinir Ağı Tabanlı	Sinir Ağı	RMSE= 43.9
Aishah, A. F. Q. A. et al. - 2011 [61]	Sağlıklı Bireylerde Açlık Kan Glukoz Düzeyini Tahmin Etmek İçin Çoklu Doğrusal Regresyon Modeli Analizi	Çoklu Doğrusal Regresyon	$R^2=0.59$
Önerdiğim Yöntem	Yapay Zeka Tabanlı Non-Invaziv Kan Şekeri Ölçümü	CatBoost Regresör	$R^2=0.71$, RMSE= 39.13

Geshwaree Huzooree ve ark. (2017), otoregresif (ARX) bir model kullandı ve 0.342 R^2 elde etti, bu da orta derecede doğruluk olduğunu gösterir. Guillermo Edinson ve ark. (2020), Yapay Zeka tekniklerini uyguladı, Yapay Sinir Ağları (YSA) zayıf performans gösterirken ($R^2 = 0.067$), Bayesyen Ağlar (BNs) iyi performans gösterdi ($R^2 = 0.634$). Nanayakkara ve ark. (2018), hibrit bir teknik (BIS + NIRS) kullandı ve 0.58 R^2 ile iyi doğruluk elde etti. Rui Sun ve ark. (2023), zaman aralığı tahmini için doğrusal regresyon kullandı ve 0.36 R^2 elde etti. Chowdhury Azimul Haque ve ark. (2020), en yüksek R^2 olan 0.68'i elde etmek için XGBoost kullandı. Pappada ve ark. (2011), sinir ağlarını kullanarak 43.9 RMSE elde etti ve Aishah ve ark. (2011), çoklu doğrusal regresyon ile 0.59 R^2 elde etti. Önerilen model, CatBoost Regresör kullanarak, diğerlerinin üstünde bir R^2 olan 0.71 ve 39.13 RMSE ile daha üstün tahmin doğruluğu ve daha düşük hata göstermektedir.

6. SONUÇ VE ÖNERİLER

Makine öğrenmesi modelleri, sinyal işleme uygulamalarında çeşitli istatistiksel özniteliklerin (features) önemini kullanarak daha doğru ve güvenilir tahminler yapabilirler. İstatistiksel öznitelikler, sinyal verisinin önemli özelliklerini yakalayıp modeli besler ve böylece modelin performansını artırır.

Özelliklerin Çıkarılması ve Temsili: İstatistiksel öznitelikler, sinyalin temel özelliklerini temsil eder. Ortalama, standart sapma, tepe değerleri gibi metrikler, sinyalin genel yapısını özetler. Bu öznitelikler, ham sinyal verisini daha kompakt ve anlamlı bir şekilde temsil ederek, modelin veriyi daha iyi anlamasını sağlar.

Model Performansını Artırma: İyi seçilmiş öznitelikler, modelin performansını büyük ölçüde artırır. Doğru öznitelikler, modelin öğrenme sürecini hızlandırır ve daha isabetli tahminler yapmasına yardımcı olur. Örneğin, sinyalin frekans bileşenleri, genlik varyasyonları veya trendleri gibi öznitelikler, sinyalin karakteristiklerini yakalayıp modelin tahmin yeteneğini geliştirir.

Gürültüyü Azaltma ve Anomalileri Belirleme: İstatistiksel öznitelikler, sinyaldeki gürültüyü azaltarak anlamlı bilgileri öne çıkarır. Aynı zamanda, anomali tespiti gibi uygulamalarda önemli olan aşırı değerler veya beklenmedik değişiklikler, bu öznitelikler sayesinde daha kolay belirlenebilir.

Veri Boyutunu Azaltma: Ham sinyal verisi genellikle yüksek boyutludur ve doğrudan işlenmesi zor olabilir. İstatistiksel öznitelikler, bu veriyi özetleyerek boyutunu azaltır ve modelin daha verimli çalışmasını sağlar.

Özelliklerin Seçimi ve Model Uyumu: Farklı istatistiksel öznitelikler, farklı modellerde farklı derecelerde önemli olabilir. Örneğin, CatBoost ve LightGBM gibi modellerde farklı özniteliklerin önem derecesi değişebilir. Bu nedenle, öznitelik seçimi, model uyumu ve performansı açısından kritik bir rol oynar.

Sonuç olarak, sinyal işleme uygulamalarında kullanılan istatistiksel öznitelikler, makine öğrenmesi modellerinin etkinliğini ve doğruluğunu artırmak için hayati öneme sahiptir. Bu öznitelikler, sinyal verisinin önemli özelliklerini yakalayarak modelin daha iyi öğrenmesine ve tahmin yapmasına yardımcı olur.

Bu çalışmada PPG sinyalleri kullanılarak kan glukoz değeri tahmin edilmeye çalışılmıştır. Sinyaller ön işlemlerden geçirilmiş, gürültü ve yapaylık özelliklerinden arındırılmıştır. Ön işlemlerden geçirilen sinyaller segmentasyon işlemine tabi tutulmuştur. Tepe (peak) noktalarına göre bölüntülenen ve ortalaması alınan sinyaller öznitelik çıkarma işlemi için hazır hale getirilmiştir.

Ön işlemlerden geçirilmiş ve segmente edilmiş sinyallere, veri setindeki önemli bilgileri belirlemek ve temsil etmek amacıyla öznitelik çıkarma işlemi uygulanmıştır. Bu işlem veri setlerindeki karmaşıklığı azaltarak, işleme daha uygun ve anlamlı bir temsili elde etmeyi hedefler. Öznitelik çıkarma işleminde istatistiksel, temporal ve spektral yöntemler kullanılmaktadır. Bu çalışmada istatistiksel ve temporal öznitelik çıkarma işlemi kullanılmıştır. Bu işlem veri setleri üzerinde çeşitli istatistiksel ve zamansal yöntemler kullanılarak verilerin özellikleri, dağılımları, eğilimleri analiz etmektedir.

Yapılan çalışmada PPG sinyallerinden 160 adet istatistiksel alan özniteliği çıkarılmış ve bunlardan hedef değişken ile arasında bağ kurulan 51 adet öznitelik seçilmiştir. Ön işlemlerden geçirilen ve segmente edilmiş ilgili öznitelik alanlarına ayrılmış sinyaller makine öğrenmesi modelleri ile eğitilmiştir. Sistem regresyon modellerinde başarı göstermiştir ve R-kare performans metrik değeri 0.718 olarak ölçülmüştür.

Bu çalışma açıklanabilirlik ve yorumlanabilirlik açısından önemli özellikler barındırmaktadır. Makine öğrenmesi modellerinden yüksek performansa sahip iki model açıklanabilir yapay zeka teknikleri (SHAP) ile incelendiğinde hem benzer hem de farklı öznitelikler olduğu görülmüştür.

Çalışmada performans verilerinden farklı olarak Feature Importance Plot (Özellik Önem Sıralama Grafiği) kullanılmıştır. Feature Importance Plot grafiği özelliklerin, yani değişkenlerin veya özniteliklerin, modelin performansına olan katkısını görselleştirmek için seçilmiştir.

Performansa en çok etki eden özelliklerin iki makine öğrenmesi modelinde de farklı sıralamaya sahip olduğu görülmüştür. Öznitelik önem sırasının incelenmesinin performansa katkısının büyük olduğu belirlenmiştir.

Sinyal özniteliklerinin sadece istatistiksel ve temporal alan özellikleri ile değil frekans alanı özellikleri ile değerlendirilmesinin sonuca pozitif olarak etki edeceği belirtilmiştir. Ayrıca örnekleme sayısının arttırılması özniteliklerin sonucu etkilemesinde büyük orana sahiptir bu yüzden örnekleme sayısının arttırılması çalışmayı derinleştirecektir. Son olarak PPG sinyalinin kaydedilmesi esnasından sinyalin gürültüden arındırılması ön işleme de kolaylık oluşturacak ve eğitim verilerinin daha sağlıklı çalışmasını sağlayacaktır. Ölçümlerin her kişide aynı sürede alınması özniteliklerin alınması ve hesaplanması esnasında etkili olduğu için doğrudan hedef değişkenler ilişkili olduğu düşünülmektedir.

Çıkarılan özniteliklerin sayısının az, hedef değişkenle olan ilişkisinin yüksek olduğu alan özelliklerinin olması ileride yapılacak IOT (Nesnelerin İnterneti) çalışmalarında kullanılmak üzere gömülü sistemlere entegre edilmesi açısından önem arz etmektedir.

7. KAYNAKLAR

Bu tez çalışmasında Vancouver atıf sistemi kullanılmıştır.

1. International Diabetes Federation. (IDF 2009c). Diabetes Prevalence.<http://www.idf.org/home/index.cfm?node=264>(erişim:12.05.2009)
2. Diyabet Çığ Gibi Büyüyor. (2008). Diyabete Bakış, 7, 6-7
3. K. Wikblad, L. Wibell, , K. Montin (1990). “The Patient's Experience Of Diabetes And Its Treatment: Construction Of An Attitude Scale By A Semantic Differential Technique.” Journal Of Advanced Nursing, 15(9), 1083-1091.
4. American Diabet Association, “Diagnosis and classification of diabetes mellitus”, Diabetes Care, 32 (1), 62–67, 2009.
5. American Diabetes Association. (2023). Standards of Medical Care in Diabetes—2023. Diabetes Care, 46(Suppl. 1): S1-S291.
6. Gusev, M., Poposka, L., Spasevski, G., Kostoska, M., Koteska, B., Simjanoska, M., Ackovska, N., Stojmenski, A., Tasic, J., Trontelj, J. 2020. Noninvasive Glucose Measurement Using Machine Learning and Neural Network Methods and Correlation with Heart Rate Variability. Journal of Sensors. <https://doi.org/10.1155/2020/9628281>
7. Sharma, N. K., Singh, S. 2012. Designing a non invasive blood glucose measurement sensor. 2012 IEEE 7th International Conference on Industrial and Information Systems, ICIIS 2012 : 2012 IEEE 7th International Conference on Industrial and Information Systems, ICIIS 2012. <https://doi.org/10.1109/ICIInfS.2012.6304818>
8. Zhao, X., Zheng, Q., Yang, Z. M. 2016. Two types of photonic crystals applied to glucose sensor. . <https://doi.org/10.1109/inec.2016.7589369>
9. Tanaka, Y., Purtil, C., Tajima, T., Seyama, M., Koizumi, H. 2017. Sensitivity improvement on CW dual-wavelength photoacoustic spectroscopy using acoustic resonant mode for 79 noninvasive glucose monitor. Proceedings of IEEE Sensors : Proceedings of IEEE Sensors. <https://doi.org/10.1109/ICSENS.2016.7808685>
10. Kossowski, T., Stasinski, R. 2016. Robust IR attenuation measurement for non-invasive glucose level analysis. International Conference on Systems, Signals, and Image 75 Processing : International Conference on Systems, Signals, and Image Processing (Vol. 2016-June). <https://doi.org/10.1109/IWSSIP.2016.7502770>
11. Beck, R. W., Riddlesworth, T., Ruedy, K., & Kollman, C. (2018). Continuous glucose monitoring versus usual care in patients with type 2 diabetes receiving multiple daily insulin injections: a randomized trial. Annals of Internal Medicine, 169(6), 379-387.
12. Karon, B. S. (2016). Why is there still no international standard for glucose in blood? Clinical Chemistry and Laboratory Medicine (CCLM), 54(6), 975-977.
13. Newman, J. D., & Turner, A. P. (2005). Home blood glucose biosensors: a commercial perspective. Biosensors and Bioelectronics, 20(12), 2435-2453.
14. Sieg, A., Guy, R. H., & Delgado-Charro, M. B. (2012). Non-invasive and minimally invasive methods for measuring glucose. In Advances in Noninvasive Electrocardiographic Monitoring Techniques (pp. 259-277). Springer, Berlin, Heidelberg.
15. Lee, Y. H., Wong, D. T., & Saliva: An emerging biofluid for early detection of diseases.

- (2009). *American Journal of Dentistry*, 22(4), 241-248.
16. Rohleder, D., von Ribbeck, H. G., Brenner, T., Giessen, H., & Lemmer, U. (2012). Noninvasive continuous glucose monitoring by photoacoustic spectroscopy. *Analytical Chemistry*, 84(16), 6558-6565.
 17. Lukaski, H. C. (1987). Methods for the assessment of human body composition: traditional and new. *American Journal of Clinical Nutrition*, 46(4), 537-556.
 18. Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3), R1-R39.
 19. Kumar, S., Buckley, J., Barton, J., Pigeon, M., Newberry, R., Rodencal, M., Hajzeraj, A., Hannon, T., Rogers, K., Casey, De., O'sullivan, D., O'Flynn, B., (2020). A Wristwatch-Based Wireless Sensor Platform for IoT Health Monitoring Applications. *Sensors*. 20. 10.3390/s20061675.
 20. Yang, Y., Beh, W., Lo, Y., Wu, A.A., & Lu, S. (2020). ECG-Aided PPG Signal Quality Assessment (SQA) System for Heart Rate Estimation. 2020 IEEE Workshop on Signal Processing Systems (SiPS), 1-6.
 21. Shokrehodaie, M., Quinones, S. 2020. Review of non - invasive glucose sensing techniques: Optical, electrical and breath acetone. *Sensors (Switzerland)*. <https://doi.org/10.3390/s20051251>
 22. Arikan, K., Burhan, H., Sahin, E., Sen, F., A sensitive, fast, selective, and reusable enzyme-free glucose sensor based on monodisperse AuNi alloy nanoparticles on activated carbon support, *Chemosphere*, 291, 3 2022 , <https://doi.org/10.1016/j.chemosphere.2021.132718>
 23. Karimi, F., Zare, N., Bekmezci, M., Akin, M., Bayat, R., Seyitoğlu, B., Arikan, K., Isik, I., Sen, F., Enzyme-free glucose detection via scalable and economical fabrication of nickel-polyvinylpyrrolidone-modified multi-walled carbon nanotubes, *Elektrochimica Acta* , 496 , 8 2024 <https://doi.org/10.1016/j.electacta.2024.144341>
 24. Baghelani, M., Abbasi, Z., Daneshmand, M., Light, P. E. 2020. Non-invasive continuous-time glucose monitoring system using a chipless printable sensor based on split ring microwave resonators. *Scientific Reports* 2020 10:1, 10(1):, 1–15. <https://doi.org/10.1038/s41598-020-69547-1>
 25. Tanaka, Yujiro, Tajima, T., Seyama, M., Waki, K. 2020. Differential Continuous Wave Photoacoustic Spectroscopy for Non-Invasive Glucose Monitoring. *IEEE Sensors Journal*, 20(8):. <https://doi.org/10.1109/JSEN.2019.2962251>
 26. Yahyaoui, A., Jamil, A., Rasheed, J., & Yesiltepe, M. (2019). A decision support system for diabetes prediction using machine learning and deep learning techniques. Paper presented at the 2019 1st International Informatics and Software Engineering Conference (UBMYK).
 27. Daskalaki, E., Prountzou, A., Diem, P., & Mougiakakou, S. G. (2012). Real-time adaptive models for the personalized prediction of glycemic profile in type 1 diabetes patients. *J Diabetes technology therapeutics*, 14(2), 168-174.
 25. Tanaka, Yujiro, Tajima, T., Seyama, M., Waki, K. 2020. Differential Continuous Wave Ph
 26. Yahyaoui, A., Jamil, A., Rasheed, J., & Yesiltepe, M. (2019). A decision support system for diabetes prediction using machine learning and deep learning techniques. Paper presented at the 2019 1st International Informatics and Software Engineering Conference (UBMYK).
 27. Daskalaki, E., Prountzou, A., Diem, P., & Mougiakakou, S. G. (2012). Real-time adaptive models for the personalized prediction of glycemic profile in type 1 diabetes patients. *J Diabetes technology therapeutics*, 14(2), 168-174.
 28. Pala, T., & Yücedağ, İ. (2016). Veri Madenciliği Tekniklerinden Sınıflandırma Kullanılarak

Tip 2 Diyabet Tanısı. Paper presented at the International Artificial Intelligence and Data Processing Symposium.

29. Wiley, M. T. (2011). Machine learning for diabetes decision support. Ohio University.
30. Sun, Q., Jankovic, M. V., Bally, L., & Mougiakakou, S. G.(2018). Predicting blood glucose with an LSTM and BiLSTM based deep neural network. Paper presented at the 2018 14th Symposium on Neural Networks and Applications (NEUREL).
31. Pérez-Gandía, C., Facchinetti, A., Sparacino, G., Cobelli, C., Gómez, E., Rigla, M., Hernando, M. (2010). Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring. *J Diabetes technology therapeutics*, 12(1), 81-88
32. Midroni, C., Leimbigler, P., Baruah, G., Kolla, M., Whitehead, A., & Fossat, Y. (2018). Predicting glycemia in type 1 diabetes patients: experiments with xg-boost. Paper presented at the KHD@ IJCAI.
33. Xie, J., & Wang, Q. (2018). Benchmark machine learning approaches with classical time series approaches on the blood glucose level prediction challenge. Paper presented at the KHD@ IJCAI
34. Gupta, S.S., Kwon, T., Hossain, S., & Kim, K. (2021). Towards non-invasive blood glucose measurement using machine learning: An all-purpose PPG system design. *Biomed. Signal Process. Control.*, 68, 102706.
35. Susana, E.; Ramli, K.; Murfi, H.; Aprianoro, N.H. Non-Invasive Classification of Blood Glucose Level for Early Detection Diabetes Based on Photoplethysmography Signal. *Information* 2022, 13, 59. <https://doi.org/10.3390/info13020059>
36. Castro P., Luis T. A., Adriana C.R., Oscar D.H., (2023). Three-wavelength PPG and support vector machine for non-invasive estimation of blood glucose. *Optical and Quantum Electronics*. 55. 10.1007/s11082-023-04927-1.
38. M. Elgendi, 2012. On the Analysis of Fingertip Photoplethysmogram Signals, *Current Cardiology Reviews*, 8, 14-25, Available: 10.2174/157340312801215782
39. M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, H. Gamboa. TSFEL: Time Series Feature Extraction Library, *SoftwareX* 11 (2020). <https://doi.org/10.1016/j.softx.2020.100456>.
40. Fraunhofer Portugal AICOS. (n.d.). TSFEL: Time Series Feature Extraction Library. GitHub Repository. Retrieved from <https://github.com/fraunhoferportugal/tsfel>
41. DeCarlo, L. T. (1997). On the meaning and use of kurtosis and skewness. *Psychological Methods*, 2(3), 292-307.
42. R. M. Kimmel, S. I. Rubin. (1997). Crest Factor. CRC Press.
43. Ali, M. (2020). PyCaret: An open-source, low-code machine learning library in Python. Available at: <https://pycaret.org>.
44. T. Nguyen, N. Tran, B. M. Nguyen, G. Nguyen, 2018. A resource usage prediction system using functional-link and genetic algorithm neural network for multivariate cloud metrics. In 2018 IEEE 11th conference on service-oriented computing and applications (SOCA), 49–56. IEEE,
45. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer Science & Business Media.
46. Root Mean Square Value A Dictionary of Physics (6 ed.). Oxford University Press. 2009. ISBN 9780199233991
47. Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied Linear Statistical*

Models. McGraw-Hill/Irwin.

48. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
49. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679-688.
50. Anna V. D., Andrey G., Gleb G., Nikita K., Liudmila O. P., Aleksandr V., "Fighting biases with dynamic boosting". arXiv:1706.09516, 2017.
51. Qi M., Guolin K., Taifeng W., Wei C., Qiwei Y., Zhi-Min M., Tie-Yan L., "A Communication-Efficient Parallel Algorithm for Decision Tree". *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, pp. 1279-1287.
52. Moore, D. S., McCabe, G. P., & Craig, B. A. (2019). *Introduction to the Practice of Statistics*. W. H. Freeman.
53. Ali, J.B., Hamdi, T., Fnaiech, N., Costanzo, V.D., Fnaiech, F., & Ginoux, J. (2018). Continuous blood glucose level prediction of Type 1 Diabetes based on Artificial Neural Network. *Biocybernetics and Biomedical Engineering*, 38, 828-840.
54. Guzman G., G. E., Burbano A., L. E., Martínez, V., & Bedoya Leiva, O. F. (2020). Application of Artificial Intelligence Techniques for the Estimation of Basal Insulin in Patients with Type I Diabetes. *International journal of endocrinology*, 2020, 7326073. <https://doi.org/10.1155/2020/7326073>
55. G. H Huzooree, K.K. Khedo, N. Joons, 2017. Glucose prediction data analytics for diabetic patients monitoring, 2017 1st International Conference on Next Generation Computing Applications (NextComp) 10.1109/NEXTCOMP.2017.8016197.
56. G. Edinson, G. Go´mez , L. E. B. Agredo, V. Mart´inez, O. F. B. Leiva. Application of Artificial Intelligence Techniques for the Estimation of Basal Insulin in Patients with Type I Diabetes, *Research Article Open Access* <https://doi.org/10.1155/2020/7326073>
57. N.D Nanayakkara, S.C Munasingha, G.P. Ruwanpathirana, 2018. Non-Invasive Blood Glucose Monitoring using a Hybrid Technique, In *Proceedings of the 2018 Moratuwa Engineering Research Conference (MERCon)*, Moratuwa, Sri Lanka, pp. 7–12. , 10.1109/MERCon.2018.8421885
58. R. Sun, Y.L. Duan, Y.M. Zhang, L.G. Feng, B. Ding, R.N. Yan, J.H. Ma, X.F. Su. Time in Range Estimation in Patients with Type 2 Diabetes is Improved by Incorporating Fasting and Postprandial Glucose Levels, *DIABETES THERAPY*, Volume 14 Issue 8 Page 1373-1386, DOI 10.1007/s13300-023-01432-2
59. S. S. Gupta, S. Hossain, C. A. Haque and K. -D. Kim, 2020. In-Vivo Estimation of Glucose Level Using PPG Signal, 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2020, pp. 733-736, doi: 10.1109/ICTC49870.2020.9289629
60. Pappada, S. M., Cameron, B. D., Rosman, P. M., Bourey, R. E., Papadimos, T. J., Olorunto, W., & Borst, M. J. (2011). Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes. *Diabetes technology & therapeutics*, 13(2), 135-141.
61. CatBoost Documentation Online : <https://catboost.ai/en/docs/>
62. LightGBM Documentation Online : <https://lightgbm.readthedocs.io/en/stable/>
63. Scikit-learn Extra Trees Documentation Online : <https://scikit-learn.org/stable/>
64. Scikit-learn Gradient Boosting Documentation Online: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostRegressor.html>

65. Scikit-learn AdaBoost Documentation Online: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostRegressor.html>
66. Scikit-learn Random Forest Documentation Online: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>
67. XGBoost Documentation Online : <https://xgboost.readthedocs.io/en/stable/>
68. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2010;33(Suppl 1)
69. American Diabetes Association. Standards of medical care in diabetes—2022. Diabetes Care. 2022;45(Suppl 1)
70. National Institute of Diabetes and Digestive and Kidney Diseases. Insulin Resistance and Prediabetes [Internet]. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health; 2018 [cited 2024 Jul 23]. Available from: <https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes/prediabetes-insulin-resistance>
71. Diyabetim.net. Kan Şekeri Ölçüm Cihazları [Internet]. Available from: <https://diyabetim.net/tag/kan-sekeri-olcum-cihazi/>
72. Soy Medikal. Kan Şekeri Ölçüm Cihazları [Internet]. Available from: <https://www.soymedikal.com.tr/sayfalar-24-Nas>
73. Kumar S, Buckley J, Barton J, Pigeon M, Newberry R, Rodencal M, Hajzeraj A, Hannon T, Rogers K, Casey D, O'Sullivan D, O'Flynn B. A wristwatch-based wireless sensor platform for IoT health monitoring applications. Sensors. 2020;20(6):1675. doi:10.3390/s20061675.
74. Richtek Technology Corporation. Medical & Healthcare Solutions [Internet]. Available from: <https://www.richtek.com/Design%20Support/Technical%20Document/AN057>
75. Türk Yükseköğrenim ve Üniversite Derneği. 23-24 Ocak Temel YB Kursu Sunumları. İstanbul: Türk Yükseköğrenim ve Üniversite Derneği; 2021 [Erişim 2024 Temmuz 23]. Erişim adresi: <https://www.tuyud.org.tr/wp-content/uploads/2021/01/23-24-Ocak-Temel-YB-Kursu-Sunumlari1-sikistirildi.pdf>
76. EDUCBA. Kurtosis Formula. [Internet]. [Erişim 2024 Temmuz 23]. Available from: <https://www.educba.com/kurtosis-formula/>
77. Study.com. Graphing a Function of the Form $f(x) = ax + b$ as an Integer Slope Explanation. [Internet]. [Place unknown]: Study.com; [Erişim 2024 Temmuz 23]. Available from: <https://study.com/skill/learn/graphing-a-function-of-the-form-f-x-ax-plus-b-as-an-integer-slope-explanation.html>
78. Faramarzi I. What is R^2 and How It Gets Negative. [Internet]. : Medium; 2024 [Erişim 2024 Temmuz 23]. Available from: <https://medium.com/@iliyafaramarzi1384/what-is-r2-and-how-it-gets-negative-e927011d92ed>
79. Analog Devices. MAX30105: Integrated Particle Sensor. [Internet]. [Place unknown]: Analog Devices; [No date] [cited 2024 Jul 23]. Available from: <https://www.analog.com/en/products/max30105.html>

8. EKLER

EK 1 Etik Kurul Onayı



BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ KLİNİK ARAŞTIRMALAR ETİK KURULU ONAYI
BOLU ABANT İZZET BAYSAL UNIVERSITY CLINICAL RESEARCHES ETHICS COMMITTEE APPROVAL

Sayı : 448
Konu: Kararlar

25.10/2023

BAŞVURU BİLGİLERİ (APPLICATION INFORMATION)	ARAŞTIRMANIN ADI (TITLE OF THE PROJECT)	Kan şekeri ölçümü için optik non-invaziv yapay zekaya dayalı akıllı ölçüm sistemi tasarımı.
	ARAŞTIRMANIN İNGİLİZCE ADI (TITLE OF THE PROJECT)	Optical non-invasive artificial intelligence based intelligent measurement system design for blood glucose measurement.
	SORUMLU ARAŞTIRMACI (PRINCIPAL INVESTIGATOR)	Prof.Dr.Kemal POLAT
	DİĞER ARAŞTIRMACILAR (OTHER INVESTIGATORS)	Dr.Öğr.Üyesi Ümit ŞENTÜRK, Dr.Züleyha ERDİN, Yük.Lis.Öğr.Gökhan ADIGÜZEL
	ARAŞTIRMA MERKEZİ (RESEARCH CENTER)	Bolu Abant İzzet Baysal Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü

KARAR (DECISION)	Karar no (Decision No): 2023/284	Tarih (Date) 10.10.2023
	Prof.Dr.Kemal POLAT'ın sorumluluğunda yapılması tasarlanan ve yukarıda başvuru bilgileri verilen araştırma dosyası ve ilgili belgelerin incelenmesi sonucunda araştırmanın gerçekleştirilmesinde etik yönden sakınca olmadığına mevcudun oy birliği/oy çokluğu ile karar verilmiştir.	

Üyeler	Uzmanlık alanı	Kurumu
KVKK kapsamında gizlenmiştir.	Genel Cerrahi	BAİBÜ Tıp Fakültesi
	Ruh Sağlığı Hastalıkları	BAİBÜ Tıp Fakültesi
	Farmakoloji	BAİBÜ Tıp Fakültesi
	Çocuk Sağlığı Hastalıkları /Nefroloji Bilim Dalı	BAİBÜ Tıp Fakültesi
	Anesteziyoloji ve Reanimasyon	BAİBÜ Tıp Fakültesi
	İç Hastalıkları Anabilim Dalı	BAİBÜ Tıp Fakültesi
	Protetik Diş Tedavisi	BAİBÜ Diş Hekimliği Fakültesi
	Hemşirelik Bölümü	BAİBÜ Sağlık Bilimler Fakültesi
	Antrenörlük Eğitimi	BAİBÜ Spor Bilimleri Fakültesi
	Halk Sağlığı	BAİBÜ Tıp Fakültesi
	Farmakolog/Eczacı	Özel Eczane (BOLU)
	Fizyoloji	BAİBÜ Tıp Fakültesi
	Avukat	Özel Hukuk Bürosu (BOLU)
	Esnaflık	Serbest Meslek (BOLU)

KVKK kapsamında gizlenmiştir.