

ARAŞTIRMA MAKALESİ

Makine Öğrenmesi Tabanlı Hematolojik Bozukluk Tespiti

Machine Learning Based Hematologic Disorder Detection

Öznur Suçeken^{1*}, Tuğçe Güzle²

¹Bilişim Sistemleri Mühendisliği Bölümü, Teknolojisi Fakültesi, Muğla Sıtkı Koçman Üniversitesi, Muğla, Türkiye.

Geliş / Received: 09.04.2025

Kabul / Accepted: 02.05.2025

***Sorumlu Yazar:** Öznur Suçeken oznursuceken@mu.edu.tr

ÖZ: Hematolojik bozukluklar, kan hücrelerinin üretimi, işlevi veya yapısında meydana gelen anormallikler sonucu ortaya çıkan, ancak erken teşhis ve uygun müdahale ile kontrol altına alınabilen ciddi sağlık problemleridir. Anemi, lösemi, lenfoma, hemofili ve trombositopeni gibi hastalıklar ile kendini gösteren bu anormallikler; bağışıklık sistemi, oksijen taşıma kapasitesi, pıhtılaşma ve doku onarımı gibi hayati fonksiyonları doğrudan etkilemektedir. Bu hastalıkların zamanında tanısı, etkili tedavi planlarının oluşturulması ve komplikasyonların önlenmesi açısından kritik önem taşımaktadır. Ancak klinik süreçlerde kalabalık veri takibi ve yorumlama zorlukları, hekim kararlarını zorlaştırabilmektedir. Bu nedenle yapay zeka destekli dijital sağlık çözümleri, tedavi yönetimlerinde önemli bir rol oynamaktadır. Bu çalışmada, hematolojik hastalıkların teşhisini kolaylaştırmak ve bireyselleştirilmiş tedavi önerileri sunmak amacıyla bir karar destek sistemi geliştirilmiştir. Sistem, hemogram (tam kan sayımı) verilerini temel alarak hastalık teşhisini yüksek doğrulukla gerçekleştirmeyi hedeflemektedir. Teşhis sürecinde Random Forest Algoritma'sı kullanılırken; beslenme, yaşam tarzı ve genel sağlık önerileri karar ağaçları yöntemiyle oluşturulmuştur. Veri setindeki sınıf dengesizliği, model performansını artırmak amacıyla Synthetic Minority Oversampling Technique (SMOTE) ile dengelenmiştir. Flask ve Node.js destekli backend altyapısı farklı kullanıcı yetkileri için React tabanlı arayüzler ile ayrıştırılarak, veri setleri MongoDB veritabanında depolanmıştır. Hematolojik bozuklukların tespiti için geliştirilen karar destek sisteminde tanı doğruluğunu artırmayı, tedavi sürecini hızlandırmayı ve hasta bakım kalitesini yükseltmeyi amaçlayan, klinik uygulamalara entegre edilebilir ve ölçeklenebilir bir yapay zeka çözümü sunulmaktadır.

Anahtar Kelimeler: Hematolojik bozukluk, hemogram, karar destek sistemi, makine öğrenimi, random forest algoritması.

ABSTRACT: Hematologic disorders are serious health problems caused by abnormalities in the production, function or structure of blood cells that can only be controlled with early diagnosis and appropriate intervention. These abnormalities, which manifest themselves in diseases such as anemia, leukemia, lymphoma, hemophilia and thrombocytopenia, directly affect vital functions such as the immune system, oxygen carrying capacity, clotting and tissue repair. Timely diagnosis of these diseases is critical to establishing effective treatment plans and preventing complications. However, crowded data tracking and interpretation difficulties in clinical processes can make physician decisions difficult. Therefore, AI-supported digital health solutions play an important role in treatment management. In this study, a decision support system was developed to facilitate the diagnosis of hematological diseases and provide individualized treatment recommendations. The system aims to perform disease diagnosis with high accuracy based on hemogram (complete blood count) data. While the Random Forest Algorithm is used in the diagnosis process, nutrition, lifestyle and general health recommendations are generated using decision trees. Class imbalance in the dataset was balanced with Synthetic Minority Oversampling Technique (SMOTE) to improve model performance. Flask and Node.js supported backend infrastructure is decomposed with React based interfaces for different user authorizations and data sets are stored in MongoDB database. In the decision support system developed for the detection of hematological disorders, a scalable artificial intelligence solution that can be integrated into clinical applications is presented, aiming to increase the accuracy of diagnosis, accelerate the treatment process and improve the quality of patient care.

Keywords: Decision support system, hematologic disorder, hemogram, machine learning, random forest algorithm.

1. GİRİŞ

Sağlık alanındaki teşhis ve tedavi süreçlerinin etkinliği, klinik başarıyı doğrudan etkileyen temel faktörlerden biridir. Özellikle karmaşık semptomlara sahip, çok faktörlü yapılarıyla tanımlanan hematolojik hastalıkların teşhisi, geleneksel yöntemlerle zaman alıcı ve hata payı yüksek bir süreç olabilir. Bu süreçte laboratuvar testlerinin manuel yorumlanması, uzman deneyimine bağlı olarak değişiklik gösterebilir ve tutarsızlıklar ortaya çıkabilir [1]. Bu gibi nedenlerle, son yıllarda sağlık hizmetlerinde dijitalleşme ve otomasyon ihtiyacı daha da artmış, özellikle yapay zeka ve makine öğrenmesi tabanlı karar destek sistemleri (KDS) ön plana çıkmıştır. Yapay zeka ve makine öğrenmesi, büyük miktarda sağlık verisini analiz etme, örüntüleri tanıma ve öngöründe bulunma yeteneği sayesinde, klinik karar verme süreçlerini daha güvenilir ve hızlı hale getirmektedir. Özellikle derin öğrenme modelleri, hematolojik bozuklukların sınıflandırılması, hücresel anormalliklerin tanımlanması ve mikroskobik görüntülerde anomali tespiti gibi uygulamalarda yüksek doğruluk oranlarına ulaşmaktadır [2]. Convolutional Neural Network (CNN) gibi modeller, morfolojik özelliklere dayalı görüntü analizi alanında sıkça kullanılmakta ve tanısal verilerin değerlendirilmesinde etkin çözümler sunmaktadır [3]. KDS'lerin yapay zeka uygulamalarına entegrasyonu ile birlikte tanı süreçlerinde standardizasyon sağlanmakta, insan hatasına dayalı yanlış pozitif veya negatif sonuçların azaltılması mümkün olmaktadır. Klinik uzmanların kararlarını desteklemek amacıyla geliştirilen bu sistemlerin önerdikleri teşhislerin ardındaki nedenleri şeffaf biçimde sunması gerekmektedir [4].

Literatürdeki akademik çalışmalar incelendiğinde makine öğrenimi ve yapay zeka tekniklerinin anemi teşhisi ve tedavisinde ne kadar etkili olabileceğini ortaya koymakta ve sağlık sektöründe bu tür teknolojilerin yaygınlaşması için önemli bir temel oluşturduğu görülmektedir. Noor ve arkadaşları yaptıkları çalışmada, non-invaziv teşhis yöntemlerine yönelik bir başka çalışmada, göz konjonktiva görüntüleri kullanılarak hemoglobinin seviyelerinin tahmininde %92 doğruluk oranına sahip bir model geliştirmişlerdir. RGB, HSV ve CIE LAB renk uzaylarından özellik çıkarılarak eğitilen

ensemble model, ekonomik ve pratik bir alternatif olarak sunulmuştur [5]. Ponnusamy ve arkadaşları, demir eksikliği anemisi teşhisinde açıklanabilir yapay zeka (XAI) yöntemlerini kullandıkları bir çalışmada, SHAP yöntemiyle hemoglobin, hematokrit ve toplam serum demiri gibi parametrelerin teşhisteki önemi analiz etmişlerdir. Random Forest ve Lojistik Regresyon algoritmaları %93 ile %100 arasında doğruluk oranları sağlamış ve modelin şeffaflığı artırılmıştır [6]. Tummala ve Parvataneni, yaptıkları çalışmada göz konjonktiva görüntüleri ve kan testleri verilerinin karşılaştırıldığı bir çalışmada, kan testlerinde AdaBoost, LGBM ve ExtraTrees algoritmaları %100 doğruluk sağlarken, göz görüntülerinde LGBM ve Perceptron modelleri %97.61 doğruluk oranına ulaşmıştır. Bu çalışma, her iki yöntemin de anemi teşhisinde etkin kullanılabileceğini göstermiştir [7]. Mugdha ve arkadaşları, hemoglobin seviyesinin tahmini ve anemi şiddetinin belirlenmesi için yaptıkları bir araştırmada, Neural Network algoritması hemoglobin tahmininde en yüksek doğruluk oranını sağlarken, Random Forest algoritması anemi şiddeti sınıflandırmasında %99 doğruluğa ulaşmıştır [8]. Rahman ve arkadaşları, anemi teşhisi için yaptıkları bir başka çalışmada, Lojistik Regresyon %95 doğruluk oranıyla en iyi performansı gösterirken, Random Forest, AdaBoost ve XGBoost gibi algoritmalar da yüksek doğruluk elde etmişlerdir [9]. Literatürdeki güncel çalışmalardan görüldüğü üzere anemi ve hematolojik bozuklukların teşhisinde makine öğrenimi algoritmalarının kullanımı sağlık alanında önemli katkılar sağlamıştır.

Bu çalışma, karar destek sistemlerinin sağlık sektöründeki potansiyelini ortaya koymakta ve teşhis-tedavi süreçlerinin nasıl optimize edilebileceğine dair bir çerçeve sunmaktadır. Geliştirilen sistem, tam kan sayımı (hemogram) verilerini analiz ederek hematolojik bozuklukların teşhisini yapmakta ve teşhis sonuçlarına ek olarak vücut kitle indeksi (VKİ) ve cinsiyet gibi parametreleri temel alarak kişiselleştirilmiş öneriler sunmaktadır. Bu yaklaşım, hastaların klinik durumlarına özgü tedavi ve yaşam tarzı önerileri geliştirilmesine olanak tanımaktadır. Sistemin teşhis aşamasında yüksek doğruluk oranı ve çok boyutlu veri setlerinde etkin performansı nedeniyle Random Forest Algoritması kullanılmıştır. Öneri mekanizması için Karar Ağaçları yöntemi tercih

edilmiştir. Veri setindeki sınıf dengesizliği, SMOTE yöntemiyle giderilerek modelin tespit yeteneği artırılmıştır. Yazılım altyapısında kullanıcı arayüzü, React kullanılarak geliştirilmiş ve kullanıcı dostu bir deneyim sunmaktadır. Backend tarafında Flask ve Node.js tabanlı API'ler, veri işleme ve analiz süreçlerini yönetmektedir. Veriler, MongoDB veritabanında saklanarak esnek ve dinamik bir veri yönetimi sağlanmıştır. Sistemde, sağlık çalışanları, hastalar ve yöneticiler için farklı kullanıcı rolleriyle özelleştirilmiş arayüzler sunularak teşhis ve öneri süreçleri kolaylaştırılmıştır. Bu tür yapay zeka tabanlı karar destek sistemlerinin yaygınlaşması yalnızca sağlık çalışanlarının iş yükünü azaltmakla kalmamakta, aynı zamanda hastaların tedavi süreçlerine daha bilinçli ve aktif katılımını da teşvik etmektedir. Bu çalışmada, hematolojik bozuklukların erken ve doğru teşhisini desteklemek amacıyla geliştirilen, makine öğrenimi temelli bir karar destek sistemi sunulmuştur. Önerilen sistem, klinik veriler üzerinden yüksek doğrulukla analiz yaparak teşhis sürecine katkı sağlamayı, açıklanabilir yapay zeka teknikleriyle klinisyenlerin iş yükünü azaltmaya ve mevcut klinik iş akışlarına entegre edilebilir bir çözüm sunmayı hedeflemektedir.

2. ANEMİ TEŞHİS VE TAVSİYE SİSTEMİ

Bu çalışmada kullanılan teşhis verileri, Kaggle veri paylaşım platformunda Aboelnaga tarafından yayımlanan "Anemia Types Classification" veri setinden alınmıştır. Kullanılan veri seti 1.281 gözlem ve 15 değişkenden oluşmaktadır [10]. Tam kan sayımı parametrelerini ve teşhis sınıflarını içermektedir. Veri setindeki bazı parametreler ve açıklamaları Tablo 1'de verilmiştir.

Cinsiyet, VKİ ve teşhis bilgilerine dayalı olarak yapay zeka tabanlı bireyselleştirilmiş beslenme önerileri ve yaşam kalitesini artırarak bireyin tedavi sürecini destekleyecek öneriler sunan bir modül geliştirilmiştir. Modülde her öneri üç parametre kombinasyonuna göre yapılandırılmıştır. Bu önerilerin oluşturulmasında herhangi bir hazır veri seti kullanılmamıştır. Teşhis, cinsiyet ve VKİ parametrelerinin birleşimi temel alınarak, literatür incelemeleri ile elde edilen bilimsel bulgular ve alan yazındaki önerilere dayalı olarak yapılandırılmış öneriler geliştirilmiştir. Her bireyin sağlık durumu, fiziksel özellikleri ve

biyolojik ihtiyaçları farklılık gösterdiğinden, önerilerin etkili ve kişiye özel olabilmesi amacıyla bu üç temel parametre kullanılmıştır.

Tablo 1: Örnek kan sayımı parametreleri

Parametre	Açıklama
WBC	Beyaz kan hücreleri
LYMP	Lenfosit yüzdesi
NEUTP	Nötrofil yüzdesi
LYMN	Lenfosit sayısı
NEUTN	Nötrofil sayısı
RBC	Kırmızı kan hücreleri
HGB	Hemogloblin konsantrasyonu
HCT	Hematokrit yüzdesi
MCV	Ortalama eritrosit hacmi

Demir eksikliği olan bir bireye demir açısından zengin gıdalar (örneğin kırmızı et, ıspanak, mercimek) ve demir emilimini artıran C vitamini kaynakları önerilir. Teşhis dikkate alınmadan yapılan genel öneriler, bireyin durumuna uygun olmayabilir ve sağlık sorunlarının devam etmesine yol açabilir. Kadınlar ve erkekler arasında hem demir eksikliği risk faktörleri hem de ihtiyaçlar bakımından önemli farklılıklar bulunmaktadır. Kadınlarda özellikle adet dönemlerinde düzenli kan kaybı yaşanmasıyla demir depolarının azalması ve demir eksikliği riskinin artması ortaya çıkabilmektedir. Bu nedenle kadınlara yönelik önerilerde daha yüksek demir alımı önerilir (örneğin kırmızı et, karaciğer, demir takviyeleri). Demir emilimini artırmak için C vitamini kaynakları (örneğin portakal, biber) eklenir. Erkeklerde demir eksikliği genellikle kronik hastalıklar, beslenme yetersizliği veya sindirim sistemi rahatsızlıklarından kaynaklanır. Bu durumda öneriler diyetle demir alımının artırılmasını, ancak kadınlara kıyasla daha düşük dozlarda yapılmasını içerir. Sindirim sistemi sağlığını desteklemek ve demir emilimini artırmak için bağırsak sağlığına yönelik besinleri de içerir [11-14].

VKİ, bireyin kilo durumu hakkında bilgi vererek beslenme önerilerinin yönlendirilmesini sağlar.

VKİ farklılıkları, önerilerin enerji dengesi, demir emilimi ve kilo yönetimi gibi faktörlere göre uyarlanmasını gerektirir. Zayıf bireylerde yetersiz beslenme, demir eksikliğine sıkça eşlik eder. Bu bireylere enerji yoğun ve demir bakımından zengin gıdalar önerilir. Kilo alımını desteklemek için karbonhidrat ve protein kaynaklarının artırılması tavsiye edilir. Normal VKİ' ye sahip bireylerde odak noktası, demir açısından zengin ancak dengeli bir diyet sunmaktır. Bu durumda, bireyin mevcut sağlıklı kiloyu koruması hedeflenir. Obezite, demir eksikliği ile ilişkili metabolik problemlere yol açabilir. Bu bireylere yönelik önerilerde kilo kontrolünü sağlarken demir alımını artıracak düşük kalorili ancak besleyici gıdalar vurgulanır. Özellikle yeşil yapraklı sebzeler, baklagiller ve yağsız protein kaynakları önerilir. VKİ dikkate alınmadığında, öneriler bireyin genel sağlık durumu ve kilo yönetimi ile uyumsuz hale gelebilir. Örneğin, zayıf bir bireye kilo alımını desteklemeyen düşük kalorili öneriler sunmak ya da obez bir bireye enerji yoğun diyet önermek, sağlık sonuçlarını olumsuz etkileyebilir. Bu üç parametrenin birlikte ele alınması öneri sisteminin hem güvenilir hem de bilimsel temellere dayalı olmasını sağlar. Kadın ve erkek bireylerin farklı fizyolojik ihtiyaçları ve VKİ 'nin enerji dengesi üzerindeki etkileri göz önüne alındığında bu parametrelerin kullanımı önerilerin doğruluğunu ve etkinliğini artırır [15-17].

2.2 Veri Ön İşleme

Sınıf dengesizliği, veri setindeki bazı sınıfların diğer sınıflara göre daha az sayıda örneğe sahip olması durumudur. Bazı sınıfların diğerlerine kıyasla daha az örneğe sahip olması, modelin çoğunluk sınıfını öğrenmeye eğilim göstermesine ve azınlık sınıflarını doğru bir şekilde tahmin edememesine yol açar. Model, çoğunluk sınıflarını daha sık tahmin eder ve azınlık sınıfları göz ardı eder. Azınlık sınıflarının tahmin edilememesi, özellikle tıbbi uygulamalarda kritik kararların yanlış alınmasına neden olabilir. Synthetic Minority Over-sampling Technique (SMOTE), sınıf dengesizliğini gidermek için kullanılan bir veri artırma yöntemidir. Bu teknik, azınlık sınıfındaki her bir veri noktasını temel alarak onun en yakın k komşusunu belirler ve bu komşular arasındaki vektörler üzerinde rastgele noktalar seçerek yeni

sentetik örnekler oluşturur. Böylece veri setine mevcut örneklerin birebir kopyası yerine, çeşitli sentetik örnekler eklenir. Bu işlem azınlık sınıflarını dengeleyerek veri setinin daha dengeli bir hale gelmesini sağlar. Bu sayede performans metriklerinde, özellikle azınlık sınıfları için belirgin iyileşmeler sağlanır [18].

Bu çalışmada, hemogram verilerinde sınıf dengesizliği tespit edilmiş ve bu sorunu gidermek için SMOTE yöntemi uygulanmıştır. Veriler %80 eğitim ve %20 test oranında bölünmüştür. Bölünme sırasında "random-state=42" parametresi kullanılmıştır. Bu parametre bölünmenin rastgele fakat tekrarlanabilir olmasını sağlayarak sonuçların tutarlılığını garanti altına almıştır. Bireyin kişisel bilgilerine uygun öneri veren modülde ise kategorik değişkenler, makine öğrenimi algoritmalarında kullanılabilir hale getirilmek için sayısal değerlere dönüştürülmüştür. Bu işlem, her bir kategorik değer için benzersiz bir sayısal etikete eşlendiği Label Encoding yöntemi kullanılarak gerçekleştirilmiştir. Bu sayede cinsiyet, VKİ kategorisi ve teşhis gibi değişkenler modeller tarafından işlenebilir hale gelmektedir.

2.3 Yazılım Altyapısı

Sistemin yazılım altyapısında Flask ve Node.js birlikte kullanılmıştır. Flask, Python tabanlı bir web framework olup, API geliştirme ve makine öğrenimi modellerinin entegrasyonu için kullanılmıştır. Flask, veri işleme ve modellerden tahmin sonuçlarının alınması gibi işlemleri yönetir. Node.js ise kullanıcı yönetimi, oturum işlemleri ve veritabanı bağlantıları gibi backend işlevlerini gerçekleştirmek için tercih edilmiştir. Node.js, MongoDB ile entegre edilmiş ve Create, Read, Update, Delete (CRUD) işlemleri başarılı bir şekilde uygulanmıştır. Sistemin arayüzü, kullanıcı dostu bir deneyim sağlamak amacıyla React framework ile geliştirilmiştir. React, dinamik ve modern web uygulamaları oluşturmak için kullanılan bir JavaScript kütüphanesidir. Admin kullanıcıları, sistemdeki tüm kullanıcıları ve verileri yönetebilirken; sağlık çalışanları hasta bilgilerini sisteme ekleyebilir. Hasta kullanıcıları ise kendilerine sunulan teşhis ve önerileri görüntüleyerek sağlık süreçlerini takip edebilir. Bu yapı her kullanıcı rolünün ihtiyaçlarına uygun bir deneyim sunarak sistemin etkinliğini artırmıştır. Backend tarafında Node.js ile yazılmış bir API kullanılmıştır.

Bu API, MongoDB veritabanına erişimi sağlayarak CRUD işlemlerini yönetir. MongoDB veritabanında kullanıcı bilgilerini, VKİ hesaplamasını ve roller içeren "User" tablosu, hastalara ait tahlil sonuçlarını ve teşhis bilgilerini içeren "Test" tablosu bulunmaktadır. JSON Web Token kullanıcı oturumlarını güvenli bir şekilde yönetmek için kullanılır. Middleware kullanıcıların rollerine göre erişim kontrolü sağlar. Örneğin, bir hastanın yalnızca kendi tahlillerine erişebilmesi, adminin tüm verilere erişebilmesi gibi kontrol mekanizmaları uygulanır. Flask Framework ile geliştirilmiş bir Python API, makine öğrenimi modellerinin entegrasyonu için kullanılmaktadır. Bu API; teşhis, genel öneri ve beslenme önerisi gibi farklı sorgular için modellerden gelen sonuçları döndürür. Kullanıcı, giriş yaptığında web token ile kimlik doğrulama yapılır ve rol bazlı yetkilendirme sağlanır. Şifreler bcrypt kullanılarak hashlenmiş şekilde saklandığı için güvenli bir oturum yönetimi gerçekleştirilir. Sağlık çalışanları tarafından girilen tahlil sonuçları, Node.js API üzerinden MongoDB'ye kaydedilir. Python API ile modelden dönen teşhis ve öneriler frontend üzerinden kullanıcıya sunulur.

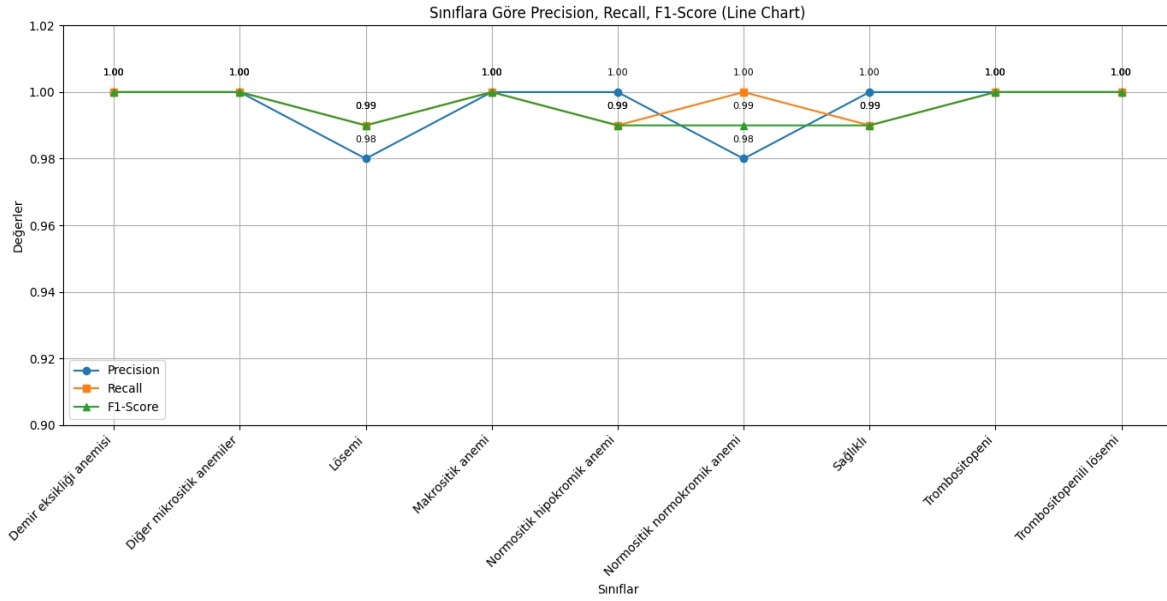
2.4 Random Forest Algoritması

Random Forest Algoritması sınıflandırma temelli tahminlerin gerçekleştirilmesinde yaygın olarak kullanılmaktadır. Random Forest, topluluk (ensemble) öğrenme yöntemlerinden biri olup, çok sayıda karar ağacının (decision tree) bir araya getirilmesiyle oluşturulan güçlü ve esnek bir sınıflandırma algoritmasıdır [19]. Her bir ağaç,

eğitim verisinin rastgele seçilmiş alt örnekleme (bootstrap) üzerinde eğitilmekte ve ağaç oluşturulurken her düğümde özniteliklerin rastgele bir alt kümesi dikkate alınmaktadır. Bu çeşitlendirme, modelin aşırı öğrenme (overfitting) eğilimini azaltırken, genelleme performansını artırmaktadır [20]. Karar ağaçlarının oluşturulmasında Gini indeksi veya bilgi kazancı (entropi) gibi ölçütler kullanılarak veri, homojenliği en üst düzeye çıkaracak şekilde bölünmektedir. Her bir karar ağacı bağımsız olarak tahminde bulunmakta ve son sınıflandırma kararı ise çoğunluk oylaması (majority voting) ile verilmektedir. Bu yapı, özellikle yüksek boyutlu ve dengesiz veri setlerinde öznitelik etkileşimlerini başarılı bir şekilde yakalayarak, kararlı ve yorumlanabilir sonuçlar üretmektedir [21]. Tıbbi veri analizi bağlamında Random Forest algoritmasının tercih edilmesinin temel nedeni, eksik verilerle başa çıkabilme yeteneği, değişken önem düzeylerini belirleyebilme özelliği ve sınıflar arası ayırım gücünün yüksek olmasıdır. Ayrıca bu algoritma azınlık sınıflarının tespiti konusunda da güçlü bir performans sergilemekte olup, dengesiz veri setlerinde uygulanabilirliği oldukça yüksektir [22]. Random Forest Algoritmasının veri seti üzerindeki performansı doğruluk, precision, recall ve F1-score metrikleriyle değerlendirilmiştir (Şekil 1 ve 2). Sistem, Random Forest algoritması kullanarak %99.66 oranında etkin bir teşhis doğruluğu elde etmiş, böylece geleneksel yöntemlere kıyasla daha hızlı ve tutarlı sonuçlar üretebilmiştir.

Sınıflara Göre Precision, Recall ve F1-Score:					
	Class	Precision	Recall	F1-Score	
0	Demir eksikliği anemisi	1.00	1.00	1.00	
1	Diğer mikrositik anemiler	1.00	1.00	1.00	
2	Lösemi	0.98	0.99	0.99	
3	Makrositik anemi	1.00	1.00	1.00	
4	Normositik hipokromik anemi	1.00	0.99	0.99	
5	Normositik normokromik anemi	0.98	1.00	0.99	
6	Sağlıklı	1.00	0.99	0.99	
7	Trombositopeni	1.00	1.00	1.00	
8	Trombositopenili lösemi	1.00	1.00	1.00	

Şekil 1: Random forest algoritmasının sınıflar üzerindeki performansı.



Şekil 2: Random forest performansının grafiği.

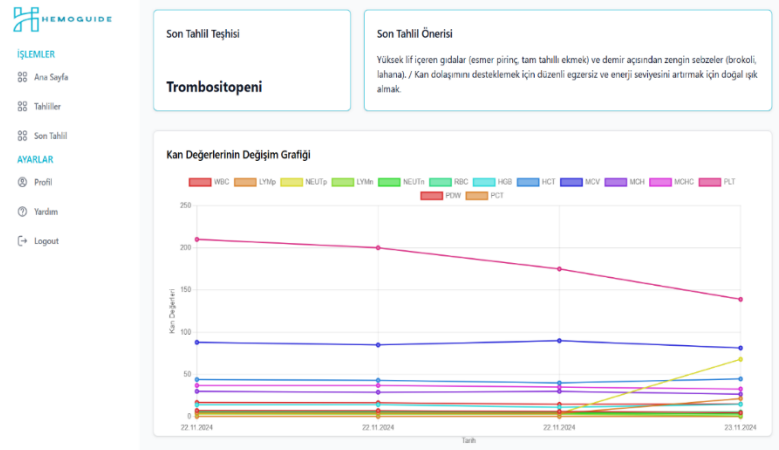
2.5 Hemoguide Uygulaması

Bu çalışmada gerçekleştirilen sistem üç temel kullanıcı rolüne hitap eden admin (Şekil 3), hasta (Şekil 4), ve sağlık çalışanı (Şekil 5), arayüzleri olarak tasarlanmıştır. Yönetici, sistemin genel yönetiminden ve diğer kullanıcıların kontrolünden sorumludur. Anasayfada sistemde kayıtlı toplam hasta sayısı, tahlil sayısı ve sağlık çalışanı sayısı bilgi kutuları şeklinde gösterilmektedir. Tahlil sonuçlarındaki teşhis türlerinin dağılımını ve hastaların cinsiyet ve VKİ kategorilerine göre dağılımlarını gösteren grafikler bulunmaktadır. Tablolar ile sisteme kayıtlı kullanıcıların bilgilerini görüntüleyebilmektedir (Şekil 6). Dinamik bir

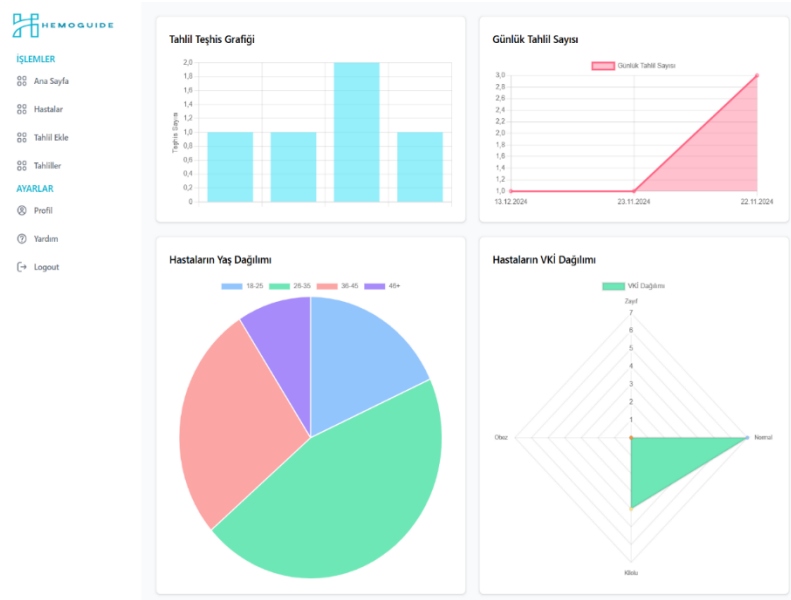
veritabanına bağlı olarak yeni hastalar ve sağlık çalışanları ekleyebilmektedir (Şekil 7). Kullanıcı hastaların geçmiş bilgileri saklanmaktadır. Bu sayede hasta geçmişteki tahlilleriyle güncel tahlillerini kıyaslayabilmektedir (Şekil 8). Güncel tahlilini detaylı olarak ayrı bir sayfada görüntüleyebilmektedir (Şekil 9). Sağlık çalışanları için tasarlanan arayüzde hasta seçimi yapılarak sisteme yeni tahliller eklenebilmektedir (Şekil 10). Hasta seçiminden sonra hastanın bilgilerine göre otomatik olarak teşhis ve öneriler sunmaktadır. Sağlık çalışanı bu öneri ve sonuçları düzenleyebilmektedir.



Şekil 3: Admin arayüzü.



Şekil 4: Hasta arayüzü.



Şekil 5: Sağlık çalışanı arayüzü.

TARİH	DOKTORUN ADI	HASTANIN ADI	TANI	ÖNERİLER
13.12.2024	Ahmet Yılmaz	Tuğçe Güzle	Demir eksikliği anemisi	İspanak, brokoli gibi yeşil yapraklı sebzeler ve meyveler (çilek, kiwi). Dozenli :
23.11.2024	Ahmet Yılmaz	Kemal Yıldız	Trombositopeni	Yüksek lif içeren gıdalar (esmer pirinç, tam tahıllı ekmek) ve demir açısından
22.11.2024	Zeynep Kaya	Meryem Çetin	Sağlıklı	Sebzeler (spanak, kabağı) ve yeterli su alımı. / Haftada en az 150 dakika orta
22.11.2024	Mehmet Aydın	Elif Güneş	İzlemi	Sebzeler (spanak, brokoli) ve antioksidan zengin meyveler (çilek). / Hafif ter
22.11.2024	Mehmet Aydın	Elif Güneş	Normosiler normokromik anemi	Sebzeler (brokoli, spanak) ve narenciye (portakal, mandalina). Yeterli su alı
22.11.2024	Ali Çelik	Elif Güneş	Normosiler normokromik anemi	Yüksek protein (tavuk, hindi) ve lif içeren yüksek sebzeler (brokoli, brokoli) /
22.11.2024	Zeynep Kaya	Meryem Çetin	Normosiler normokromik anemi	Sebzeler (brokoli, spanak) ve narenciye (portakal, mandalina). Yeterli su alı
22.11.2024	Ahmet Yılmaz	Kemal Yıldız	Normosiler normokromik anemi	Protein kaynakları (tavuk, balık) ve lifli gıdalar (tam buğday ürünleri). / Kan d
22.11.2024	Fatma Güneş	Elif Güneş	Sağlıklı	Sebzeler (spanak, kabağı) ve yeterli su alımı. / Haftada en az 150 dakika orta
22.11.2024	Emine Demir	Elif Yılmaz	Sağlıklı	Sebzeler (spanak, kabağı) ve yeterli su alımı. / Haftada en az 150 dakika orta
22.11.2024	Mehmet Aydın	Serkan Özdemir	Sağlıklı	Yüksek protein (hindi, balık) ve sebzeler. / Günlük adım hedefi koyarak aktif
22.11.2024	Ahmet Yılmaz	Kemal Yıldız	Sağlıklı	Dengeli protein (tavuk, balık) ve lifli gıdalar. / Fiziksel olarak aktif kalmak için
22.11.2024	Ahmet Yılmaz	Kemal Yıldız	Sağlıklı	Dengeli protein (tavuk, balık) ve lifli gıdalar. / Fiziksel olarak aktif kalmak için

Şekil 6: Yönetici kullanıcı görüntüleme arayüzü.

HEMOGUIDE

İŞLEMLER

- Ana Sayfa
- Hastalar
- Sağlık Çalışanları
- Tahilliler
- Sağlık Çalışanı Ekle
- Hasta Ekle

AYARLAR

- Profil
- Yardım
- Logout

Hasta Ekle

Ad: Soyadı:

Doğum Tarihi: Boy (cm):

Ağırlık (kg): Cinsiyet:

Şifre: Email:

Telefon:

Ekle

Şekil 7: Yönetici kullanıcı ekleme arayüzü.

TARİH	TANI	ÖNERİ
22.11.2024	Sağlıklı	Dengeli protein (tavuk, balık) ve lifli gıdalar. / Fiziksel olarak aktif kalmak için düzenli yürüyüşler ve düzenli uyku.
22.11.2024	Sağlıklı	Dengeli protein (tavuk, balık) ve lifli gıdalar. / Fiziksel olarak aktif kalmak için düzenli yürüyüşler ve düzenli uyku.
22.11.2024	Normositer normokromik anemi	Protein kaynaktan (tavuk, balık) ve lifli gıdalar (tam buğday ürünleri). / Kan dolaşımını artırmak için hafif koşu veya yüzme gibi d
23.11.2024	Trombositopeni	Yüksek lif içeren gıdalar (esmer pirinç, tam tahıllı ekmek) ve demir açısından zengin sebzeler (brokoli, lahan). / Kan dolaşımını d

Şekil 8: Hasta geçmiş tahliller görüntüleme arayüzü.

Son Tahlil Bilgileri			
Doktorun Adı:	Ahmet Yılmaz	Tarih:	23.11.2024
Tanı:	Trombositopeni	Öneri:	Yüksek lif içeren gıdalar (esmer pirinç, tam tahıllı ekmek) ve demir açısından zengin sebzeler (brokoli, lahan). / Kan dolaşımını desteklemek için düzenli egzersiz ve enerji seviyesini artırmak için doğal ışık almak.
WBC:	4.7	LYM%:	21.5
NEUT%:	68	LYM#:	1
NEUT#:	3.2	RBC:	5.5
HGB:	14.7	HCT:	44.7
MCV:	81.4	MCH:	26.7
MCHC:	32.8	PLT:	139
PDW:	14.9	PCT:	0.14

Şekil 9: Hasta son tahlil görüntüleme arayüzü.

Şekil 10: Sağlık çalışanı tahlil ekleme arayüzü.

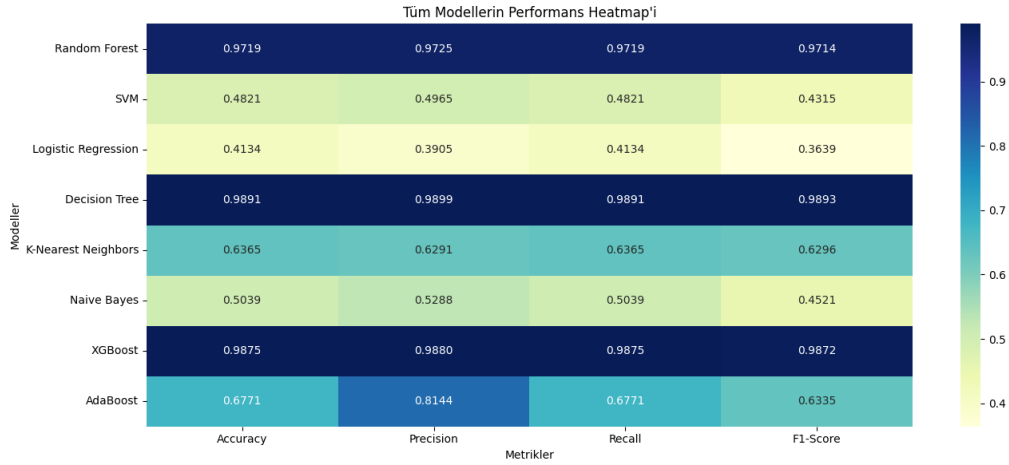
3. TARTIŞMA VE SONUÇ

Bu çalışmada hematolojik bozuklukların teşhisi ve bireyselleştirilmiş tedavi stratejilerinin belirlenmesi amacıyla geliştirilen makine öğrenimi tabanlı karar destek sistemi, hemogram verilerinin analizine dayalı olarak yüksek doğruluk oranıyla tanı koyabilme ve kişiye özgü beslenme ile yaşam tarzı önerileri sunma kapasitesini göstermiştir. SMOTE yöntemi uygulanmadan önce ve uygulandıktan sonra çeşitli makine öğrenmesi algoritmaları (Random Forest, SVM, Logistic Regression, Decision Tree, K-Nearest Neighbors, Naive Bayes, XGBoost, AdaBoost) uygulanarak performans karşılaştırmaları yapılmıştır (Şekil 11–14). SMOTE yöntemi ile dengelenmiş veri seti kullanılarak elde edilen doğruluk oranları sırasıyla Random Forest için %99,66, SVM için %63,96, Logistic Regression için %66,28, Decision Tree için %99,50, K-Nearest Neighbors için %86,61, Naive Bayes için %63,30, XGBoost için %98,70 ve AdaBoost için %67,70

olarak hesaplanmıştır. Analiz sonuçları, SMOTE yöntemiyle dengesiz veri probleminin etkili bir şekilde giderildiğini ve özellikle bazı sınıflandırma algoritmalarının performansında iyileşmeler sağlandığını göstermektedir. Özellikle Random Forest ve Decision Tree gibi ağaç tabanlı yöntemlerin SMOTE sonrası belirgin bir şekilde yüksek doğruluk değerlerine ulaşması, bu algoritmaların sınıf dengesizliklerine karşı daha dayanıklı olduğunu ortaya koymaktadır. Random Forest algoritması, %99,66 doğruluk oranı ile tüm diğer yöntemlere göre en başarılı model olarak öne çıkmıştır. Bu yüksek başarı oranı, modelin karmaşık veri yapılarını öğrenme kapasitesinin yüksekliğini ve SMOTE ile desteklenmiş dengeli veriler üzerinde güçlü bir genelleme yeteneğine sahip olduğunu göstermektedir. Bu bağlamda, özellikle sağlık gibi hata toleransının düşük olduğu alanlarda, Random Forest gibi sağlam ve güvenilir modellerin kullanılması önerilmektedir.

Tüm Modellerin Detaylı Performans Tablosu:				
Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.971919	0.972495	0.971919	0.971386
SVM	0.482059	0.496525	0.482059	0.431502
Logistic Regression	0.413417	0.390540	0.413417	0.363896
Decision Tree	0.989080	0.989898	0.989080	0.989318
K-Nearest Neighbors	0.636505	0.629106	0.636505	0.629648
Naive Bayes	0.503900	0.528845	0.503900	0.452123
XGBoost	0.987520	0.987990	0.987520	0.987234
AdaBoost	0.677067	0.814432	0.677067	0.633539

Şekil 11: SMOTE Yöntemi Öncesi Model Performansları.

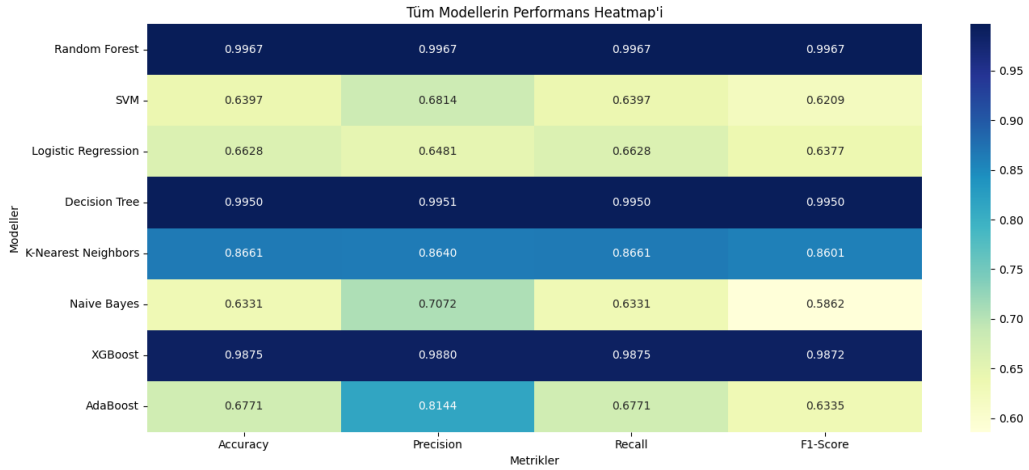


Şekil 12: SMOTE yöntemi öncesi model performanslarının grafiği.

Tüm Modellerin Detaylı Performans Tablosu:

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.996694	0.996748	0.996694	0.996697
SVM	0.639669	0.681431	0.639669	0.620903
Logistic Regression	0.662810	0.648051	0.662810	0.637733
Decision Tree	0.995041	0.995115	0.995041	0.995041
K-Nearest Neighbors	0.866116	0.863989	0.866116	0.860136
Naive Bayes	0.633058	0.707210	0.633058	0.586197
XGBoost	0.987520	0.987990	0.987520	0.987234
AdaBoost	0.677067	0.814432	0.677067	0.633539

Şekil 13: SMOTE yöntemi sonrası model performansları.



Şekil 14: SMOTE yöntemi sonrası model performanslarının grafiği.

Sistem, hastaların teşhis bilgileri, VKİ ve demografik özelliklerini dikkate alarak kişiselleştirilmiş beslenme ve yaşam tarzı önerileri sunmuş, böylece tedavi süreçlerinin optimizasyonuna önemli katkılar sağlamıştır.

Bununla birlikte, nadir görülen sınıflarda performans düşüklüğü gözlemlenmiş ve modelin daha çeşitli ve kapsamlı veri kümeleriyle geliştirilmesi gerekliliği ortaya çıkmıştır. Uzman görüşlerinin entegrasyonu ve farklı

coğrafi/demografik popülasyonlardan elde edilecek verilerle sistemin güncellenmesi, önerilerin evrensel geçerliliğini artırabilir. Sistemin klinik uygulamalardaki potansiyeli, sağlık çalışanlarının karar verme süreçlerini destekleyerek tanı ve tedavi protokollerini iyileştirmesiyle öne çıkmaktadır. Sonuç olarak, bu çalışma, yapay zeka tabanlı KDS'lerin sağlık sektöründeki dönüştürücü etkisini vurgulamakta olup, veriye dayalı tıbbın gelişimine yönelik önemli bir adım teşkil etmektedir. Sistemin ölçeklenebilirliği ve adaptasyonu, gelecekte daha geniş kullanıcı kitlelerine ulaşmasını ve klinik etkinliğini artırmasını mümkün kılacağı öngörülmektedir.

Yazarların Katkıları: İlk yazar Öznur SUÇEKEN analiz ve yorum, makale yazımı ve denetleme hususlarında, ikinci yazar Tuğçe GÜZLE fikir, literatür taraması, veri toplama ve tasarım hususlarında çalışmaya katkıda bulunmuştur.

Çıkar Çatışmaları: Yazarlar çalışmada herhangi bir çıkar çatışması olmadığını belirtmektedirler.

4. KAYNAKLAR

- [1] M. N. Garcia-Casal, O. Dary, M. E. Jefferds, and S. R. Pasricha, "Diagnosing anemia: Challenges selecting methods, addressing underlying causes, and implementing actions at the public health level," *Ann. N. Y. Acad. Sci.*, vol. 1524, no. 1, pp. 37-50, 2023.
- [2] Y. El Alaoui, A. Elomri, M. Qaraqe, R. Padmanabhan, R. Y. Taha, H. El Omri, and O. Aboumarzouk, "A review of artificial intelligence applications in hematology management: current practices and future prospects," *J. Med. Internet Res.*, vol. 24, no. 7, p. e36490, 2022.
- [3] A. Srisuwananukorn, M. E. Salama, and A. T. Pearson, "Deep learning applications in visual data for benign and malignant hematologic conditions: a systematic review and visual glossary," *Haematologica*, vol. 108, no. 8, p. 1993, 2023.
- [4] Y. Zhang, Y. Weng, and J. Lund, "Applications of explainable artificial intelligence in diagnosis and surgery," *Diagnostics*, vol. 12, no. 2, p. 237, 2022.
- [5] N. B. Noor, U. A. Oyshi, A. Das, and K. Iqbal, "A systematic approach to predict anemia from eye conjunctiva images," in *Proc. 26th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Cox's Bazar, Bangladesh, Dec. 13-15, 2023, pp. 1-5.
- [6] U. Ponnusamy, D. Darshan, and N. Sampathila, "Approaching explainable artificial intelligence methods in the diagnosis of iron deficiency anemia using blood parameters," in *Proc. Int. Conf. Recent Adv. Inf. Technol. Sustain. Dev. (ICRAIS)*, Manipal, India, Nov. 6-7, 2023, pp. 201-206.
- [7] A. Tummala and K. Parvataneni, "Comparative study of AI-based anemia prediction using eye conjunctiva images and blood tests," in *Proc. Int. Workshop Artif. Intell. Image Process. (IWAIIIP)*, Yogyakarta, Indonesia, Dec. 1-2, 2023, pp. 87-91.
- [8] A. G. Mugdha, F. T. Pinki, and S. K. Talukdhar, "Hemoglobin estimation and anemia severity prediction using machine learning algorithms," in *Proc. 5th Int. Conf. Sustain. Technol. Ind. 5.0 (STI)*, Dhaka, Bangladesh, Dec. 9-10, 2023, pp. 1-6.
- [9] M. M. Rahman, M. U. Mojumdar, H. A. Shifa, N. R. Chakraborty, N. P. Stenin, and M. A. Hasan, "Anemia disease prediction using machine learning techniques and performance analysis," in *Proc. 11th Int. Conf. Comput. Sustain. Global Dev. (INDIACom)*, New Delhi, India, Feb. 28-Mar. 1, 2024, pp. 1276-1282.
- [10] E. Aboelnaga, "Anemia types classification," *Kaggle, Dataset*, 2022. [Online]. Available: <https://www.kaggle.com/datasets/eboelnaga/anemia-types-classification>
- [11] M. Levi, M. Simonetti, E. Marconi, O. Brignoli, M. Cancian, A. Masotti, and F. Lapi, "Gender differences in determinants of iron-deficiency anemia: a population-based study conducted in four European countries," *Ann. Hematol.*, vol. 98, no. 7, pp. 1573-1582, 2019.
- [12] R. Krisnanda, "Vitamin C helps in the absorption of iron in iron deficiency anemia," *J. Penelit. Perawat. Prof.*, vol. 2, no. 3, pp. 279-286, 2020.
- [13] J. Deng, L. Ramelli, P. Y. Li, A. Eshaghpour, G. E. M. Schuenemann, and M. A. Crowther, "Efficacy of vitamin C with iron supplementation in iron deficiency anemia patients: A systematic review and meta-analysis," *Blood*, vol. 142, p. 1091, 2023.
- [14] J. Deng, L. Ramelli, P. Y. Li, A. Eshaghpour, G. E. M. Schuenemann, and M. A. Crowther, "Efficacy of vitamin C with Fe supplementation

- in patients with iron deficiency anemia: A systematic review and meta-analysis," *Blood Vessels Thromb. Hemost.*, vol. 1, no. 4, p. 100023, 2024.
- [15] Y. Qin, A. Melse-Boonstra, X. Pan, B. Yuan, Y. Dai, J. Zhao, and Z. Shi, "Anemia in relation to body mass index and waist circumference among Chinese women," *Nutr. J.*, vol. 12, no. 10, pp. 1-3, 2013.
- [16] A. C. Cepeda-Lopez, S. J. Osendarp, A. Melse-Boonstra, I. Aeberli, F. Gonzalez-Salazar, E. Feskens, and M. B. Zimmermann, "Sharply higher rates of iron deficiency in obese Mexican women and children are predicted by obesity-related inflammation rather than by differences in dietary iron intake," *Am. J. Clin. Nutr.*, vol. 93, no. 5, pp. 975-983, 2011.
- [17] A. Peña-Fernández, M. Higuera, E. Segura, M. D. Evans, and M. D. L. Á. Peña, "Inadequacies of dietary iron intake in normal- and overweight young university students from Leicester, England," *Biol. Life Sci. Forum*, vol. 38, no. 1, p. 6, 2025.
- [18] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321-357, 2002.
- [19] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5-32, 2001.
- [20] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18-22, 2002.
- [21] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3133-3181, 2014.
- [22] H. Chen, et al., "A random forest model based classification scheme for medical diagnosis," *Comput. Math. Methods Med.*, 2017.