



TÜBİTAK 2209-A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEKLEME PROGRAMI

Başvuru formunun Arial 9 yazı tipinde, her bir konu başlığı altında verilen açıklamalar göz önünde bulundurularak hazırlanması ve ekler hariç toplam 20 sayfayı geçmemesi beklenir (Alt sınır bulunmamaktadır). Değerlendirme araştırma önerisinin bilimsel niteliği, yöntemi, proje yönetimi ve yaygın etkisi başlıkları üzerinden yapılacaktır.

ARAŞTIRMA ÖNERİSİ FORMU

2025 Yılı

Güz Dönem Başvurusu

2209-A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEKLEME PROGRAMI
ARAŞTIRMA ÖNERİSİ FORMU

A. GENEL BİLGİLER

Başvuru Sahibinin Adı Soyadı: HÜSEYİN KAYA
Araştırma Önerisinin Başlığı: Safra Taşı Oluşumundaki Kritik Farklılıkların Belirlenmesi Amacıyla Benzer Sağlık Profillerinin k-En Yakın Komşu (k-NN) Algoritması ile Analizi
Danışmanın Adı Soyadı: PINAR KARADAYI ATAŞ
Araştırmanın Yürütüleceği Kurum/Kuruluş: İSTANBUL AREL ÜNİVERSİTESİ

ÖZET

Araştırma önerisi özetinin (1) bilimsel nitelik; (2) yöntem; (3) proje yönetimi ve (4) yaygın etki hakkında bilgileri kapsamı beklenir. Bu bölümün en son yazılması önerilir.

Özet

Bu araştırma, bireylerin demografik, biyoimpedans ve laboratuvar verilerini kullanarak safra taşı hastalığı riskini öngörebilen makine öğrenmesi tabanlı bir tahmin modeli geliştirmeyi amaçlamaktadır. Çalışmada, Ankara VM Medical Park Hastanesi Dahiliye Polikliniği'nden 2022–2023 yılları arasında toplanan ve 38 özellik içeren etik onaylı veri seti kullanılmıştır. Veri seti, eksiksiz ve dengeli olması bakımından istatistiksel analiz ve modelleme çalışmaları için uygun bir yapıya sahiptir.

Bilimsel açıdan proje, klinik görüntüleme yöntemlerine (örneğin ultrasonografi) başvurmadan yalnızca biyolojik ölçümler ve temel kan parametreleriyle safra taşı riskini belirleyebilmeyi hedefleyen non-invaziv bir yaklaşım önermektedir. Literatürde, safra taşı hastalığının dünya genelinde artan bir sağlık yükü oluşturduğu ve obezite, insülin direnci, yaş, cinsiyet ve yaşam tarzı gibi faktörlerle güçlü bir ilişkisi olduğu vurgulanmaktadır (World Gastroenterology Organisation, 2024; Jones, 2024). Bu proje, bu risk faktörlerini makine öğrenmesi algoritmalarıyla ilişkilendirerek erken tanı olasılığını artırmayı ve klinisyenlere karar destek mekanizması sunmayı hedeflemektedir.

Çalışmada kullanılacak k-NN, SVM, Random Forest, Gradient Boosting ve Lojistik Regresyon algoritmaları, veri setinin sayısal, eksiksiz ve dengeli yapısı ile yüksek düzeyde uyumludur. Özellikle SVM ve Gradient Boosting modelleri, düşük örneklem sayısına rağmen karmaşık ilişkileri modelleyebilme yeteneğiyle öne çıkmaktadır. MLP ve Naïve Bayes algoritmaları ise karşılaştırmalı analiz amacıyla kullanılacaktır. Veri boyutu ve model karmaşıklığı göz önünde bulundurularak aşırı öğrenmeyi önlemek amacıyla çapraz doğrulama (cross-validation) ve parametre optimizasyonu yöntemleri uygulanacaktır. Veri setinin doğruluk oranının artması amacıyla bahsi geçmeyen birçok modelin kullanılması/test edilmesi olasıdır.

Proje yönetimi kapsamında çalışma beş aşamada yürütülmektedir: literatür taraması, veri ön işleme, model geliştirme, performans analizi ve raporlama. Literatür taraması aşamasında, safra taşı hastalığı ve yapay zekâ tabanlı sağlık modelleri üzerine güncel çalışmalar incelenmektedir. Veri ön işleme aşamasında, eksik veya hatalı veriler düzeltilip değişkenler uygun formata dönüştürülmektedir. Model geliştirme aşamasında farklı makine öğrenmesi algoritmaları denenerek en yüksek başarı oranına sahip model belirlenmektedir. Performans analizi aşamasında, modeller doğruluk, duyarlılık ve F1-score gibi metriklerle değerlendirilmektedir. Son aşamada ise bulgular raporlanarak genel sonuçlar ve gelecekteki çalışmalara yönelik öneriler sunulmaktadır. Tüm bu aşamaların 6 - 9 ay içerisinde tamamlanması planlanmaktadır. Bu süreçte bahsedilen tüm aşamalar itinayla tamamlanıp projenin kusursuz olacak şekilde devam etmesi sağlanacaktır.

Yaygın etki açısından proje, sağlıkta yapay zekâ tabanlı analizlerin kullanımını teşvik ederek erken tanı, risk analizi ve önleyici hekimlik alanlarında önemli katkılar sunacaktır. Elde edilecek model, özellikle tarama programlarında düşük maliyetli, hızlı ve non-invaziv bir değerlendirme aracı olarak kullanılabilecek potansiyele sahiptir. Ayrıca proje, Türkiye'nin 12. Kalkınma Planı ve 2030 Sanayi ve Teknoloji Stratejisi ile uyumlu biçimde, yerli ve yenilikçi sağlık teknolojilerinin geliştirilmesine destek olacaktır.

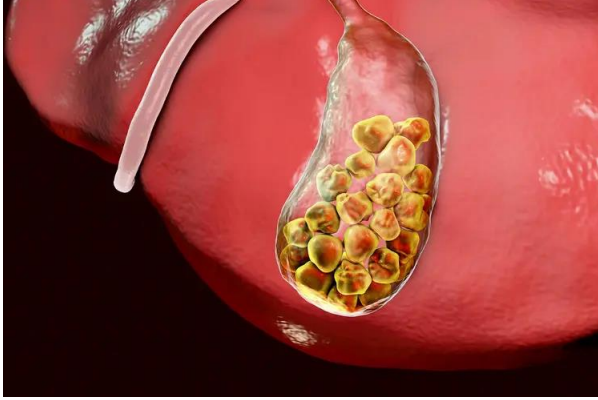
Anahtar Kelimeler: Safra Taşı, Makine Öğrenmesi, k-NN, Klinik Veri Analizi

1. ARAŞTIRMA ÖNERİSİNİN BİLİMSSEL NİTELİĞİ

1.1. Konunun Önemi ve Araştırma Önerisinin Bilimsel Niteliği

Araştırma önerisinde ele alınan konunun kapsamı, sınırları ve önemi ortaya konulur. Araştırma önerisi kapsamında yapılacak çalışmalarla literatürdeki hangi eksikliğin nasıl giderileceği veya hangi soruna nasıl bir çözüm getirileceği ilgili literatüre atıfla açıklanarak araştırma önerisinin bilimsel niteliği ortaya konulur. Araştırma sorusu ve varsa hipotez(ler)i tanımlanır.

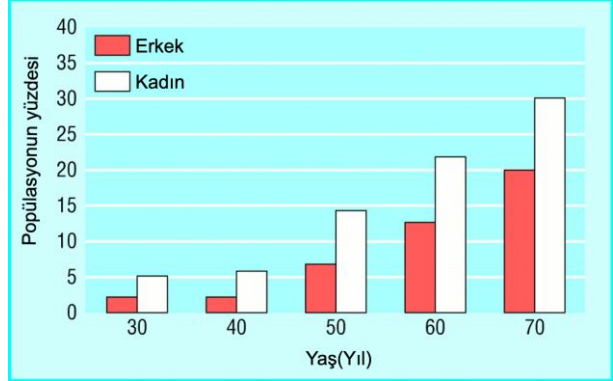
Projenin konusu, 12. Kalkınma Planı ve 2030 Sanayi ve Teknoloji Stratejisi'nde yer alan kritik teknoloji alanları ile öncelikli Ar-Ge ve yenilik konuları ile ilişkili ise, ilişkilendirilme sebebi ve ilgili alana sağlayacağı yararlar açıklanmalıdır.



Safra taşı, safra kesesinde veya safra yollarında safra bileşenlerinin kristalleşerek veya çökerek halk dilinde taş olarak belirtilen katı yapılar oluşturmaları sonucu meydana gelen tıbbi bir durumdur. Safra, karaciğer tarafından üretilen ve **yağların sindiriminde görev alan** bir sıvıdır; safra taşları genellikle safradaki kolesterol, bilirubin ya da diğer bileşenlerin dengesizliği nedeniyle meydana gelir. Safra kesesi yeterince boşalamadığında veya safradaki çözünmeyen bileşen miktarı azaldığında kristalleşme artabilir, taşlar zamanla büyüyebilir. Taşlar boyut ve sayı olarak farklılık gösterebilir; bazıları kum tanesi büyüklüğünde ve sessiz kalabilirken bazıları büyük ve semptomatik olabilir. Çoğu hasta taşlara rağmen **belirti göstermeyebilir**; ancak taşlar safra yollarına

göç ettiğinde karın sağ üst bölgesinde ağrı, sarılık, bulantı, safra kesesi iltihabı veya pankreatit gibi komplikasyonlar oluşabilir. Safra taşının tedavisi spesifik duruma bağlı olarak gözlem, ilaç tedavisi veya safra kesesinin alınması gibi cerrahi operasyon içerebilir. Bu durum hastanın yaşına, cinsiyetine ve yaşadığı coğrafi bölgeye göre değişiklik gösterebilir.

Son yıllarda yapılan meta analize göre, dünyada genel nüfusta safra taşı hastalığının prevalansı **yaklaşık %6** olarak raporlanmıştır; bu oran coğrafi bölgelere göre değişkenlik göstermektedir. Ayrıca, yapılan sistematik derlemeler ve epidemiyolojik veriler taş hastalığının özellikle yaşla birlikte arttığını, kadınlarda erkeklerden daha yaygın olduğunu göstermektedir. Örneğin batılı ülkelerde yetişkin nüfusun **yaklaşık %10-15**'inin safra taşı taşıdığı, yaş ilerlemesiyle bu oranın daha yüksek seviyelere çıkabileceği belirtilmektedir. Bu durum, hem sağlık sistemleri üzerinde artan yük hem de erken tanı ve önleme stratejilerinin önemini artırmaktadır. Projede bu durumlar göz önüne alınarak hastanın daha konforlu ve daha özgür bir şekilde sağlık sistemine erişmesi hedeflenmiştir.



Bu proje, safra taşı hastalığının dünya genelinde yaygın görülmesi ve erken tanı için hâlen büyük ölçüde görüntüleme yöntemlerine bağımlı olunması nedeniyle önemli bir araştırma ihtiyacını karşılamaktadır. Safra taşları çoğu zaman belirti vermeden ilerleyip ciddi komplikasyonlara yol açabildiğinden, non-invaziv, **düşük maliyetli** ve veri temelli bir **erken uyarı sisteminin** geliştirilmesi hem klinik hem toplumsal açıdan büyük değer taşımaktadır. Bu çalışma, biyoimpedans ve laboratuvar verilerini bir arada değerlendirerek makine öğrenmesi algoritmaları aracılığıyla safra taşı riskini yüksek doğrulukla tahmin etmeyi amaçlamakta, böylece literatürdeki klasik görüntüleme tabanlı yaklaşımlara yenilikçi bir alternatif sunmaktadır. Aynı veri seti üzerinde k-En Yakın Komşu (k-NN) ve diğer makine öğrenmesi algoritmalarının karşılaştırmalı olarak uygulanması, projenin hem bilimsel hem metodolojik özgünlüğünü oluşturmaktadır. Projenin amaçlarında belirtildiği gibi en iyi model tespit edilerek hastalığın tahmini ve gerçeğe yakın bir risk skoru belirnebilmesi hedeflenmiştir. Bu yönüyle çalışma, Türkiye'nin 12. Kalkınma Planı ve **2030 Sanayi ve Teknoloji Stratejisi**'nde vurgulanan "**Sağlıkta Yapay Zekâ ve Veri Odaklı Uygulamalar**" hedefleriyle örtüşmekte, klinik karar destek sistemlerine katkı sağlayarak erken teşhis, hasta takibi ve sağlık hizmetlerinin verimliliği açısından önemli bir potansiyel sunmaktadır.

1.2. Amaç ve Hedefler

Araştırma önerisinin amacı ve hedefleri açık, ölçülebilir, gerçekçi ve ulaşılabilir nitelikte olacak şekilde yazılır.

Bu araştırmanın temel amacı, bireylerin demografik, biyoimpedans ve laboratuvar verilerini kullanarak safra taşı hastalığı riskini makine öğrenmesi yöntemleriyle tahmin edebilen, doğruluğu yüksek bir model geliştirmektir. Bu doğrultuda, k-En Yakın Komşu (k-NN) algoritması başta olmak üzere, çok katmanlı algılayıcı (MLP), destek vektör makineleri (SVM), karar ağacı, rastgele orman (Random Forest), uyarlamalı artırma (AdaBoost), gradyan artırma (Gradient Boosting), naif Bayes ve lojistik regresyon teknikleri uygulanacaktır.

Araştırmanın özel hedefleri arasında:

1. Safra taşı hastalığıyla en güçlü ilişkili biyobelirteçleri istatistiksel ve model temelli analizlerle belirlemek,
2. Farklı makine öğrenmesi algoritmalarının performanslarını sistematik biçimde kıyaslamak,
3. Elde edilen en iyi modeli klinik karar destek sistemlerine entegre edilebilecek bir yapıya dönüştürmek,
4. Erken teşhis için kullanılabilecek güvenilir bir risk değerlendirme aracı geliştirmek

yer almaktadır.

Projenin genel hedefi ise, yapay zekâ temelli tanı modellerinin yerli sağlık uygulamalarında kullanımını teşvik ederek, Türkiye'nin sağlık teknolojileri alanındaki dijital dönüşümüne katkı sağlamaktır. Bu bağlamda çalışma, hem akademik literatüre özgün bir katkı sunmakta hem de gelecekte geliştirilebilecek klinik karar destek sistemlerine zemin hazırlamaktadır.

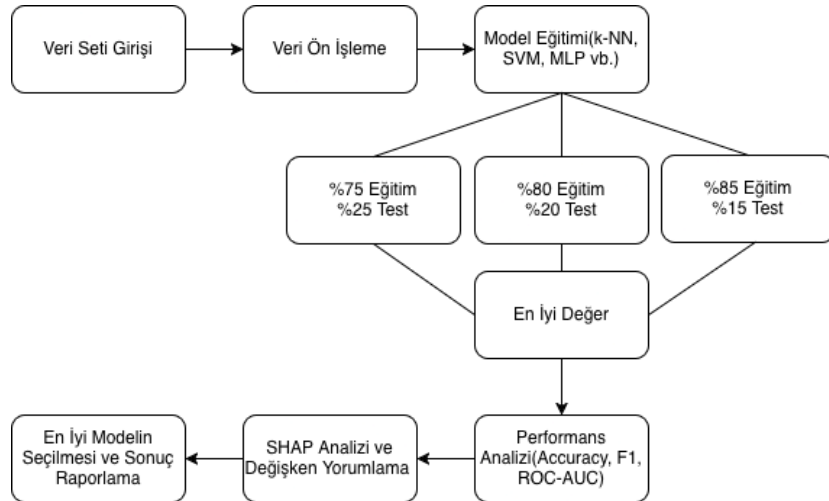
2. YÖNTEM

Araştırmada uygulanacak yöntem ve araştırma tekniklerinin, amaç ve hedeflere ulaşmaya ne düzeyde elverişli olduğu ilgili literatüre atıf yapılarak ortaya konulur.

Yöntem bölümünün; araştırma tasarımı, bağımlı ve bağımsız değişkenler, istatistiksel yöntemler vb. unsurları içermesi gerekir. Araştırma önerisinde herhangi bir ön çalışma veya fizibilite yapıldıysa bunların sunulması beklenir. Araştırma önerisinde sunulan yöntemlerin çalışma takvimi ile ilişkilendirilmesi gerekir.

Bu araştırmada, safra taşı hastalığının tahminine yönelik olarak gözetimli makine öğrenmesi algoritmaları uygulanacaktır. Araştırma tasarımı, retrospektif bir klinik veri analizi üzerine kuruludur ve Ankara VM Medical Park Hastanesi Dahiliye Polikliniği'nden 2022–2023 yılları arasında toplanan, 2319 bireye ait 38 değişken içeren etik onaylı bir veri seti kullanılacaktır. Veri setinde eksik değer bulunmamakta olup, sınıf dağılımı dengelidir.

Bağımlı değişken "Gallstone Status (Safra Taşı Varlığı)" olup, bireyin safra taşı tanısı alıp almadığını göstermektedir. Bağımsız değişkenler ise yaş, cinsiyet, vücut kitle indeksi (BMI), toplam su, ekstraselüler su, intraselüler su, kas kütlesi, yağ oranı, visceral yağ alanı, karaciğer yağlanması, glukoz, kolesterol, HDL, LDL, trigliserid, AST, ALT, ALP, kreatinin, GFR, CRP, hemoglobin ve D vitamini düzeyleri gibi demografik, biyoimpedans ve laboratuvar verileridir.



Veri ön işleme aşamasında kategorik değişkenler Label Encoding yöntemiyle sayısal değerlere dönüştürülecek, tüm değişkenler Min-Max normalizasyonu ile 0–1 aralığına ölçeklenecektir. Ardından veri kümesi test edilerek en iyi şekilde ayrılacak, model performansı test verisi üzerinde değerlendirilecektir.

Kullanılan Modeller ve Yöntemler

Bu araştırmada, safra taşı hastalığı riskini öngörebilmek amacıyla farklı makine öğrenmesi algoritmaları uygulanmıştır. Her bir algoritma, verinin yapısına ve problem türüne uygun özgün bir yaklaşım sunmakta; bu sayede modellerin karşılaştırmalı performans analiziyle en yüksek tahmin doğruluğuna ulaşılması hedeflenmektedir.

1. **k-En Yakın Komşu (k-NN) algoritması**, örneklerin sınıflandırılmasında benzerlik temelli bir yaklaşım kullanır. Yeni bir örnek, özellik uzayında kendisine en yakın komşuların çoğunluk sınıfına göre etiketlenir. Bu modelin tercih edilme nedeni, özellikle biyoimpedans ve laboratuvar verilerinde gözlenen karmaşık doğrusal olmayan ilişkileri basit bir uzaklık metriğiyle etkili biçimde yansıtabilmesidir. k-NN, yorumlanabilirliği yüksek bir yöntem olduğundan, hastalık riskini açıklayan temel özelliklerin belirlenmesinde güçlü bir referans modeli olarak kullanılacaktır.

$$d(x, x_i) = \sqrt{\sum_{j=1}^n (x_j - x_{ij})^2}$$

2. **Çok Katmanlı Algılayıcı (MLP – Neural Network)**, doğrusal olmayan ilişkileri modelleme kapasitesi yüksek bir derin öğrenme yöntemidir. Birden fazla gizli katman ve aktivasyon fonksiyonu kullanarak karmaşık örüntüleri öğrenebilir. Bu çalışmada MLP'nin seçilme nedeni, biyolojik ve klinik değişkenler arasındaki karmaşık etkileşimleri modelleyebilmesi ve yüksek doğruluk potansiyelidir. Özellikle yeterli veri çeşitliliği bulunduğu anda, MLP'nin safra taşı riski tahmininde daha genellenebilir sonuçlar üreteceği öngörülmektedir.

$$h_j = f\left(\sum_{i=1}^n w_{ij} x_i + b_j\right)$$
$$\hat{y} = \sigma\left(\sum_{j=1}^m v_j h_j + c\right)$$

3. **Destek Vektör Makineleri (SVM)**, veri noktalarını en iyi şekilde ayıran hiper düzlemi bulmayı hedefler. Doğrusal ve doğrusal olmayan ayrımları kernel fonksiyonları aracılığıyla öğrenebilir. Bu çalışmada SVM'nin tercih edilme nedeni, özellikle yüksek boyutlu ve karmaşık veri setlerinde güçlü genelleme yeteneğine sahip olmasıdır. Safra taşı hastalığında çok sayıda biyobelirteç bulunduğu andan, SVM bu değişkenler arasındaki sınırları optimize ederek yüksek sınıflandırma başarısı sağlayabilir.

$$w \cdot x + b = 0$$

w : ağırlık vektörü, b : bias, y_i : sınıf etiketi.

4. **Karar Ağacı (Decision Tree)**, veriyi dallara ayırarak sınıflandırma yapan sezgisel bir yöntemdir. Her dal, belirli bir özelliğe göre veri setini alt gruplara böler ve nihai tahmin yaprağı ulaşıldığında elde edilir. Karar ağacı yöntemi, özellikle tıbbi verilerde hangi değişkenlerin karara en fazla katkı sağladığını açıklayabilmesi bakımından değerlidir. Bu nedenle, modelin yorumlanabilirliğini artırmak amacıyla tercih edilmiştir.

$$\text{Gini}(t) = 1 - \sum_{i=1}^c p_i^2$$

p_i : düğümdeki sınıf oranı.

5. **Rastgele Orman (Random Forest)**, birden fazla karar ağacının birleşiminden oluşan bir topluluk (ensemble) öğrenme yöntemidir. Her bir ağaç, rastgele seçilen örnekler ve değişkenler üzerinden eğitilerek çeşitlilik yaratır. Bu modelin seçilme nedeni, yüksek doğruluk sağlaması ve aşırı öğrenmeye (overfitting) karşı dayanıklı olmasıdır. Safra taşı riskinde birden çok faktörün etkileşimli yapısı göz önüne alındığında, Random Forest modelinin genelleme kabiliyeti yüksek bir sonuç elde etmesi beklenmektedir.

$$\hat{y} = \text{mode}(h_1(x), h_2(x), \dots, h_n(x))$$

$h_i(x)$: i. ağacın çıktısı.

6. **Uyarlamalı Artırma (AdaBoost)**, zayıf sınıflandırıcıların ardışık olarak eğitilmesiyle güçlü bir tahmin modeli oluşturan bir topluluk yöntemidir. Her iterasyonda hatalı sınıflandırılmış örneklerin ağırlığı artırılarak modelin daha hassas hale gelmesi sağlanır. AdaBoost'un bu araştırmada kullanılması nedeni, düşük hata oranı ve yüksek duyarlılık sağlamasıdır. Özellikle sınıf dağılımında küçük dengesizlikler mevcutsa AdaBoost bu farkı dengeleyebilir.

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right)$$

ϵ_t : t'ninci modelin hata oranı.

7. **Gradyan Artırma (Gradient Boosting)**, hataları kademeli olarak azaltmak için ardışık modeller oluşturur ve önceki tahmin hatalarından öğrenir. Bu yöntem, karmaşık veri setlerinde güçlü bir tahmin performansı sağlar. Çalışmada Gradient Boosting'in seçilme nedeni, parametre optimizasyonu ile yüksek doğruluk, özgüllük ve dengeli performans elde etme kapasitesidir.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

γ_m : öğrenme oranı.

8. **Naif Bayes (Naïve Bayes)**, Bayes teoremini temel alan olasılıksal bir sınıflandırma algoritmasıdır. Özellikler arasındaki bağımsızlık varsayımıyla çalışır ve küçük veri setlerinde dahi istikrarlı sonuçlar üretir. Bu modelin tercih edilme nedeni, basitliği, hız avantajı ve özellikle ön tanı süreçlerinde hızlı tahmin üretebilme kapasitesidir.

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

9. **Lojistik Regresyon (Logistic Regression)**, olasılık temelli doğrusal bir sınıflandırma yöntemidir. Model, bağımsız değişkenlerle hedef değişken arasındaki ilişkiyi lojistik fonksiyon aracılığıyla açıklar. Bu çalışmada lojistik regresyon, temel bir referans modeli olarak kullanılacak ve diğer karmaşık algoritmaların performanslarının kıyaslanmasında kontrol grubu görevi görecektir.

$$P(y = 1|x) = \frac{1}{1 + e^{-(w \cdot x + b)}}$$

Model performanslarının değerlendirilmesinde doğruluk (Accuracy), duyarlılık (Recall), özgüllük (Specificity), F1-Score ve ROC-AUC metrikleri kullanılacaktır. Ayrıca modellerin karar mekanizmalarının şeffaflığını artırmak amacıyla SHAP (SHapley Additive exPlanations) yöntemi uygulanacaktır. SHAP, her özelliğin model kararına katkısını nicel olarak açıklayarak, özellikle SVM ve MLP gibi karmaşık algoritmaların klinik açıdan yorumlanabilir hale gelmesini sağlayacaktır.

Safra Taşı Hastalığı Risk Tahmin Süreci (Pseudocode)

Algorithm 1 Makine Öğrenmesi Tabanlı Safra Taşı Riski Tahmin Süreci

Require: Veri seti

Ensure: En iyi model ve değerlendirme sonuçları

- 1: **Veri Ön İşleme:**
- 2: Eksik verileri ortalama ile doldur: $x_{ij} = \text{mean}(X_j)$ eğer x_{ij} boşsa
- 3: Kategorik değişkenleri kodla, sayısal verileri ölçekle: $x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$
- 4: Veriyi ayır: $D_{\text{train}}, D_{\text{test}} = \text{split}(D, 0.8)$
- 5: **Model Eğitimi ve Değerlendirme:**
- 6: **for** all algoritma $\in \{\text{kNN, MLP, SVM, RF, DT, NB, LR}\}$ **do**
- 7: $M \leftarrow \text{TrainModel}(\text{algoritma}, D_{\text{train}})$
- 8: $\hat{Y} \leftarrow \text{Predict}(M, D_{\text{test}})$
- 9: Hesapla: $\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$, $F1 = \frac{2PR}{P+R}$
- 10: Kaydet sonuçları: $\text{Results}[\text{algoritma}]$
- 11: **end for**
- 12: **Model Seçimi ve İnce Ayar:**
- 13: $M_{\text{best}} \leftarrow \arg \max_M (F1)$
- 14: $M_{\text{final}} \leftarrow \text{HyperparamTune}(M_{\text{best}}, D_{\text{train}})$
- 15: **Açıklanabilirlik ve Raporlama:**
- 16: SHAP değerlerini hesapla: $\phi_j = \text{SHAP}(M_{\text{final}}, D_{\text{test}})$
- 17: Özellik önem sıralamasını çıkar.
- 18: Performans metriklerini tablo ve grafik olarak özetle.

Bu çalışmada sunulan sahte kod (pseudo-code), projenin genel işleyiş mantığını ve planlanan algoritmik süreci açıklayıcı niteliktedir. Kod, araştırmanın temel adımlarını şematik biçimde ortaya koymakta olup projenin ilerleyişine, veri yapısına ve modelleme sonuçlarına bağlı olarak güncellenebilir esnek bir yapıya sahiptir. Bu nedenle, söz konusu pseudo kod nihai uygulamayı temsil etmemekte, projenin metodolojik çerçevesini ve sistematik akışını kavramsal düzeyde tanımlamak amacıyla hazırlanmıştır.

Elde edilen sonuçlar, **Scikit-learn (v1.4)**, **Pandas**, **NumPy**, **Matplotlib** ve **SHAP** kütüphaneleri kullanılarak analiz edilecektir. Tüm modellerin performansları **accuracy**, **recall**, **specificity**, **F1-score** ve **ROC-AUC** metrikleri üzerinden karşılaştırılacak, bu değerlendirme sonucunda en yüksek genellenebilirliğe sahip model belirlenecektir. Ayrıca istatistiksel incelemelerde korelasyon analizi, özellik önem sıralaması (feature importance), SHAP tabanlı değişken etki analizi ve uygun durumlarda hipotez testleri (t-testi, Mann-Whitney U testi) uygulanarak, demografik, biyoimpedans ve laboratuvar değişkenlerinin safra taşı oluşumuna olan etkileri değerlendirilecektir.

2209/A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEĞİ PROGRAMI
ARAŞTIRMA ÖNERİSİ FORMU

3. PROJE YÖNETİMİ

3.1 Çalışma Takvimi

Araştırmada yer alacak başlıca faaliyetler, her bir faaliyetin hangi sürede gerçekleştirileceği, başarı ölçütü ve araştırmacının başarısına katkısı "Çalışma Takvimi" doldurularak sunulur.

ÇALIŞMA TAKVİMİ (*)

Tarih Aralığı	Faaliyetler**	Kim(ler) Tarafından Gerçekleştirileceği	Başarı Ölçütü ve Araştırmacının Başarısına Katkısı***
01/12/2025 - 31/01/2026	Veri seti hazırlığı ve potansiyel ek veri arayışı	Tüm proje üyeleri	%15
01/02/2026 - 31/02/2026	Bulunan tüm veri setlerinin ön işleminin yapılması	Tüm proje üyeleri	%20
01/03/2026 - 15/04/2026	Kullanılacak veri setleri ile algoritmaların yazılması	Tüm proje üyeleri	%40
16/04/2026 - 31/06/2026	Algoritmanın doğruluğu ve test aşamaları	Tüm proje üyeleri	%15
01/07/2026 - 01/08/2026	Projenin sonlandırılması ve hazırlıkların yapılması	Tüm proje üyeleri	%10

(*) Çizelgedeki satırlar ve sütunlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

(**) Literatür taraması, sonuç raporu hazırlama aşamaları, araştırma sonuçlarının paylaşımı, ve malzeme alımı ayrı birer iş adımı olarak gösterilmemelidir.

(***) Başarı ölçütü, ölçülebilir ve izlenebilir nitelikte olacak şekilde nicel veya nitel ölçütlerle (ifade, sayı, yüzde, vb.) belirtilir. Bu sütundaki değerlerin toplamı 100 olmalıdır.

3.2 Risk Yönetimi

Araştırmanın başarısını olumsuz yönde etkileyebilecek riskler ve bu risklerle karşılaşıldığında araştırmanın başarıyla yürütülmesini sağlamak için alınacak tedbirler (B Planı) aşağıdaki Risk Yönetimi Tablosu'nda ifade edilir. B Plan(lar)ının uygulanması araştırmanın temel hedeflerinden sapmaya yol açmamalıdır. B Plan(lar)ına geçilmesi durumunda yöntem değişikliğine gidiliyor ise bu durum detaylandırılmalıdır.

RİSK YÖNETİMİ TABLOSU*

En Önemli Riskler	Alınacak Tedbirler (B Planı)
Güncel veya Yerel veri seti bulunamaması	Üniversite veri tabanlarına erişim veya üniversiteden veri tabanı erişim hakkı isteme
Bulunan veri setlerinin yeteri kadar değişken barındıramaması	Sentetik veri ekleme
Yeteri kadar iyi başarı oranı oluşturulamaması/istenilen orana ulaşamamak	Yeni modeller eklenip test aşamasına eklenmesi
Literatürde kaynaklara erişememek	Üniversiteden kaynaklara erişim talebi

(*) Tablodaki satırlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

2209/A ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEĞİ PROGRAMI
ARAŞTIRMA ÖNERİSİ FORMU

3.3 Araştırma Olanakları

Bu bölümde projenin yürütüleceği kurum ve kuruluşlarda var olan ve projede kullanılacak olan altyapı/ekipman (laboratuvar, araç, makine-teçhizat, vb.) olanakları belirtilir.

ARAŞTIRMA OLANAKLARI TABLOSU (*)

Kuruluşta Bulunan Altyapı/Ekipman Türü, Modeli (Laboratuvar, Araç, Makine-Teçhizat, vb.)	Projede Kullanım Amacı
İnternet	Veri seti araştırması ve makale bulma
Bilgisayar	Algoritmaların oluşturulması
Laboratuvar	Çalışmaların yürütülmesi

(*) Tablodaki satırlar gerektiği kadar genişletilebilir ve çoğaltılabilir.

4. ARAŞTIRMA ÖNERİSİNİN YAYGIN ETKİSİ

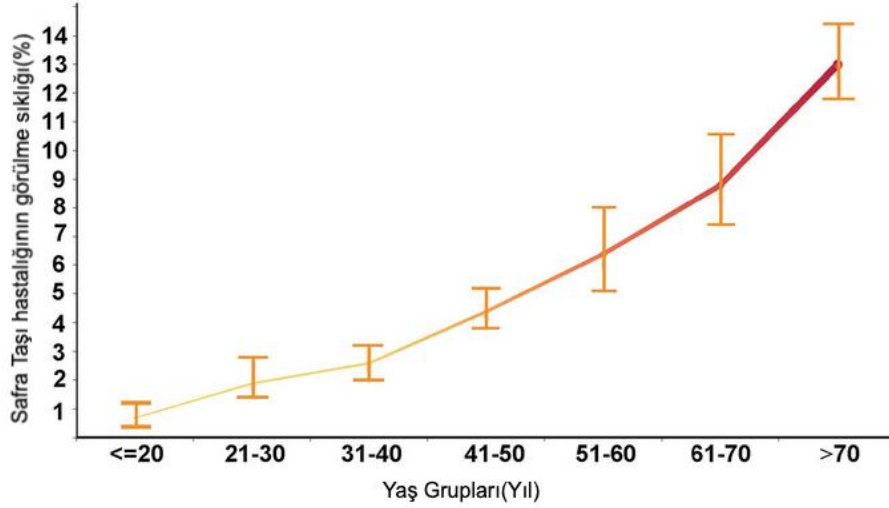
Araştırma önerisi kapsamındaki çalışmadan elde edilmesi öngörülen çıktılar amaçlarına göre belirlenen kategorilere ayrılarak belirtilir; ölçülebilir ve gerçekçi hedeflere dayandırılır.

Çıktı, Etki ve Kazanımlar	Öngörülen Çıktı(lar), Etki(ler) ve Kazanım(lar)
Bilimsel/Akademik Çıktılar (Ulusal/Uluslararası Makale, Kitap Bölümü, Kitap, Bildiri vb.)	Safra taşı hastalığının erken teşhisi, hastalara daha verimli bilgi verilebilmesi ve safra taşı hastalığındaki temel sorunlara bakış açısı oluşturma.
Ekonomik/Ticari/Sosyal Çıktılar (Ürün, Prototip, Patent, Faydalı Model, Tescil, Görsel/İşitsel Arşiv, Envanter/Veri Tabanı, Çalıştay, Eğitim, Bilimsel Etkinlik vb.)	Sağlık alanında yerli yazılım çözümlerinin gelişimine katkı sağlayarak erken tanı, maliyet ve test azaltımı ve klinik verimlilik açısından sosyal ve ekonomik fayda sağlama.
Yeni Proje(ler) Oluşturmasına Yönelik Çıktılar (Ulusal/Uluslararası Yeni Proje vb.)	Bu proje sonrasında bazı verilere dayanılarak safra taşı hastalığının daha spesifik sebeplerle nitelendirildiği yeni veri setleri ve ya makaleler kullanılarak yapay zekâ ile entegre edilerek hastalarda kullanılması.

5. BELİRTMEK İSTEDİĞİNİZ DİĞER KONULAR

Sadece araştırma önerisinin değerlendirilmesine katkı sağlayabilecek bilgi/veri (grafik, tablo, vb.) eklenebilir.

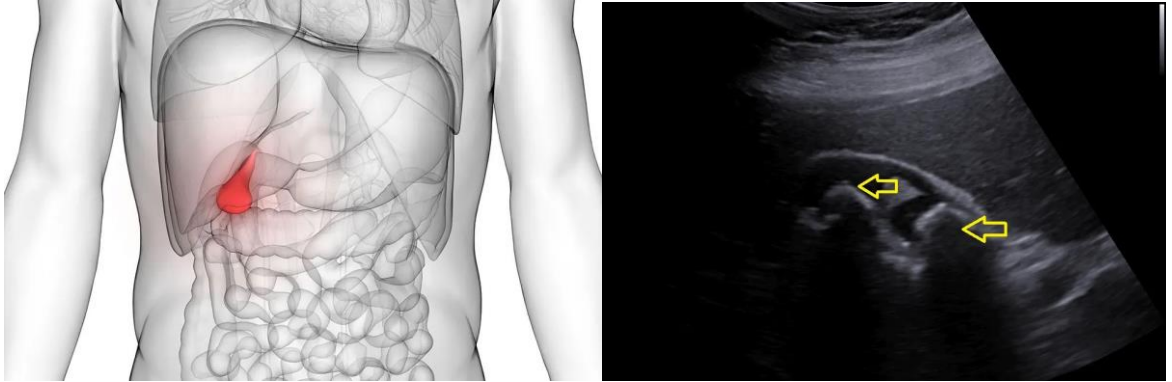
Aşağıdaki tablo, yaş gruplarına göre safra taşı görülme sıklığını göstermekte olup hastalığın yaşla birlikte belirgin biçimde arttığını ortaya koymaktadır.



Şekil 1.

Yaş gruplarına göre safra taşı hastalığının görülme sıklığını gösteren istatistiksel veri.

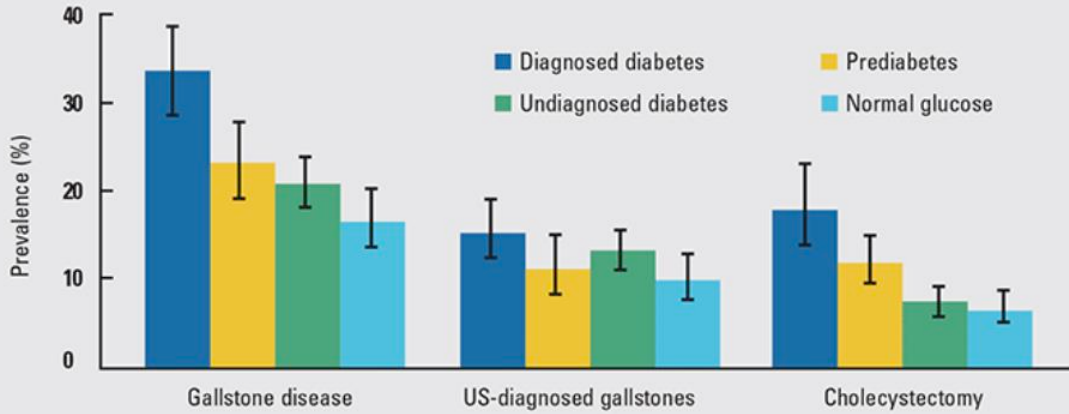
Verilere göre 20–30 yaş aralığında oldukça düşük olan safra taşı oranı, 40 yaş üzeri bireylerde anlamlı düzeyde artış göstermekte, 60 yaş ve sonrasında ise en yüksek seviyeye ulaşmaktadır. Bu durum, yaşın safra taşı oluşumunda önemli bir risk faktörü olduğunu desteklemekte ve proje kapsamında geliştirilecek tahmin modellerinde yaş değişkeninin belirleyici rolünü vurgulamaktadır.



Safra kesesi, karaciğerin alt yüzeyinde yer alan, sindirim sürecinde önemli rol oynayan küçük bir organdır. Karaciğer tarafından üretilen safra, bu kesede depolanır ve yağların sindirimi sırasında ince bağırsağa iletilir. Üstteki görselde safra kesesinin karın boşluğundaki anatomik konumu gösterilmektedir.

Üstteki röntgen görüntüsünde safra taşlarının (kolelitiazis) tipik görünümü izlenmektedir. Safra taşları, genellikle kolesterol, bilirubin veya kalsiyum tuzlarının birikimi sonucu oluşur ve zamanla safra akışını engelleyerek ağrı, bulantı, sarılık ve akut kolesistit gibi klinik tablolara yol açabilir. Bu tür görseller, hastalığın patofizyolojisini anlamak ve tanısal süreçte ultrasonografi veya BT görüntülemesinin önemini vurgulamak açısından değerlidir.

Age-Standardized Prevalence of Gallstone Disease, US-Diagnosed Gallstones, and Cholecystectomy by Diabetes Status



^a U.S. data for the period 1988–1994. US: ultrasound. Source: Reference 7.

Şekil 2.

ABD genel nüfusunda diyabet durumuna göre safra taşı hastalığı prevalansı (NHANES III verileri)

Üçüncü Ulusal Sağlık ve Beslenme İnceleme Çalışması (NHANES III) kapsamında **20–74 yaş arası** yetişkin bireylerde yapılan safra kesesi ultrasonografisi analizlerine göre, safra taşı hastalığının görülme sıklığı diyabetin varlığına göre anlamlı farklılık göstermektedir. Tanı almış diyabet hastalarında safra taşı prevalansı **%33,3 ± 2,6** iken, tanı almamış diyabetlilerde bu oran **%23,3 ± 2,2** olarak saptanmıştır. Prediyabet grubunda prevalans **%20,8 ± 1,5**, normal glukoz düzeyine sahip bireylerde ise **%16,7 ± 1,7** olarak belirlenmiştir.

NHANES III çalışmasında diyabet; tanı konmuş (sağlık profesyoneli tarafından bildirilen), tanı almamış (A1C ≥ 6.5% veya açlık glukozu ≥ 126 mg/dL), prediyabet (A1C %5.7–6.4 veya açlık glukozu 100–125 mg/dL) ve normal glukoz (A1C < %5.7 ve açlık glukozu < 100 mg/dL) olarak sınıflandırılmıştır. Bu bulgular, diyabet varlığının **safra taşı gelişiminde önemli bir risk faktörü** olduğunu göstermektedir. Günümüzün popüler hastalıklarından olan **diyabetin safra taşıyla doğrudan olan bağlantısı** projenin gelecekteki yerini belirliyor.



Şekil 3.

Safra taşı ağrısı (biliyer kolik) yaşayan birey

Safra taşları, safra kesesi veya safra kanallarında tıkanmaya neden olduğunda şiddetli karın ağrısı ile kendini gösterir. Bu ağrı genellikle **biliyer kolik** olarak adlandırılır ve karının sağ üst kısmında, kaburga altında yoğun bir şekilde hissedilir. Ağrı omuz veya sırtta yayılabilir, genellikle yağlı bir öğünden sonra başlar ve birkaç dakikadan birkaç saate kadar sürebilir.

Ağrıya eşlik eden bulantı, kusma, terleme ve karında gerginlik hissi tipiktir. Eğer taş safra akışını tamamen engellerse, **akut kolesistit** (safra kesesi iltihabı), **kolanjit** (safra yolu iltihabı) veya **akut pankreatit** (pankreas iltihabı) gibi ciddi komplikasyonlar gelişebilir. Bu durumlar acil tıbbi müdahale gerektirir ve tedavi edilmezse sepsis veya çoklu organ yetmezliği gibi hayati riskler ortaya çıkabilir. Bu nedenle görseldeki gibi safra taşı ağrısı çeken bireylerin, belirtileri tekrar etmeden önce tıbbi değerlendirmeye başvurmaları büyük önem taşımaktadır.

6. EKLER

EK-1: KAYNAKLAR

- [1] World Gastroenterology Organisation. The Growing Global Burden of Gallstone Disease. E-WGN Expert Point of View, World Gastroenterology Organisation. Erişim adresi: <https://www.worldgastroenterology.org/publications/e-wgn/e-wgn-expert-point-of-view-articles-collection/the-growing-global-burden-of-gallstone-disease>
- [2] Gallstone disease: epidemiologic, mechanistic and potential therapeutical strategies.. Erişim adresi: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11309733/>
- [3] WebMD. Gallstones. WebMD, n.d. Erişim adresi: <https://www.webmd.com/digestive-disorders/gallstones>
- [4] Jones, M. W. Gallstones (Cholelithiasis). StatPearls, National Center for Biotechnology Information (NCBI), 2024. Erişim adresi: <https://www.ncbi.nlm.nih.gov/books/NBK459370/>
- [5] Sun, H., Zhang, L., & Chen, Y. "Factors Influencing Gallstone Formation: A Review of the Pathogenesis." Biomolecules, vol. 12, no. 4, 2022, p. 550. Erişim adresi: <https://www.mdpi.com/2218-273X/12/4/550>
- [6] Portincasa, P., Di Ciaula, A., Bonfrate, L., Stella, A., & Garruti, G. "Metabolic Dysfunction-Associated Gallstone Disease: Expecting More from Critical Care Manifestations." Internal and Emergency Medicine, vol. 18, no. 2, 2023, pp. 293–304. Erişim adresi: <https://link.springer.com/article/10.1007/s11739-023-03355-z>
- [7] Abdu, S. M., Ahmed, M., & Kedir, S. "Prevalence of Gallstone Disease in Africa: A Systematic Review." BMJ Open Gastroenterology, vol. 12, no. 1, 2025, e001441. Erişim adresi: <https://bmjopengastro.bmj.com/content/12/1/e001441>
- [8] Latenstein, C. S. S., et al. "Tailoring Diagnosis and Treatment in Symptomatic Gallstone Disease: A Review of Trials and Guidelines." The British Journal of Surgery, vol. 109, no. 9, 2022, pp. 832–841. Erişim adresi: <https://academic.oup.com/bjs/article/109/9/832/6595861>
- [9] Ye, Z., Wang, T., & Zhao, L. "Physical Activity and Risk of Cholelithiasis: A Narrative Review." Frontiers in Medicine, vol. 11, 2024, 1485097. Erişim adresi: <https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2024.1485097/full>
- [10] Swobodnik, W.; Soloway, R. D.; Ditschuneit, H. (eds.). Gallstone Disease: Pathophysiology and Therapeutic Approaches. Springer, 1990.
- [11] Cox, M. R.; Eslick, G. D.; Padbury, R. (eds.). The Management of Gallstone Disease: A Practical and Evidence-Based Approach. Springer International Publishing, 2018.
- [12] Portincasa, P. (chap.). "Gallstones." In: Yamada's Textbook of Gastroenterology, (ed. Yamada). Wiley, 2022.
- [13] Williams, V. (2021, December 17). Diabetes-Related Gastrointestinal Emergencies: Gallstones and Acute Pancreatitis. U.S. Pharmacist, 46(12), HS2–HS9. Erişim adresi: <https://www.uspharmacist.com/article/diabetesrelated-gastrointestinal-emergencies-gallstones-and-acute-pancreatitis>
- [14] The Scan Clinic. No Gallstones Diet – Complete Guide 101. The Scan Clinic. Erişim adresi: <https://thescanclinic.co.uk/health/gallstones-diet-complete-guide/>