FORBES TIP DERGİSİ

FORBES JOURNAL OF MEDICINE

Cilt / Volume: 6 Sayı / Issue: 2 Ağustos / August 2025

www.forbestip.org
e-ISSN: 2757-5241

Sağlıklı Yarınlar için Bilimin İzindeyiz

FORBES
TIP DERGİSİ - JOURNAL OF MEDICINE

e-ISSN: 2757-5241

Ağustos / August 2025
Cilt / Volume: 6
Sayı / Issue: 2

Forbes Tıp Dergisi, İzmir Buca Seyfi Demirsoy Eğitim ve Araştırma Hastanesi'nin yayın organıdır.

Dergi dört ayda bir yılda 3 sayı (Nisan, Ağustos, Aralık) olarak yayımlanan açık erişim, ücretsiz ve hakemli bir dergidir.

"Amaç ve Kapsam", "Yayın Etiği", "Yazarlar için Kontrol Listesi" ve "Makale Hazırlama" hakkında daha fazla bilgi için lütfen derginin web sayfasını **(https://forbestip.org)** ziyaret edin.

Derginin editöryal ve yayın süreçleri ile etik kuralları ICMJE, WAME, CSE, COPE, EASE ve NISO gibi uluslararası kuruluşların kurallarına uygun olarak şekillenmektedir.

Forbes Tıp Dergisi, **Ulakbim TR Dizin, EBSCO, Türk Medline, J-Gate, Türkiye Atıf Dizini, Gale** ve **DOAJ** tarafından indekslenmektedir.

Dergi elektronik ortamda yayımlanmaktadır.

Kurucu: Buca Seyfi Demirsoy Eğitim ve Araştırma Hastanesi Başhekimi Prof. Dr. M. Yekta Öncel

Baş Editör: Prof. Dr. M. Yekta Öncel

Forbes Journal of Medicine is the publication of İzmir Buca Seyfi Demirsoy Training and Research Hospital.

It is published three times a year (April, August, December) FJM is an open access, free and peer-reviewed Journal

Please refer to the journal's webpage **(https://forbestip.org)** for "Aims and Scope", "Ethical Policy", "Checklist for Authors" and "Manuscript Preparation".

The editorial and publication process of Forbes Journal of Medicine are shaped in accordance with the guidelines of the ICMJE, WAME, CSE, COPE, EASE, and NISO.

Forbes Journal of Medicine is indexed in **Ulakbim TR Dizin, Turkey Citation Index, EBSCO, Türk Medline, J-Gate, Gale** and **DOAJ**.

The journal is published electronically.

Owner: M. Yekta Öncel, M.D. Prof. on Behalf of Buca Seyfi Demirsoy Training and Research Hospital Head Physician

Editor in Chief: M. Yekta Öncel, M.D. Prof.

Yönetim Yeri/Administrative Office
İzmir Buca Seyfi Demirsoy EAH, Başhekimlik, Buca/İzmir
Tel: +90 (232) 452 52 52 /4293 **E-mail:** forbestipdergisi@gmail.com

Yayın Türü/Publication Type
Yaygın Süreli/Scholarly Periodical

Yayınevi/Publisher
GALENOS YAYINEVİ
Molla Gürani Mah. Kaçamak Sk. No: 21/1 34093 Fındıkzade, Fatih, İstanbul, Türkiye
Tel: +90 (530) 177 30 97 **E-mail:** info@galenos.com.tr/yayin@galenos.com.tr **Web:** www.galenos.com.tr **Yayıncı Sertifika No:** 14521

Yayınlanma Tarihi/Online Publication Date
Ağustos 2025/August 2025

FORBES
TIP DERGİSİ - JOURNAL OF MEDICINE

Yayın Kurulu

FORBES
TIP DERGİSİ - JOURNAL OF MEDICINE

Editorial Board

FORBES
TIP DERGİSİ - JOURNAL OF MEDICINE

Editöryal Danışma Kurulu

FORBES
TIP DERGİSİ - JOURNAL OF MEDICINE

Editorial Advisory Board

FORBES
TIP DERGİSİ - JOURNAL OF MEDICINE

İçindekiler/Contents

FORBES
TIP DERGİSİ - JOURNAL OF MEDICINE

İçindekiler/Contents

DOI: 10.4274/forbes.galenos.2025.00378
Forbes J Med 2025;6(2):97-102

Derleme / Review

Candida (Candidozyma) Auris: Kolonizasyondan İnvaziv Enfeksiyona Giden Süreç

Candida (Candidozyma) Auris: From Colonization to Invasive Disease

Burcu AÇIKALIN ARIKAN, Nurbanu SEZAK

Demokrasi Üniversitesi Buca Seyfi Demirsoy Eğitim ve Araştırma Hastanesi, Enfeksiyon Hastalıkları ve Klinik Mikrobiyoloji Kliniği, Anabilim Dalı, Türkiye

Atıf: Açıkalın Arıkan B, Sezak N. Candida (Candidozyma) auris: from colonization to invasive disease. Forbes J Med. 2025;6(2):97-102

ÖZ

Günümüzde sağlık kuruluşlarında Candidozyma auris (C. auris) enfeksiyonları, direnç ve önlenmesi güç çapraz bulaş riski nedeniyle endişe edici hale gelmiştir. Ülkemizde de 2021 yılından itibaren birçok ilde C. auris enfeksiyonları ve salgınları bildirilmeye başlanmıştır. Hastadan hastaya çok hızlı yayılım gösteren bu Candida türü, antifungallere dirençli olması nedeni ile de kontrol altına alınması zor bir sağlık bakımı ilişkili enfeksiyon olarak karşımıza çıkmaktadır. Uzun süre antibiyotik kullanan, santral ve periferik vaskuler kateter gibi etkenin kolonize olup kolayca tutunabileceği cihazlarla takip edilen, ileri yaş, cerrahi girişim geçiren, immunsupresif hastalarda kolayca kolonize olup enfeksiyona sebep olmaktadır. Yapılan çalışmalarda, kolonizasyonun aylarca ve yıllarca devam edebildiği bildirilmektedir. Kan dolaşımı enfeksiyonları, miyokardit, cerrahi alan enfeksiyonu, yanık enfeksiyonları, kateter yerinde apse, otit, menenjit, kemik eklem enfeksiyonları gibi birçok enfeksiyona sebep olabilmektedir. Enfekte veya kolonize hastaların mümkün olduğunca tek kişilik odalarda izlenmesi, hastaya, hasta yakınlarına ve sağlık personellerine el hijyeni ve izolasyon önlemleri ile ilgili eğitim verilmesi gerekmektedir.

Anahtar Kelimeler: Candida, Candidozyma auris, antifungal direnç, çoklu ilaca dirençli mantarlar, enfeksiyon kontrolü

ABSTRACT

Currently, Candida auris (C. auris) (infections in healthcare settings have become increasingly concerning due to their antifungal resistance and the difficulties associated with preventing cross-contamination. Since 2021, C. auris cases and outbreaks have been reported in several provinces across our country. This rapidly spreading Candida species represents a healthcare-associated infection that is particularly challenging to control due to its resistance profile. It easily colonizes and causes infections in patients with prolonged antibiotic use, those with indwelling devices such as central and peripheral vascular catheters that facilitate colonization and adherence, elderly individuals, surgical patients, and immunosuppressed individuals. Studies have shown that colonization can persist for months or even years. C. auris may lead to a variety of infections, including bloodstream infections, myocarditis, surgical site infections, burn wound infections, catheter site abscesses, otitis, meningitis, and bone and joint infections. Whenever possible, infected or colonized patients should be cared for in single-occupancy rooms, and education on hand hygiene and isolation precautions should be provided to patients, their families, and healthcare workers.

Keywords: Candida, Candidozyma auris, antifungal resistance, multidrug-resistant fungi, infection control

Geliş/Received: 17.06.2025
Kabul/Accepted: 08.07.2025
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Sorumlu Yazar/ Corresponding Author:
Dr. Burcu AÇIKALIN ARIKAN, Demokrasi Üniversitesi Buca Seyfi Demirsoy Eğitim ve Araştırma Hastanesi, Enfeksiyon Hastalıkları ve Klinik Mikrobiyoloji Anabilim Dalı, İzmir, Türkiye
dr.hburcua@gmail.com
ORCID: 0000-0002-4955-0789

GİRİŞ

Kandida türlerinin etken olduğu ciddi enfeksiyonların görülme oranı son yıllarda giderek artmaktadır. Özellikle sağlık bakımı ilişkili enfeksiyonlar arasında önemli bir yere sahip olan invaziv kandidiyaz, başta yoğun bakım ünitelerinde izlenenler olmak üzere bağışıklığı baskılanmış hastalar için ciddi bir morbidite ve mortalite nedenidir. Bu durumun hem modern sağlık hizmetlerinin başarısı hem de karmaşıklığıyla doğrudan ilişkili olduğunu söylemek mümkündür. Uzun süreli ve geniş spektrumlu antibiyotik kullanımı, hastanede yatış süresinin uzaması, invaziv cihaz kullanımının artması, parenteral beslenme gibi faktörler hem mukoza bariyerlerini bozmakta hem de bakteriyel florayı baskılayarak kandideminin gelişimine zemin hazırlamaktadır.[1] Hematolojik maligniteler, solid organ nakli, kanser kemoterapisi ve insan bağışıklık yetmezliği virüsü/edinilmiş bağışıklık yetmezliği sendromu gibi bağışıklığı baskılanmış hasta popülasyonu, prematür yenidoğanlar ve kronik hastalıkları bulunan çocuklar da giderek artan oranda dissemine kandidiyaz gelişimi için risk altındadır.[2]

Son yıllarda ortaya çıkan Candidozyma auris (*C. auris*) gibi çoklu antifungal direnç gösteren türler, bu artışı daha da karmaşık hale getirmiştir. *C. auris*, hem çevresel direnç hem de antifungal ajanlara karşı toleransı nedeniyle salgınlara neden olabilmekte, geleneksel enfeksiyon kontrol önlemleriyle elimine edilmesi zor olmaktadır.[3] Bu nedenle yalnızca bireysel değil, kurumsal düzeyde sağlık hizmeti sunumunu tehdit etmektedir. Tanı yöntemlerinde yaşanan teknolojik gelişmeler sayesinde daha önce tanınamayan ya da yanlış sınıflandırılan Candida türleri artık doğru biçimde tanımlanabilmektedir. Matrix yardımlı lazer desorpsiyon/iyonizasyon-uçuş zamanlı kütle spektrometrisi (MALDI-TOF MS), polimeraz zincir reaksiyonu (PZR) tabanlı teknikler ve yeni nesil sekanslama yöntemleri, dissemine kandidiyaz olgularının daha fazla tanınmasına olanak sağlamıştır.[4] Bu durum bildirilen olgu sayılarında artışa yol açsa da aynı zamanda klinik farkındalığın da yükseldiğini göstermektedir.

İnvaziv Candida enfeksiyonlarındaki artış yalnızca mikroorganizmanın biyolojik özelliklerinden değil, aynı zamanda modern tıbbın uygulamaları, demografik geçişler ve küresel sağlık sistemlerinin dinamiklerinden kaynaklanan çok boyutlu bir sorundur. Bu durum, enfeksiyon kontrol önlemlerinin güçlendirilmesini, antifungal tedavi protokollerinin yeniden gözden geçirilmesini ve hastane içi sürveyansın etkin bir şekilde yürütülmesini zorunlu kılmaktadır.[5]

Günümüzde sağlık kuruluşlarında *C. auris* enfeksiyonları, direnç ve önlenmesi güç çapraz bulaş riski nedeniyle endişe edici hale gelmiştir.[1,2] Candida haemulonii kompleksi içerisinde yer alan *C. auris*, ilk olarak 2007 yılında Japonya'da bir hastanın dış kulak yolu akıntısından izole edilmiş, 2009 yılında tanımlanmış ve latincede kulak anlamına gelen "auris" olarak isimlendirilmiştir.[4] Daha sonra birçok kıtada ve ülkede *C. auris* invaziv enfeksiyonları ve salgınları bildirilmiştir.[5,6] Ülkemizde de 2021 yılından itibaren birçok ilde *C. auris* enfeksiyonları ve salgınları bildirilmeye başlanmıştır.[7] Çok hızlı yayılım gösteren bu candida türü, antifungallere dirençli olması nedeni ile de kontrol altına alınımı zor olan bir sağlık bakımı ilişkili enfeksiyon olarak karşımıza çıkmaktadır.[8-10] Bu derlemede Candida auris yerine Candidozyma auris ismi kullanılacak olup, *C. auris* mikrobiyolojisi, epidemiyolojisi, kolonizasyonu, enfeksiyonları, tanısı, tedavisi ve önleme yaklaşımları anlatılacaktır.

Mikroorganizmanın özellikleri

C. auris ilk keşfedildiğinde Candida auris olarak isimlendirilmiş olup, Candida cinsine ait bir maya türü olarak tanımlanmıştır. İlk olarak Japonya'da bir hastanın dış kulak yolundan izole edildiği için bu ismi almıştır.[3] *C. auris* tek, çift veya grup halinde olabilen hücrelere sahip tomurcuklanan bir mayadır. Hücreler oval, elipsoidal ve 2,5-5,0 mikron boyutundadır. *C. auris* nadiren hif veya psödohif oluşturur ancak germ tüp oluşturmaz.[11] Maya genomik deoksiribonükleik asit analizi, Candida ruelliae, Candida haemulonii, C. duobushaemulonii ve C. pseudohaemulonii'ye yakın filogenetik profillere sahip farklı bir tür olduğunu göstermiştir.[12] *C. auris* enfeksiyonlarıyla ilişkili virülans faktörleri tam olarak anlaşılamamış olmasına rağmen birkaç virülans stratejisi bildirilmektedir. Candida türlerinde biyofilm oluşumu önemli bir virulans faktörüdür ancak *C. auris*'in biyofilm üretimi C. albicans'a göre daha düşüktür. Ancak yine de yüzeylere yapışma yeteneği gösterir. *C. auris*'in enfeksiyona neden olmak için kullandığı başarılı stratejiler arasında ürettiği enzimler ile nötrofil saldırısından kaçma ve yaygın olarak kullanılan antifungal ilaçlarla tedaviye direnme yeteneği yer almaktadır. Doğal immüniteden diğer Candida türlerine kıyasla daha kolay kaçabilir.[131]

Epidemiyolojisi

C. auris enfeksiyonları ve salgınları 2009 yılında Japonya'da izole edildiği andan itibaren artarak farklı ülkelerden bildirilmeye devam etmektedir. 2013 yılına gelindiğinde *C. auris* Hindistan, Güney Kore, Güney Afrika ve Venezuela'da tanımlanmış ve o zamandan beri küresel olarak yayılmıştır.[9,14,15] Türkiye'de de 2021 yılından itibaren birçok ilde *C. auris* enfeksiyonları ve salgınları bildirilmeye başlanmıştır.[8-10] 2023 yılında sadece Amerika Birleşik Devletleri'nde 4514 yeni olgu bildirilmiştir.[14] Etken, Dünya Sağlık Örgütü tarafından kritik öncelik grubunda sınıflandırılmış ve bir hastaneye bir kez yerleştikten sonra

eliminasyonunun neredeyse imkansız olduğu belirtilmiştir.[15] Etkenin tanımlanmasından sonraki ilk yıllarda sporadik olgu bildirimleri yayınlanırken günümüzde bu durum şekil değiştirmiş ve hastane içi salgınlar bildirilmeye başlanmıştır.[16] Türkiye Cumhuriyeti Sağlık Bakanlığı tarafından yayınlanan 1 Eylül-31 Kasım 2024 tarihleri arasındaki sürveyans verilerine göre 116 hastaneden, 237'si Sağlık hizmeti ilişkili enfeksiyon (SHİE), 278'i SHİE dışı olmak üzere toplam 515 *C. auris* enfeksiyonu bildirimi yapılmıştır. En çok bildirim yapan hastaneler, eğitim ve araştırma hastaneleri ile devlet hastaneleridir. Bildirimler en çok İstanbul ilinden, ardından Ankara ve Adana'dan yapılmıştır. Ülkemizde enfeksiyon bildirimlerinin sağlıklı ve güvenilir olabilmesi için her hastanede tanı koyma kapasitelerinin güçlendirilmesi ve sürveyansın eksiksiz yapılması gerekmektedir.[17]

Kontaminasyon, Kolonizasyon ve Enfeksiyon

C. auris türünün, insan patojeni olarak ortaya çıkmadan önce doğada var olduğu varsayılmaktadır. Sulak alanlarda, denizler ve nehirlerde aylarca kalabildiği, su ile teması olan plastik ve cam yüzeylerde biyofilm oluşturarak uzun süreler kolonize olabildiği bilinmektedir. Çevre koşullarına ve dezenfektanlara dirençli yapısı, hastanede kullanılan ekipmanlar ve çevresel yüzey kontaminasyonu aracılığı ile hastadan hastaya hızla yayılmasına sebep olur.[18-20] Eldivenler, önlükler, hasta monitörleri, tansiyon aleti manşonları, pansuman arabaları ve yataklar yüzey kolonizasyonunun en sık görüldüğü yerlerdir.[20,21]

Uzun süre antibiyotik kullanan, santral ve periferik kateter gibi etkenin kolonize olup kolayca tutunabileceği cihazlarla takip edilen, ileri yaş, cerrahi girişim geçiren, immünosüpresif hastalarda kolayca kolonize olup enfeksiyona sebep olmaktadır. Özellikle aksilla ve kasık derisinde kolonizasyonu olan hastalar bulaş kaynağıdır. Maruziyetten birkaç saat sonra kolonizasyon oluşabilmekte ve aylarca devam edebilmektedir. *C. auris*, vücudun tüm boşluklarından, sıvılarından, deriden (özellikle aksiller bölge ve kasık bölgesi), tüm organlarından izole edilmiştir. Deri, ürogenital bölge ve solunum yolu sekresyonlarının kültürlerindeki üremeler, öncelikle kolonizasyon lehine değerlendirilmelidir. Diğer tüm etkenlerde olduğu gibi, kolonizasyon, enfeksiyon gelişiminin öncülüdür. Çalışmalarda, kolonize olan hastaların %18-25'inde enfeksiyon gelişebildiği bildirilmektedir. Birden fazla bölgede kolonizasyonu olan hastalarda enfeksiyon gelişme riski daha fazladır.[22-24] En sık kan dolaşımı enfeksiyonları, miyokardit, cerrahi alan enfeksiyonu, yanık enfeksiyonları, kateter yerinde apse gelişimi, otit, menenjit, kemik eklem enfeksiyonları geliştiği bildirilmiştir.[3,4,10,25,26] Uzun süre hastane yatışı olan, antibiyoterapiyle yanıt alınamayan diyabetik ayak enfeksiyonlarında da etken akla gelmelidir.[10] *C. auris* enfeksiyonları ile ilişkili risk faktörleri Tablo 1'de özetlenmiştir.[24] Etkenin *C. auris* olduğu enfeksiyonlarda mortalite oranının %39 civarında olduğu ancak kandidemi gelişmesi durumunda riskin artarak %45 olduğu bildirilmektedir.[27]

Etkenin Tanımlanması ve Değerlendirilmesi

Aksilla, kasık, dış kulak gibi florası olan vücut bölgelerinden süruntü şeklinde alınan numunelerde etken saptanması, öncelikle kolonizasyon lehine düşünülmelidir. Enfeksiyonu destekleyen bulgular olduğunda alınan ve steril vücut bölgeleri ya da steril vücut sıvılarından aspirasyon yöntemiyle alınan kültürlerde etken izole edildiğinde enfeksiyon olarak değerlendirilmelidir.[18,24]

Tablo 1. Candidozyma auris enfeksiyonları ile ilişkili risk faktörleri[24]

Risk faktörleri	Açıklama
Santral venöz kateter varlığı	Vasküler giriş yolu enfeksiyon riskini artırır
Kalıcı idrar kateteri	Kolonizasyon riski ve idrar yolu enfeksiyonu gelişim riski artar
İmmünsüpresif durum	HIV, hematolojik malignite, solid tümör, transplantasyon, nötropeni, kemoterapi, kortikosteroid kullanımı
Diyabetes mellitus	Bağışıklığın baskılanmasıyla enfeksiyon gelişim riski artar
Kronik böbrek hastalığı	Enfeksiyona yatkınlık artışı
Geniş spektrumlu antibiyotik/antifungal maruziyeti	Dirençli organizmaların seçilimini artırır
Eşlik eden bakteriyemi veya kandidüri	İkincil enfeksiyonlar ve yayılım riski artar
Parenteral beslenme	Kontaminasyon gelişimi ve enfeksiyon gelişim riski artar
Kan transfüzyonu	Bağışıklık sistemi etkilenimi ve enfeksiyon riski artar
Hemodiyaliz	Kateter ve cihaz kaynaklı enfeksiyon riski artar
Otuz gün içinde ameliyat	Cerrahi girişimler sonrası enfeksiyon riski artar
Yoğun bakım ünitesine kabul	Yoğun invaziv girişimler ve dirençli mikroorganizma maruziyeti artar

HIV: İnsan bağışıklık yetmezliği virüsü

Ticari olarak renklendirici maddeler eklenmiş katı besiyerlerinde, *C. auris* kolonileri genellikle beyaz veya pembe görünür, ancak bazı koloniler kırmızı veya mor görünebilir. Etkenin *in vitro* koşullarda üretilebilmesi için ideal inkübasyon sıcaklığı 40-42 °C'dir. Otomatize kan kültürü sistemlerinde Candida türlerinin büyümesi genellikle 1-3 gün sürer ve agar besiyerine ekim yapıldıktan sonra tanımlama için 1-2 gün daha gerekir. Tanımlama için geleneksel fenotipik yöntemler veya yarı otomatize sistemler kullanıldığında, tür düzeyinde tanımlamalarda hatalar olabilir. Bu yöntemler arasında VITEK 2 YST, API 20C, BD Phoenix maya tanımlama sistemi ve MicroScan bulunur. Bu yöntemlerle, *C. auris*, yanlış olarak, C. catenulata, *C. haemulonii*, *C. duobushaemulonii* ve *C. sake* olarak tanımlanabilir. Tablo 2'de *C. auris*'in yanlış tanımlanabileceği organizmalar ve tanımlanma yöntemi verilmiştir.[28] En doğru tanımlama, veri tabanına *C. auris* eklenmiş olan MALDI-TOF MS yöntemiyle yapılır. PZR gibi moleküler yöntemler de tanımlama için önerilmektedir.[28-30]

Tarama ve Sürveyans

Hastalık kontrol ve önleme merkezi, ilk olgu tanımlandıktan sonra kolonizasyon için riskli hastaların sağlık kuruluşlarında taranmasını önermektedir. Tarama için iki taraflı aksilla ve kasık bölgelerinden alınacak deri süründülerinden ekim yapılmalıdır. Alınan örneklerden etkenin tespiti için PZR ya da MALDI-TOF MS yöntemlerinin kullanılması önerilmektedir. *C. auris* ile kolonize veya enfekte hastaya bakım veren sağlık personeli tarafından bakım alan, aynı odayı veya aynı üniteyi paylaşan diğer hastalar, temizlik ve dezenfeksiyon konusunda endişe duyulan tıbbi cihaz ile teması olanlar kolonizasyon yönünden taranmalıdır. Etkenin endemik olduğu sağlık kurumlarından gelen hastaların da taranması önerilmektedir. Kolonizasyon saptanan hastaların tedavi alması önerilmemektedir. Bu hastalar, taburcu olana kadar temas izolasyon önlemleri altında izlenmelidir. Yapılan çalışmalarda, kolonizasyonun aylarca ve yıllarca devam edebildiği bildirilmekte bu nedenle tekrarlayan kültürlerle kolonizasyon durumunun takibi veya izolasyonun sonlandırılması konularında net bir yaklaşım ve öneri yoktur.[20,21,31]

Tablo 2. Candida auris tanımlanma yöntemi ve yanlış tanımlandığı organizmalar[30]

Tanımlanma yöntemi	Yanlış tanımlandığı organizma
Vitek 2 YST	Candida haemulonii Candida duobushaemulonii
API 20C	Rhodotorula glutinis Candida sake
API ID 32C	Candida intermedia Candida sake Saccharomyces kluyveri
BD Phoenix yeast identification system	Candida haemulonii Candida catenulata
MicroScan	Candida famata Candida guilliermondii Candida lusitaniae Candida parapsilosis
RapID™ YEAST PLUS	Candida parapsilosis

Tedavi ve Antifungal Direnç

C. auris'i diğer Candida türlerinden ayıran ve klinisyenleri zorlayan en önemli özelliği, antifungal ilaçlara olan direncidir. En yüksek direnç sırasıyla flukonazole (%87-100) ve amfoterisin B'ye (%8-35) karşı bildirilmiştir.[27,32,33] Direncin en az olduğu ilaç grubu ekinokandinlerdir. Bu nedenle gelişen enfeksiyonun tedavisinde ilk seçenek, ekinokandinlerdir (anidulafungin, kaspofungin, mikafungin). Etken saptandığında antifungal duyarlılık testi yapılması önerilmektedir. Hastanın önceki antifungal maruziyet öyküsü etkenin direnç durumunu etkileyebileceği gibi, tedavinin seyri sırasında da direnç gelişimi gözlenmiştir. Bu sebeple tedavi öncesi antifungal duyarlılık testleri görülmeli, tedavi yanıtının alınamadığı durumlarda da tekrarlanmalıdır. Literatürde, giderek artan oranlarda ekinokandin direnci bildirilmektedir. Yapılan çalışmalarda %0-8 oranında ekinokandin direnci bildirilmiştir. Hindistan ve Kuveyt'te ekinokandine dirençli olgu sayıları giderek artmaktadır.[34,35] Ekinokandin direnci saptandığında veya tedavi altında beşinci günden itibaren ekinokandin tedavisine yanıt alınamadığı durumlarda lipozomal amfoterisin B tedavisi verilmesi önerilmektedir.[27,32,33] Giderek artan ekinokandin dirençleri yeni tedavi rejimleri için ileri çalışmaların gerekliliğini arttırmış olup araştırılan moleküllerden biri de farnesoldür. Yapılan çalışmalarda farnesolün *C. auris*'in efluks pompaları ve biyofilm yapısı üzerine etkinliği gösterilmiştir.[36] Farnesol monoterapisinin veya flukonazol ile kombinasyonunun gelecek için umut verici tedavi stratejileri olabileceği yönünde veriler mevcuttur.[36,37] İlaç direnci devam ettiği sürece yeni antifungal stratejileri konusunda çalışmalara ihtiyaç var gibi görünmektedir.

Korunma ve Hijyen

Çoklu ilaca dirençli enfeksiyonlarda olduğu gibi *C. auris* enfeksiyonları için de yayılımı önlemek için alınacak önlemler, standart temas izolasyon uygulamalarının etkin şekilde yapılması, el hijyeni kurallarına uyum, çevresel dezenfeksiyonun uygun ve etkili ürünler ile sağlanması, hasta transferi sırasında hastanın durumunun bildirilmesidir. Enfekte veya kolonize hastaların mümkün olduğunca tek kişilik odalarda izlenmesi, hastaya, hasta yakınlarına ve sağlık personellerine el hijyeni ve izolasyon önlemleri ile ilgili eğitim verilmesi gerekmektedir. Kullanılan tıbbi cihazların hastaya özgü ya da tek

kullanımlık olması gerekmektedir. Çevre temizliği için *C. auris*'e karşı etkili ve Çevre Koruma Ajansı'na (EPA) kayıtlı bir dezenfektan kullanılması önerilmektedir. Listede yer alan ürünlere ulaşılamazsa, C. difficile sporlarına karşı etkili EPA'ya kayıtlı bir dezenfektan kullanabileceğini bildirilmektedir. Alkollü veya alkolsüz kuaterner amonyum bileşikleri içeren ürünler yerine perasetik asit bazlı dezenfektanlar kullanılmalıdır. Ultrason problarının hidrojen peroksit içeren mendillerle temizlenmesi önerilmektedir. Hastanedeki diğer tıbbi cihazların veya yüzeylerin dezenfeksiyonu için alkol bazlı dezenfektanların kullanımına devam edilebilir. Bu ürünlere ek olarak ultraviyole ışınlama ve buharlaştırılmış hidrojen peroksit de çevre temizliği için kullanılabilir. Ancak halihazırda bu uygulamaların süresi ve uygulama sıklığı ile ilgili net veriler yoktur. Temizlik işlemlerinde kontaminasyonun sık olduğu hasta yatak başları, yatak yanları, monitör yüzeyleri ve kablolarının ayrıntılı olarak temizlenmesine özen gösterilmelidir.[38,39]

Dipnotlar

Yazarlık Katkıları

Cerrahi ve Medikal Uygulama: B.A.A., N.S., Konsept: B.A.A., N.S., Dizayn: B.A.A., N.S., Veri Toplama veya İşleme: B.A.A., Analiz veya Yorumlama: B.A.A., Literatür Arama: B.A.A., N.S., Yazan: B.A.A.

Çıkar Çatışması: Yazarlar bu makale ile ilgili olarak herhangi bir çıkar çatışması bildirmemiştir.

Finansal Destek: Çalışmamız için hiçbir kurum ya da kişiden finansal destek alınmamıştır.

KAYNAKLAR

1. Pappas PG, Kauffman CA, Andes DR, et al. Clinical Practice Guideline for the Management of Candidiasis: 2016 Update by the Infectious Diseases Society of America. Clin Infect Dis. 2016;62:e1-50.
2. Colombo AL, de Almeida Júnior JN, Slavin MA, Chen SC, Sorrell TC. Candida and invasive mould diseases in non-neutropenic critically ill patients and patients with haematological cancer. Lancet Infect Dis. 2017;17:e344-56.
3. Satoh K, Makimura K, Hasumi Y, Nishiyama Y, Uchida K, Yamaguchi H. Candida auris sp. nov., a novel ascomycetous yeast isolated from the external ear canal of an inpatient in a Japanese hospital. Microbiol Immunol. 2009;53:41-4.
4. Lee WG, Shin JH, Uh Y, et al. First three reported cases of nosocomial fungemia caused by Candida auris. J Clin Microbiol. 2011;49:3139-42.
5. Calvo B, Melo AS, Perozo-Mena A, et al. First report of Candida auris in America: clinical and microbiological aspects of 18 episodes of candidemia. J Infect. 2016;73:369-74.
6. van Schalkwyk E, Mpembe RS, Thomas J, et al. Epidemiologic shift in Candidemia driven by Candida auris, South Africa, 2016-2017[1]. Emerg Infect Dis. 2019;25:1698-1707.
7. Kurt AF, Kuskucu MA, Balkan II, et al. Candida auris fungemia and a local spread taken under control with infection control measures: first report from Turkey. Indian J Med Microbiol. 2021;39:228-30.
8. Bölükbaşı Y, Erköse Genç G, Orhun G, et al. Türkiye'de ilk COVID-19 pozitif Candida auris fungemi olgusu [First case of COVID-19 positive Candida auris fungemia in Turkey]. Mikrobiyol Bul. 2021;55:648-55.
9. Kömeç S, Karabıçak N, Ceylan AN, Gülmez A, Özalp O. Türkiye İstanbul'dan bildirilen üç Candida auris olgusu [Three Candida auris case reports from Istanbul, Turkey]. Mikrobiyol Bul. 2021;55:452-60. Turkish.
10. Kulaklı K, Arslan N, Gürsan O, Özkütük A. İzmir'den ilk Candida auris izolasyonu: amputasyon ile sonuçlanan polimikrobiyal diyabetik ayak enfeksiyonu. Turk Mikrobiyol Cemiy Derg. 2023;53:47-54.
11. Muñoz JF, Gade L, Chow NA, et al. Genomic insights into multidrug-resistance, mating and virulence in Candida auris and related emerging species. Nat Commun. 2018;9:5346
12. Chowdhary A, Sharma C, Duggal S, et al. New clonal strain of Candida auris, Delhi, India. Emerg Infect Dis. 2013;19:1670-3.
13. Hernando-Ortiz A, Mateo E, Perez-Rodriguez A, de Groot PWJ, Quindós G, Eraso E. Virulence of Candida auris from different clinical origins in Caenorhabditis elegans and Galleria mellonella host models. Virulence. 2021;12:1063-75.
14. Centers for Disease Control and Prevention. Tracking C. auris. Son Erişim Tarihi: 25.07.2025. Erişim Linki: https://www.cdc.gov/candida-auris/tracking-c-auris/index.html
15. Burki T. WHO publish fungal priority pathogens list. Lancet Microbe. 2023;4:e74.
16. Eix EF, Nett JE. Candida auris: epidemiology and antifungal strategy. Annu Rev Med. 2025;76:57-67.
17. T.C. Sağlık Bakanlığı Halk Sağlığı Genel Müdürlüğü. Sağlık hizmetlerinde Candida auris. 2025. Available from: https://hsgm.saglik.gov.tr/depo/birimler/bulasici-hastaliklar-ve-erken-uyari-db/Dokumanlar/Afisler-Brosurler/2025/Ek_Saglik_Hizmetlerinde_C._auris.pdf
18. Chowdhary A, Sharma C, Meis JF. Candida auris: a rapidly emerging cause of hospital-acquired multidrug-resistant fungal infections globally. PLoS Pathog. 2017;13:e1006290.
19. Casadevall A, Kontoyiannis DP, Robert V. On the emergence of Candida auris: climate change, azoles, swamps, and birds. mBio. 2019;10:e01397-19.
20. Akinbobola A, Kean R, Quilliam RS. Plastic pollution as a novel reservoir for the environmental survival of the drug resistant fungal pathogen Candida auris. Mar Pollut Bull. 2024;198:115841.
21. Harris A, Sansom SE, Hayden MK, et al. Risk factors for contamination of healthcare personnel gloves and gown with Candida auris (C. auris). OFID. 2025;12(Suppl 1):ofae631.1447.
22. Briano F, Magnasco L, Sepulcri C, et al. Candida auris Candidemia in critically Ill, colonized patients: cumulative incidence and risk factors. Infect Dis Ther. 2022;11:1149-60.
23. Schelenz S, Hagen F, Rhodes JL, et al. First hospital outbreak of the globally emerging Candida auris in a European hospital. Antimicrob Resist Infect Control. 2016;5:35.
24. Cortegiani A, Misseri G, Fasciana T, Giammanco A, Giarratano A, Chowdhary A. Epidemiology, clinical characteristics, resistance, and treatment of infections by Candida auris. J Intensive Care. 2018;6:69.

25. Sarma S, Upadhyay S. Current perspective on emergence, diagnosis and drug resistance in Candida auris. Infect Drug Resist. 2017;10:155-65.
26. Choi HI, An J, Hwang JJ, Moon SY, Son JS. Otomastoiditis caused by Candida auris: case report and literature review. Mycoses. 2017;60:488-92.
27. Chen J, Tian S, Han X, et al. Is the superbug fungus really so scary? A systematic review and meta-analysis of global epidemiology and mortality of Candida auris. BMC Infect Dis. 2020;20:827.
28. Zhang SX, Carroll KC, Lewis S, et al. Multicenter evaluation of a PCR-based digital microfluidics and electrochemical detection system for the rapid identification of 15 fungal pathogens directly from positive blood cultures. J Clin Microbiol. 2020;58:e02096-19.
29. Centers for Disease Control and Prevention. Identification of C. auris. Son Erişim Tarihi: 25.07.2025. Erişim Linki:https://www.cdc.gov/candida-auris/hcp/laboratories/identification-of-c-auris.html
30. Centers for Disease Control and Prevention. Guidance for detection of C. auris colonization. Son Erişim Tarihi: 25.07.2025. Erişim Linki: https://www.cdc.gov/candida-auris/hcp/laboratories/detection-colonization.html
31. Kim HY PhD, Nguyen TA MSc, et al. Candida auris-a systematic review to inform the world health organization fungal priority pathogens list. Med Mycol. 2024;62:myae042.
32. Centers for Disease Control and Prevention. Clinical treatment of C. auris infections. Available from: https://www.cdc.gov/candida-auris/hcp/clinical-care/index.html
33. Ahmad S, Khan Z, Al-Sweih N, Alfouzan W, Joseph L. Candida auris in various hospitals across Kuwait and their susceptibility and molecular basis of resistance to antifungal drugs. Mycoses. 2020;63:104-12.
34. Kordalewska M, Lee A, Park S, et al. Understanding echinocandin resistance in the emerging pathogen Candida auris. Antimicrob Agents Chemother. 2018;62:e00238-18.
35. Srivastava V, Ahmad A. Abrogation of pathogenic attributes in drug resistant Candida auris strains by farnesol. PLoS One. 2020;15:e0233102.
36. Dekkerová J, Černáková L, Kendra S, et al. Farnesol boosts the antifungal effect of fluconazole and modulates resistance in Candida auris through Regulation of the CDR1 and ERG11 genes. J Fungi (Basel). 2022;8:783.
37. Černáková L, Roudbary M, Brás S, Tafaj S, Rodrigues CF. Candida auris: a quick review on identification, current treatments, and challenges. Int J Mol Sci. 2021;22:4470.
38. Aldejohann AM, Wiese-Posselt M, Gastmeier P, Kurzai O. Expert recommendations for prevention and management of Candida auris transmission. Mycoses. 2022;65:590-8.
39. Centers for Disease Control and Prevention. Infection Control Guidance: Candida auris. Son Erişim Tarihi: 25.07.2025. Erişim Linki: https://www.cdc.gov/candida-auris/hcp/infection-control/index.html

DOI: 10.4274/forbes.galenos.2025.23590
Forbes J Med 2025;6(2):103-12

Derleme / Review

TUKMOS'un Güncel Uygulamaları Çerçevesinde Türkiye'de Tıpta Uzmanlık Eğitiminin Mevcut Durumu

The Current State of Medical Specialty Training in Türkiye within the Framework of Recent TUKMOS Practices

Murat DEMİRCİ[1,2], Mehmet KOCA[3], Özlem COŞKUN[1]

[1]Gazi Üniversitesi Tıp Fakültesi, Tıp Eğitimi ve Bilişimi Anabilim Dalı, Ankara, Türkiye
[2]Marmara Üniversitesi, Pendik Eğitim ve Araştırma Hastanesi, Kardiyoloji Kliniği, İstanbul, Türkiye
[3]T.C. Sağlık Bakanlığı, Ankara, Türkiye

Atıf: Demirci M, Koca M, Coşkun Ö. The current state of medical specialty training in Türkiye within the framework of recent TUKMOS practices. Forbes J Med. 2025;6(2):103-12

ÖZ

Tıpta Uzmanlık Kurulu Müfredat Oluşturma Sistemi (TUKMOS), Türkiye'de uzmanlık eğitiminin kalitesini artırmak ve ulusal ölçekte standardizasyonu sağlamak amacıyla oluşturulmuştur. 2011 yılından itibaren zorunlu olan bu sistem, tıp alanında 88 uzmanlık dalı için çekirdek müfredat geliştirmiştir; bunların 43'ü ana dal, 45'i ise yan dal uzmanlık branşlarını kapsamaktadır. Bu derlemede, TUKMOS'un tarihsel gelişimi, müfredat yapısı, uygulamadaki karşılıkları ve eğitime olan etkileri değerlendirilmiştir. Çalışma, ilgili literatür ve güncel resmi belgeler esas alınarak hazırlanan bir geleneksel derlemedir. Klinik ve girişimsel yetkinliklere yönelik öğrenim hedefleri tanımlanmıştır. Ancak sahadaki bulgular, bu müfredatların tüm kurumlarda etkin şekilde uygulanamadığını göstermektedir. Eğitimcilerin sürece yeterince katılmaması, ölçme ve değerlendirme araçlarının sınırlı kullanılması, dış rotasyonlardaki işleyiş sorunları ve bilimsel araştırma eğitimine yeterli önem verilmemesi, karşılaşılan başlıca sorunlardandır. TUKMOS'un zorunlu hale gelmesi, sadece müfredatların hazırlanmış olmasını değil, aynı zamanda bu müfredatların sahada etkili ve sürdürülebilir bir şekilde uygulanmasını da zorunlu kılmaktadır. Bu nedenle, eğiticiler ve uzmanlık öğrencileri için TUKMOS farkındalığının artırılması, denetim ve izleme süreçlerinin güçlendirilmesi, uzmanlık eğitiminin kalitesini artırmada önemlidir.

Anahtar Kelimeler: TUKMOS, tıpta uzmanlık eğitimi, çekirdek müfredat, klinik yetkinlik, ölçme ve değerlendirme

ABSTRACT

The Medical Specialization Board Curriculum Development System (TUKMOS) was established in Turkey to enhance the quality of specialty training in medicine and to ensure standardization on a national scale. This system, which has been mandatory since 2011, has developed core curricula for 88 medical specialties, of which 43 are primary specialties and 45 are subspecialties. This review evaluates the historical development of TUKMOS, the structure of its curricula, their practical implementation, and their impact on medical education. This is a traditional review based on the relevant literature and current official documents. Learning objectives related to clinical and procedural competencies have been defined. However, field data indicate that these curricula are not effectively implemented across all institutions. The main challenges include insufficient involvement of educators, limited use of assessment and evaluation tools, problems in the functionality of external rotations, and the lack of adequate emphasis on scientific research training. The mandatory nature of TUKMOS requires not only the existence of these curricula but also their effective and sustainable implementation in practice. Therefore, raising awareness of TUKMOS among educators and medical residents, and strengthening supervision and monitoring processes, are important steps for improving the quality of specialty training.

Keywords: TUKMOS, medical specialty training, core curriculum, clinical competency, assessment and evaluation

Geliş/Received: 13.07.2025
Kabul/Accepted: 23.07.2025
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Sorumlu Yazar/ Corresponding Author:
Dr. Murat DEMİRCİ,
Gazi Üniversitesi Tıp Fakültesi, Tıp Eğitimi ve Bilişimi Anabilim Dalı, Ankara; Marmara Üniversitesi, Pendik Eğitim ve Araştırma Hastanesi, Kardiyoloji Kliniği, İstanbul, Türkiye
✉ drmuratdemirci@gmail.com
ORCID: 0000-0002-8835-9557

GİRİŞ

Tıpta Uzmanlık Kurulu Müfredat Oluşturma ve Standart Belirleme Sistemi (TUKMOS), Türkiye'de tıpta uzmanlık eğitiminin standardizasyonunu sağlamak amacıyla kurulmuş bir sistemdir.[1] TUKMOS'un temel hedefi, uzmanlık eğitimi alanında ulusal düzeyde standartları belirleyerek eğitim kalitesini artırmak ve farklı kurumlar arasında öğrenim sürecinde bütünlüğü sağlamaktır.

TUKMOS'un etkin biçimde hayata geçirilmesi, eğitim seviyesini artırmakla birlikte, genel sağlık hizmetlerinin verimliliğini de destekleyebilir. Özellikle iyi tanımlanmış öğrenme hedeflerine dayalı çıktı temelli uzmanlık eğitimi; koruyucu hekimliğin güçlendirilmesi, hastalıkların erken tanı ve etkili yönetimi ile hastane yatışlarının azaltılması, mortalite ve morbidite oranlarının düşürülmesi gibi temel sağlık çıktılarının iyileştirilmesine doğrudan katkı sağlayabilir.[2,3] Bu yaklaşım, uluslararası düzeyde de kabul görmüş Kanadalı Uzmanlar için Tıp Eğitimi Yönergeleri ve Tıpta Uzmanlık Eğitimi Akreditasyon Konseyi gibi eğitim modelleriyle uyumludur ve mezunların yetkinlik odaklı yetiştirilmesini öngörmektedir.[4,5] Bu nedenle, TUKMOS müfredatlarının güncel bilimsel bilgilerle uyumlu olması ve pratikte etkili bir şekilde uygulanabilmesi hem bireylerin hem de toplumun sağlığını korumada önemli bir rol oynamaktadır.

Türkiye'de uzmanlık eğitiminde standartlaştırma çalışmaları uzun süredir devam etmekle birlikte, resmî anlamda TUKMOS'un yasal temeli, 2009 yılında yayımlanan "Tıpta ve Diş Hekimliğinde Uzmanlık Eğitimi Yönetmeliği" ile atılmıştır.[6] Bu yönetmelikle birlikte uzmanlık dalları için zorunlu hale getirilen çekirdek müfredatlar oluşturulmaya başlanmıştır. Bu geleneksel derlemede, TUKMOS'un tarihsel gelişimi, güncel uygulamaları ve tıpta uzmanlık eğitimine etkilerinin genel hatlarıyla değerlendirilmesi amaçlanmıştır.

TUKMOS Müfredatlarının Tarihsel Gelişimi ve Revizyon Süreçleri

TUKMOS'un ilk oluşturduğu müfredat örnekleri (v.0.5-v.1.0), zorunluluğu bulunmayan, tavsiye amaçlı belgeler olarak kurumların tercihine göre uygulanmaktaydı.[7] Bu erken dönem taslaklarda yalnızca konu başlıkları yer almakta, öğrenme amaçları ile klinik ve girişimsel becerilere dair detaylara yer verilmemişti. Ancak 2013 yılında yürürlüğe giren ve zorunlu uygulama sürecini başlatan Versiyon 2.0 ile birlikte, uzmanlık dallarının çekirdek müfredatlarının standardizasyonu yönünde önemli bir adım atıldı. Bu versiyonla birlikte her branş için klinik ve girişimsel yetkinlikler, öğrenim hedefleri, rotasyon planları, öğrenme-öğretme yöntemleri, ölçme-değerlendirme araçları ve eğitici yeterlilikleri ayrıntılı olarak tanımlanmıştır.[7] Bu versiyonun uzmanlık eğitimine temel katkısı, standardizasyonu sağlamaya yönelik somut yetkinlik tanımlamaları ve ölçme değerlendirme araçlarının oluşturulmasıdır. Versiyon 2.0'ın getirdiği bu ana iskelet, sonraki tüm güncellemelerde korunmuş ve müfredatların temel yapısını oluşturmaya devam etmiştir. Versiyon 2.1 (2016) ve 2.3 (2017) ile özellikle minimum olgu sayısı ve yetkinlik düzeylerinin tanımı güçlendirilmiş, bazı alanlarda olgu odaklı yaklaşım öne çıkarılmıştır.[7] Bazı branşlarda Versiyon 2.3 kullanılmaya devam etmektedir. Halen pek çok uzmanlık dalında yürürlükte olan 2.4 sürümü 2019 yılında uygulanmaya başlamış ölçme değerlendirme yöntemlerine daha fazla yer verilmiştir.[7] En son olarak, Versiyon 2.5, 2024 yılı itibarıyla taslak olarak hazırlanmış ve ilgili branş derneklerinin görüşleri doğrultusunda uluslararası rehberler ile uyumlu hale gelecek şekilde kapsamlı bir revizyon sürecine girilmiştir. Bu süreç sonucunda onaylanan yeni müfredatlar, 2025 yılı haziran ayından itibaren resmiyet kazanarak bazı bölümler için kullanıma sunulmuştur.[8]

Müfredat Hazırlama Süreci

Uzmanlık eğitiminin planlanması, sürekli geliştirilmesi ve kalite güvencesinin sağlanması görevini üstlenen Tıpta Uzmanlık Kurulu (TUK), Sağlık Bakanlığı bünyesinde faaliyet göstermektedir.[9] TUKMOS müfredatları, TUK'un koordinasyonunda, her branşa özgü oluşturulan akademik çalışma ekipleri tarafından hazırlanmaktadır. Bu komisyonlar; üniversitelerde ve araştırma hastanelerinde görev yapan öğretim üyeleri, alanında yetkin uzman hekimler ve ilgili uzmanlık derneklerinin temsilcilerinden oluşmaktadır.[9]

Müfredat geliştirme sürecinde temel ilke; uluslararası standartlar, bilimsel gelişmeler ile toplumsal sağlık gereksinimlerinin göz önünde bulundurulduğu çok boyutlu bir değerlendirmedir.[10] Süreç, çok paydaşlı ve kademeli bir yöntemle ilerler. Öncelikle her uzmanlık alanında; öğrenci ve eğitici beklentileri, alanın bilimsel ve teknolojik gelişimi, klinik uygulama gereklilikleri ve global uzmanlık eğitim standartları ışığında bir eğitim planı önerisi geliştirilir.

Hazırlanan taslak, TUK tarafından detaylı biçimde değerlendirilir; içerik ve yapı bakımından uygun bulunması durumunda onay sürecine alınır, gerek görülürse ilgili komisyona revizyon önerileriyle geri gönderilir. Bu aşamada sadece akademik paydaşların değil, aynı zamanda sağlık otoritelerinin ve meslek örgütlerinin görüşleri alınarak toplu bir uzlaşı sağlanması amaçlanır. Son halini alan müfredatlar, TUK'un onayıyla birlikte resmen yürürlüğe girer.[9]

Uygulama aşamasında eğitim kurumları ile sorumlu eğiticiler ana aktör olarak görev yapar; kurumlar, TUKMOS tarafından tanımlanan öğrenme hedefleri ve yetkinlik

alanlarını eksiksiz biçimde gerçekleştirmekle yükümlüdür.[9] Müfredatlar doğrultusunda klinik rotasyonlar, teorik eğitim oturumları, uygulamalı beceri eğitimleri ve diğer öğretim etkinlikleri planlı ve düzenli şekilde uygulanmalıdır.

TUKMOS'ta Yer Alan Uzmanlık Dalları ve Eğitim Süreleri

TUKMOS'un güncel müfredat düzenlemeleri doğrultusunda, Türkiye'de toplam 97 uzmanlık dalı için çekirdek müfredat oluşturulmuştur.[8] Bu dallardan 88'si tıp, 9'u ise diş hekimliği alanına aittir. Tıp alanındaki müfredatların 43'ü ana dal, 45'i ise yan dal uzmanlıklarını kapsamaktadır. Ana dal uzmanlıklarında eğitim süresi 3 ile 5 yıl arasında olup, yan dallarda bu süre 2 ile 3 yıl arasında değişmektedir (Tablo 1 ve 2).[11]

Tıpta uzmanlık eğitimi süreleri 2011 yılında 1219 sayılı kanuna ekli çizelge yoluyla düzenlenmiştir.[12] Ancak, TUK kararıyla 2014'te Anesteziyoloji ve Reanimasyon (karar

Tablo 1. Ana dal uzmanlıkları, eğitim ve rotasyon süreleri

Branş	Son versiyon	Eğitim süresi (yıl)	Rotasyon süresi (ay)
Acil Tıp	2.4	4	9
Adli Tıp	2.4	4	6
Ağız, Yüz ve Çene Cerrahisi	2.0	5	15
Aile Hekimliği	2.4	3	18
Anatomi	2.3	3	6
Anesteziyoloji ve Reanimasyon	2.4	5	2
Askeri Sağlık Hizmetleri	2.1	3	6
Beyin ve Sinir Cerrahisi	2.3	5	4
Çocuk Cerrahisi	2.5	5	6
Çocuk Sağlığı ve Hastalıkları	2.4	4	4
Çocuk ve Ergen Ruh Sağlığı ve Hastalıkları	2.4	4	12
Deri ve Zührevi Hastalıkları	2.4	4	7
Enfeksiyon Hastalıkları ve Klinik Mikrobiyoloji	2.5	5	12
Fiziksel Tıp ve Rehabilitasyon	2.3	4	9
Fizyoloji	2.3	4	4
Genel Cerrahi	2.5	5	13
Göğüs Cerrahisi	2.4	5	12
Göğüs Hastalıkları	2.5	4	14
Göz Hastalıkları	2.3	4	1
Halk Sağlığı	2.4	4	6
Hava ve Uzay Hekimliği	2.4	3	6
Histoloji ve embriyoloji	2.4	4	3
İç Hastalıkları	2.5	4	8
Kadın Hastalıkları ve Doğum	2.3	4	6
Kalp ve Damar Cerrahisi	2.4	5	11
Kardiyoloji	2.4	5	15
Kulak Burun Boğaz Hastalıkları	2.4	5	4
Nöroloji	2.5	5	12
Nükleer Tıp	2.3	4	4
Ortopedi ve Travmatoloji	2.4	5	9
Plastik, Rekonstrüktif ve Estetik Cerrahi	2.4	5	7
Radyasyon Onkolojisi	2.4	5	14
Radyoloji	2.4	5	2

Tablo 1. Devamı

Branş	Son Versiyon	Eğitim süresi (yıl)	Rotasyon süresi (ay)
Ruh Sağlığı ve Hastalıkları	2.4	4	7
Spor Hekimliği	2.3	4	17
Sualtı Hekimliği ve Hiperbarik Tıp	2.3	3	5
Tıbbi Biyokimya	2.3	4	7
Tıbbi Ekoloji ve Hidroklimatoloji	2.3	3	12
Tıbbi Farmakoloji	2.4	4	5
Tıbbi Genetik	2.4	4	6
Tıbbi Mikrobiyoloji	2.4	4	6
Tıbbi Patoloji	2.4	4	0
Üroloji	2.4	5	6

Eğitim süresi "yıl", rotasyon süresi ise "ay" olarak belirtilmiştir. Rotasyon süresi "0" olarak verilen branşlarda, ilgili uzmanlık alanında zorunlu dış rotasyon bulunmamaktadır

Tablo 2. Yan dal uzmanlıkları, eğitim ve rotasyon süreleri

Branş	Son Versiyon	Eğitim süresi (yıl)	Rotasyon süresi (ay)
Algoloji	2.4	2	7
Askeri Psikiyatri	2.1	2	4
Cerrahi Onkoloji	2.3	2	4
Çevre Sağlığı	2.4	2	0
Çocuk Acil	2.3	3	1
Çocuk Endokrinolojisi	2.4	3	3
Çocuk Enfeksiyon Hastalıkları	2.3	3	2
Çocuk Gastroenterolojisi	2.3	3	3
Çocuk Genetik Hastalıkları	2.3	3	3
Çocuk Göğüs Hastalıkları	2.3	3	2
Çocuk Hematolojisi ve Onkolojisi	2.3	3	0
Çocuk İmmünolojisi ve Alerji Hastalıkları	2.4	3	1
Çocuk Kalp ve Damar Cerrahisi	2.3	2	4
Çocuk Kardiyolojisi	2.4	3	3
Çocuk Metabolizma Hastalıkları	2.3	3	2
Çocuk Nefrolojisi	2.3	3	0
Çocuk Nörolojisi	2.4	3	10
Çocuk Radyolojisi	2.3	2	0
Çocuk Romatolojisi	2.3	3	4
Çocuk Ürolojisi	2.3	3	3
Çocuk Yoğun Bakımı	2.3	3	1
El Cerrahisi	2.3	2	4
Endokrinoloji ve Metabolizma Hastalıkları	2.4	3	2
Epidemiyoloji	2.4	2	0
Gastroenteroloji	2.3	3	3
Gastroenteroloji Cerrahisi	2.3	2	0
Gelişimsel Pediatri	2.3	3	8

Tablo 2. Devamı

Branş	Son Versiyon	Eğitim süresi (yıl)	Rotasyon süresi (ay)
Geriatri	2.4	3	4
Harp Cerrahisi	2.1	2	9
Hematoloji	2.3	3	0
İmmünoloji ve Alerji Hastalıkları	2.3	3	0
İş ve Meslek Hastalıkları	2.3	3	17
Jinekolojik Onkoloji Cerrahisi	2.3	3	10
Klinik Nörofizyoloji	2.3	2	0
Nefroloji	2.3	3	1
Neonatoloji	2.3	3	3
Perinatoloji	2.4	3	3
Romatoloji	2.4	3	0
Sitopatoloji	2.3	2	0
Temel İmmünoloji	2.3	2	3
Tıbbi Mikoloji	2.3	2	4
Tıbbi Onkoloji	2.4	3	1
Tıbbi Parazitoloji	2.3	2	1
Tıbbi Viroloji	2.4	2	3
Yoğun Bakım	2.3	3	9

Eğitim süresi "yıl", rotasyon süresi ise "ay" olarak belirtilmiştir. Rotasyon süresi "0" olarak verilen branşlarda, ilgili uzmanlık alanında zorunlu dış rotasyon bulunmamaktadır

numarası: 488), 2016'da Kulak Burun Boğaz Hastalıkları (karar numarası: 651); 2017'de Radyoloji (karar numarası: 729), Kardiyoloji (karar numarası: 731), ve Radyasyon Onkolojisi (karar numarası: 916), 2018'de Nöroloji (karar numarası: 982) 4 yıldan 5 yıla çıkarılmıştır.[13] Ayrıca, 2022 yılında Fizyoloji (karar numarası: 2091) ile Histoloji ve Embriyoloji (karar numarası: 2105) alanlarında eğitim süresi 3 yıldan 4 yıla yükseltilmiştir.[13]

TUKMOS Kapsamında Hekim Yetkinlik Alanları ve Uygulama Düzeyleri

TUKMOS, uzmanlık eğitimi kapsamında hekimlerin sahip olması gereken yedi temel yetkinliği tanımlamaktadır.[1] Bu yetkinliklerin ilk altısı mesleki gelişimin farklı yönlerini kapsayan, birbiriyle iç içe geçmiş beceri alanlarından oluşmaktadır. "Yönetici" yetkinliği, hekimlerin sağlık sistemini anlayarak insan ve kaynak yönetimi gibi süreçlerde etkin rol üstlenmelerini hedeflerken; "Ekip Üyesi" rolü, multidisipliner ekiplerle iş birliği içinde çalışarak hasta bakımında uyumlu ve koordineli hareket edebilmeyi gerektirir. "Sağlık Koruyucusu" başlığı altında, toplum sağlığına katkı sunan, koruyucu hekimlik uygulamalarında aktif görev alan bireyler yetiştirilmesi amaçlanmaktadır. Hekimlerin iletişim becerilerini kapsayan "İletişim Kuran" yetkinliği ise hasta ve hasta yakınlarıyla açık, empatik ve etkili bir iletişim kurma sorumluluğunu taşır. "Değer ve Sorumluluk Sahibi" yetkinliği, etik kurallara bağlılık, hasta mahremiyetine saygı ve profesyonel tutum gibi mesleki sorumlulukları içerirken; "Öğrenen ve Öğreten" yetkinliği, hekimlerin yaşam boyu öğrenme sürecine açık olmalarını ve bilgi aktarımı yoluyla eğitici rolünü üstlenmelerini öngörmektedir. Bu yetkinlik alanları birlikte ele alındığında, hekimlerin sadece klinik becerilerde değil, aynı zamanda mesleki kimlik ve sorumluluk bilincinde de gelişim göstermesi amaçlanmaktadır.

TUKMOS'un çekirdek müfredatlarında yer alan klinik ve girişimsel yetkinlikler, uzmanlık öğrencilerinin bilgi ve becerilerinin yapılandırılmış biçimde geliştirilmesini hedeflemektedir. Klinik yetkinlikler için dört ana düzey (B, T, ETT, TT) ve iki tamamlayıcı düzey (A ve K) tanımlanmıştır. "B" düzeyi, hastalığa ön tanı koyma ve doğru yönlendirme yapma becerisini ifade ederken; "T" düzeyi, tanı koyma ve tedaviye yönlendirme yeterliliğini tanımlar. "ETT", ekip çalışması içinde tanı ve tedaviyi yönetme; "TT" ise bağımsız olarak tüm süreci yürütme düzeyidir. Tamamlayıcı düzeylerden "A", acil durumlarda doğru tanı ve müdahaleyi; "K" ise koruyucu sağlık hizmetlerinin tüm basamaklarında gerekli önlemleri alabilmeyi içerir. Öte yandan girişimsel

yetkinlikler için belirlenen dört düzey; bilgi sahibi olma (1. düzey), acil durumlarda denetim altında uygulama (2. düzey), tipik olgularda bağımsız girişim yapma (3. düzey) ve her tür olguda tam yetkinlik gösterme (4. düzey) basamaklarından oluşmaktadır. Bu yapı, öğrencinin gelişim sürecini izlemeyi kolaylaştırmakla kalmaz, aynı zamanda kurumlar arası standartlaşmanın sağlanmasına da önemli katkı sunar (Tablo 3).[1]

Bu sistematik yapı, uzmanlık öğrencilerinin klinik ve girişimsel yetkinliklere ulaşım süreçlerini yapılandırılmış bir şekilde izlemeyi ve değerlendirmeyi mümkün kılarken; aynı zamanda kurumlar arasında uygulama farklılıklarını azaltarak ulusal düzeyde eğitimde standartlaşmayı desteklemektedir. Ancak sahada elde edilen veriler, bu yapıların uygulamada yeterince etkin kullanılmadığını göstermektedir. Örneğin, Kardiyoloji uzmanlık öğrencilerinin çekirdek eğitim programına göre eğitimin ilk yarısında edinmesi beklenen klinik yetkinliklere ilişkin özdeğerlendirme sonuçları, hedeflenen düzeyin altında kalmaktadır.[14] Ayrıca poliklinik verilerine dayalı çalışmalarda, bazı branşlarda ilk beş tanının hasta başvurularının %70-90'ını oluşturduğu, ancak Acil Tıp ve Aile Hekimliği gibi alanlarda bu oranın 30%'lara kadar düştüğü gösterilmiştir.[15] Bu durum, klinik yetkinliklerin, müfredatlarda yaygın ve yüksek karşılanma olasılığı olan tanılar dikkate alınarak yapılandırılması gerektiğine işaret etmektedir.

TUKMOS Kapsamında Öğrenme ve Öğretme Yöntemleri

TUKMOS, öğrenme ve öğretme yöntemlerini üç ana başlık altında sınıflandırmaktadır:

- Yapılandırılmış Eğitim Etkinlikleri
- Uygulamalı Eğitim Etkinlikleri
- Bağımsız ve Keşfederek Öğrenme Etkinlikleri

Yapılandırılmış Eğitim Etkinlikleri

Uzmanlık öğrencilerine teorik bilgilerin sistematik bir biçimde sunulmasını hedefleyen, planlı ve eğitici denetiminde yürütülen ders, seminer, makale saatleri, olgu sunumları ve konsey gibi multidisipliner etkinlikleri kapsamaktadır. Bu tür yapılandırılmış etkinlikler, ağırlıklı olarak bilişsel öğrenme kuramlarına (örneğin; Bloom'un Bilişsel Alan Taksonomisi) dayanmaktadır.[16] Çünkü bilgi aktarımı, kavramsal çerçevenin oluşturulması ve klinik karar verme süreçlerinde mantıksal akıl yürütmenin geliştirilmesi ön plandadır.[16] Bu çerçevede, yapılandırılmış eğitim uygulamalarına yönelik başarılı örneklerden biri, üroloji uzmanlık eğitimi için geliştirilen ve TUKMOS çekirdek müfredatı temel alınarak hazırlanan E-öğrenme Uzmanlık Eğitim Programı'dır. Program, çevrimiçi dersler ve ölçme-değerlendirme araçları aracılığıyla tüm ülke genelinde yüksek memnuniyet ve katılım oranlarıyla uygulanmıştır.[17]

Uygulamalı Eğitim Etkinlikleri

Uzmanlık öğrencilerinin klinik pratikte doğrudan deneyim kazanmasını amaçlayan hasta vizitleri, yatak başı eğitimler, girişimsel uygulamalar ve simülasyon temelli etkinlikleri kapsamaktadır. Bu etkinlikler ise deneyimsel öğrenme kuramı (örneğin; Kolb'un Deneyimsel Öğrenme Döngüsü) ile ilişkilendirilebilir.[18] Uygulamalı eğitimlerde, öğrenenler bilgiyi aktif şekilde deneyimleyerek, uygulama ve gözlem yoluyla bilgi ve becerilerini geliştirmektedir. Bu yaklaşım, yaparak öğrenme ilkesine dayanır ve klinik beceri

Tablo 3. TUKMOS'ta tanımlanan klinik ve girişimsel yetkinliklerin öğrenim hedefi düzeyleri

Klinik yetkinlikler	
B	Hastalığa ön tanı koyma ve gerekli durumda hastaya zarar vermeyecek şekilde ve doğru zamanda, doğru yere sevk edebilecek bilgiye sahip olma düzeyini ifade eder
T	Hastaya tanı koyma ve sonrasında tedavi için yönlendirebilme düzeyini ifade eder
TT	Ekip çalışmasının gerektirdiği durumlar dışında herhangi bir desteğe gereksinim duymadan hastanın tanı ve tedavisinin tüm sürecini yönetebilme düzeyini ifade eder
ETT	Ekip çalışması yaparak hastanın tanı ve tedavisinin tüm sürecini yönetebilme düzeyini ifade eder
A	Hastanın acil durum tanısını koymak ve hastalığa özel acil tedavi girişimini uygulayabilme düzeyini ifade eder
K	Hastanın birincil, ikincil ve üçüncül korunma gereksinimlerini tanımlamayı ve gerekli koruyucu önlemleri alabilme düzeyini ifade eder
Girişimsel yetkinlikler	
1	Girişimin nasıl yapıldığı konusunda bilgi sahibi olma ve bu konuda gerektiğinde açıklama yapabilme düzeyini ifade eder
2	Acil bir durumda, kılavuz veya yönerge eşliğinde veya gözetim ve denetim altında bu girişimi yapabilme düzeyini ifade eder
3	Karmaşık olmayan, sık görülen tipik olgularda girişimi uygulayabilme düzeyini ifade eder.
4	Karmaşık olsun veya olmasın her tür olguda girişimi uygulayabilme düzeyini ifade eder.

kazanımını güçlendirir. Acil Tıp uzmanlık öğrencileriyle yapılan bir çalışmada, TUKMOS müfredatında tanımlı transkütanöz pacemaker uygulamasına yönelik kısa süreli teorik ve pratik eğitimin, bilgi düzeyi ve müdahale becerilerinde anlamlı gelişme sağladığı belirlenmiştir.[19] Benzer şekilde, Aile Hekimliği uzmanlık öğrencilerine temel yaşam desteği, ileri yaşam desteği, doğum yönetimi, hipertansiyona yaklaşım ve febril konvülsiyon yönetimi gibi temel girişimsel beceriler simülasyon merkezinde hem teorik hem de uygulamalı biçimde öğretilmiş; eğitim öncesi ve sonrası yapılan değerlendirmelerde bilgi düzeyinde istatistiksel olarak anlamlı artış saptanmıştır.[20] Bu bulgular, uygulamalı ve simülasyon temelli eğitimlerin, uzmanlık öğrencilerinin bilgi ve klinik becerilerini güçlendirmede etkili bir yöntem olduğunu ortaya koymaktadır.

Bağımsız ve Keşfederek Öğrenme Etkinlikleri

Uzmanlık öğrencisinin kendi öğrenme sorumluluğunu almasını sağlayan; yatan hasta takibi, ayaktan hasta takibi, akran öğrenmesi, literatür okuma, araştırma yapma gibi öğrenen merkezli yaklaşımları kapsamaktadır. Bu yöntemler özellikle yetişkin öğrenme kuramları (andragoji) ile temellendirilmiştir.[21] Yetişkin öğrenme kuramlarına göre bireyler, kendi öğrenme süreçlerinde aktif rol alır, deneyimlerinden yola çıkarak anlam üretir ve içsel motivasyonla öğrenirler. Bu yaklaşım, uzmanlık öğrencilerinin yaşam boyu öğrenme becerilerinin gelişimine katkı sağlamaktadır.

TUKMOS Kapsamında Eğitici ve Mekan Standartları

Tıpta uzmanlık eğitiminin niteliği, yalnızca müfredatın içeriği ya da süresiyle değil; bu eğitimin sunulduğu fiziksel ortam ve süreci yöneten eğiticilerin yetkinliğiyle doğrudan ilişkilidir. TUKMOS, bu doğrultuda hem eğitici profiline hem de eğitim ortamlarına ilişkin standartlar tanımlayarak kalite güvencesini sağlamayı hedeflemektedir.

Her uzmanlık dalı için eğitici sayısı ve niteliklerine dair temel koşullar tanımlanmıştır.[1] Genellikle bir programın akredite edilebilmesi için en az iki eğiticinin bulunması gerekir. Bu kişilerin akademik yetkinliğe ve klinik deneyime sahip olmaları, yalnızca bilgi aktarıcı değil, aynı zamanda rol model, klinik mentor ve değerlendirici olarak da aktif görev almaları beklenmektedir.[4,5] Bu nedenle, TUKMOS müfredatları eğiticilerin mesleki gelişim, geri bildirim verme ve eğitim yöntemleri konusunda yetkin olmalarını zorunlu kılan bir yaklaşım benimsemektedir.

Yapılan çalışmalarda, asistanların büyük bir kısmı eğitimin önemli bir bölümünü kıdemli asistanlardan öğrendiklerini belirtmiş; eğiticilerin eğitim sürecine katılımının sınırlı olduğu ve erişimin kısıtlı kaldığı ifade edilmiştir.[15,22] Oysa yapılandırılmış geri bildirim mekanizmaları, yalnızca öğrenmeyi desteklemekle kalmaz; aynı zamanda öz değerlendirme becerilerinin gelişimini de sağlar.[23] Bu bağlamda, eğiticilere yönelik sürekli mesleki gelişim programlarının ulusal düzeyde yaygınlaştırılması büyük önem taşımaktadır.

Uzmanlık eğitiminin verildiği fiziki ortam da kaliteyi etkileyen bir diğer kritik faktördür. TUKMOS, eğitim birimlerinin yeterli fiziksel kapasiteye, olgu çeşitliliğine, laboratuvar ve görüntüleme altyapısına, girişimsel uygulamalara elverişli klinik koşullara ve gerektiğinde simülasyon destekli eğitim alanlarına sahip olmasını öngörmektedir.[1] Ayrıca, uygun dinlenme alanları, çalışma salonları ve akademik toplantı odaları da eğitim ortamının ayrılmaz bileşenleri olarak kabul edilmektedir.

Bu standartlara ilişkin yapılan değerlendirmelerde, örneğin Hacettepe Üniversitesi'nde yürütülen bir çalışmada, uzmanlık öğrencileri mevcut eğitim ortamlarını genel olarak olumlu bulmakla birlikte, geliştirilmesi gereken yönler olduğunu ifade etmiştir.[24] Eğitici sayısının artırılması, düzenli seminerlerin yapılması ve bilimsel etkinliklerin teşvik edilmesi gibi önerilmiştir.[25]

TUKMOS Kapsamında Rotasyonlar: Etkinlik ve Sorunlar

TUKMOS, her uzmanlık dalı için belirli sürelerde zorunlu dış rotasyonlar tanımlamıştır. Bu rotasyonların süresi 1 ila 18 ay arasında değişmekte olup bazı branşlarda ise dış rotasyon bulunmamaktadır (Tablo 1 ve 2).[26] Rotasyonların temel amacı, uzmanlık öğrencilerine farklı klinik ortamlarda çok yönlü deneyim kazandırmak ve yetkinlik alanlarını genişletmektir. Ancak sahada yapılan çeşitli araştırmalar, bu rotasyonların uygulamada hem süre hem de içerik açısından hedeflenen standartlara ulaşmakta zorlandığını göstermektedir. Örneğin, Göğüs Cerrahisi asistanlarıyla yapılan bir çalışmada, katılımcıların çoğu rotasyonların verimsiz geçtiğini, hedeflerden haberdar olmadıklarını ve eğitim sürecinde aktif görev almadıklarını ifade etmiştir.[27] Benzer şekilde, Yoğun Bakım yan dal asistanları arasında yürütülen bir çalışmada, rotasyonların içerik ve işleyişinin klinikler arasında önemli farklılıklar gösterdiği ve öğrenim hedeflerine ulaşılıp ulaşılmadığının düzenli biçimde değerlendirilmediği vurgulanmıştır.[28]

Olgu çeşitliliği sınırlı olan Tıbbi Biyokimya, Adli Tıp ve Çocuk Nefrolojisi gibi bazı alanlarda ise rotasyon içeriklerinin müfredata uygunluğu sorgulanmaktadır.[29-31] Genel olarak yapılan değerlendirmeler, TUKMOS tarafından tanımlanan rotasyonların çoğu zaman biçimsel olarak yürütüldüğünü; özellikle anabilim dalı dışındaki rotasyonlarda içerik bütünlüğü, denetim, eğitici katılımı ve ölçülebilir öğrenme hedeflerinin yeterince sağlanamadığını ortaya koymaktadır.

Bu bulgular, rotasyonların eğitim sürecine katkısını artırmak için yapısal iyileştirmelere ihtiyaç olduğunu

göstermektedir. Öncelikle, rotasyonların içeriği kurumlar arası standart bir program ile uyumlu hale getirilmeli; rotasyonlarda asistanların yalnızca gözlemci değil, aktif uygulayıcı rol alması sağlanmalıdır. Eğitici sorumlulukları netleştirilmeli, rotasyon süreçlerinin düzenli olarak dış denetimlerle izlenmesi gerekmektedir. Ayrıca, her rotasyon için spesifik öğrenme hedefleri ve değerlendirme kriterleri tanımlanmalı, bu hedeflere ulaşılıp ulaşılmadığı düzenli olarak raporlanmalıdır.

TUKMOS Kapsamında Ölçme-Değerlendirme

Tıpta uzmanlık eğitiminin başarısı, yalnızca müfredatın kapsamı ve süresiyle değil, öğrenme çıktılarının geçerli ve güvenilir biçimde değerlendirilmesiyle de doğrudan ilişkilidir.[32] Bu nedenle, eğitim programlarının etkililiğini ölçmek ve uzmanlık öğrencilerinin yeterlik düzeylerini değerlendirmek için çok boyutlu ölçme-değerlendirme yaklaşımlarına ihtiyaç vardır. Kuramsal bilgiyi ölçen çoktan seçmeli testlerin yanı sıra, objektif yapılandırılmış klinik sınavlar (OSCE), olgu temelli değerlendirmeler, simülasyon uygulamaları ve işyeri tabanlı yöntemler [Mini-Klinik Değerlendirme (mini-CEX), Doğrudan Gözlemli Klinik Uygulama (DOPS), 360 derece değerlendirme] önerilmektedir.[1]

TUKMOS kapsamında Kardiyoloji uzmanlık eğitimine yönelik yapılan bir çalışmada, en yaygın kullanılan değerlendirme araçlarının asistan karnesi, olgu temelli tartışma ve DOPS olduğu belirlenmiştir.[33] Buna karşın, mini-CEX, çok kaynaklı geri bildirim ve OSCE gibi yapılandırılmış araçların kullanım oranlarının düşük olduğu bildirilmiştir. Katılımcıların büyük çoğunluğu işyeri temelli değerlendirme yöntemlerinin yararlı olduğunu kabul etmekle birlikte, bu yöntemlerin Türkiye koşullarına uyarlanması gerektiğini vurgulamıştır.[33] TUKMOS müfredatları kapsamında bilgi, beceri ve tutum boyutlarına yönelik değerlendirme ölçütleri tanımlanmış olsa da, bu araçların sahadaki kullanım sıklığı ve uygulama biçimi kurumlar arasında belirgin farklılık göstermektedir. Kardiyoloji alanında yapılan başka bir çalışmada, özdeğerlendirme verilerinin, eğitim programının etkililiğini izlemek ve iyileştirmek açısından yararlı olduğu bulunmuştur.[14]

Göğüs Hastalıklarında, asistanların müfredatta tanımlı hedeflere ulaşmakta zorlandığı ve temel klinik yeterliklerde mezuniyet sonrası eksiklikler yaşadığı bildirilmiştir.[34] Bu durumun temel nedenlerinden biri, ölçme-değerlendirme uygulamalarının sistematik ve yapılandırılmış biçimde planlanamaması ve uygulanamamasıdır. Tıpta Uzmanlık Sınavı üzerine yapılan analizlerde, sınav sorularının temel hekimlik yeterliklerini ölçmekte yetersiz kaldığı, ayrıca ulusal çekirdek eğitim programı ve TUKMOS müfredatlarıyla tam olarak örtüşmediği ortaya konmuştur.[35] Bu sonuçlar, hem uzmanlık öncesi hem de sonrası değerlendirme süreçlerinin gözden geçirilmesi gerektiğini düşündürmektedir.

Uzmanlık eğitiminin kalitesini artırmak için ölçme ve değerlendirme süreçlerinin güçlendirilmesi gereklidir. Mevcut uygulamaların daha etkin olabilmesi için; işyeri temelli değerlendirme (mini-CEX, DOPS, 360 derece geri bildirim vb.) araçlarının yaygınlaştırılması ve düzenli olarak kullanılması gerekmektedir. Ayrıca, her kurumda ölçme-değerlendirme sürecini koordine edecek ve standardizasyonu sağlayacak sorumlu kişiler atanmalı; bu süreçler dijital portfolyo sistemleriyle izlenebilir hale getirilmelidir.

Uzmanlık Eğitiminde Bilimsel Araştırma Eğitimi

Uzmanlık öğrencileri, birçok çalışmada bilimsel araştırmalara ilgi duyduklarını ifade etmelerine rağmen, araştırma sürecine aktif katılım oranı düşüktür. Bu durumun temel nedenleri arasında zaman yetersizliği, maddi destek eksikliği ve danışmanlık desteğinin sınırlı olması öne çıkmaktadır.[36-38] Bir çalışmada, uzmanlık öğrencilerinin %66'sı eğitim programlarında araştırma faaliyetleri için zaman ayrılmadığını, %48,1'i ise hiç araştırma eğitimi almadığını belirtmiştir. Katılımcıların yarısından fazlası araştırmaya istekli olduğunu ifade etse de, yapısal destek eksikliği bu isteği pratiğe dönüştürmede engel oluşturmaktadır.[38]

TUKMOS, bilimsel araştırma kültürünü desteklemeyi ve uzmanlık öğrencilerinin araştırma becerilerini geliştirmeyi hedeflemektedir. Ancak, saha verileri bu hedeflerin uygulamada yeterince karşılık bulmadığını göstermektedir.[36-38] Eğitim sürecinde araştırmaya yönelik danışmanlık mekanizmalarının güçlendirilmesi, zaman planlamasının eğitim programlarına entegre edilmesi ve yapılandırılmış araştırma eğitimi modüllerinin yaygınlaştırılması, bu alandaki niteliksel iyileşme sağlayabilir. Buna ek olarak, asistanların araştırmalara aktif katılımı için gerekli teşvik sistemleri (ödül vb.) oluşturulmalı; eğiticilere yönelik mentorluk eğitimleri yaygınlaştırılmalıdır. Kurum içi ve kurumlar arası ortak projelerin desteklenmesi ile araştırma kültürünün geliştirilmesi sağlanabilir.

Ana Dal-Yan Dal ve Branşlar Arası Yetki Sınırı Sorunları

Günümüzde klinik uygulama alanlarının giderek örtüşmesi, ana ve yan dal ilişkilerinde yetki ve sorumluluk sınırlarında belirsizliğe yol açmaktadır.[39-41] Bu belirsizlik, hem uzmanlık eğitiminin kalitesini hem de sağlık hizmetlerinin etkinliğini olumsuz yönde etkilemektedir. Kadın Hastalıkları ve Doğum ile Perinatoloji, Kulak Burun Boğaz Hastalıkları ile Plastik, Rekonstrüktif ve Estetik Cerrahi gibi branşlarda klinik sorumluluk alanları sıklıkla çakışmakta; belirli girişimlerin hangi uzmanlık alanının sorumluluğunda olduğu konusunda karışıklıklar yaşanmaktadır.

Türkiye'de Kadın Doğum hekimleriyle yapılan bir çalışmada, perinatoloji uzmanlarının görev tanımlarına ilişkin farkındalığın yetersiz olduğu belirlenmiştir.[39] Özellikle özel sektörde görev yapan Kadın Doğum uzmanlarının fetal anomali, çoğul gebelik ve plasenta anomalileri gibi durumlarda perinatolojiye konsültasyon oranlarının kamuya göre anlamlı ölçüde daha düşük olduğu saptanmıştır.[39] Bu belirsizlik yalnızca hastane düzeyindeki uygulamalarla sınırlı değildir. Toplum sağlığı merkezlerinde görev yapan Halk Sağlığı uzmanlarının, TUKMOS müfredatında yer almayan defin işlemlerine yönelik hizmet sunmaları da benzer bir uygulama-müfredat uyumsuzluğunu ortaya koymaktadır.[40] Buna ek olarak, İş ve Meslek Hastalıkları yan dalında TUKMOS müfredatı tanımlanmış olsa da, görev tanımları, özlük hakları ve saha uygulamaları konusundaki boşluklar devam etmektedir.[41] Bu durum, müfredatların yalnızca içeriksel değil, aynı zamanda uygulanabilirlik ve mevzuat düzeyinde de güçlendirilmesi gerekmektedir.

Tüm bu bulgular, ana dal ve yan dal uzmanlık alanları arasında görev, yetki ve sorumluluk sınırlarının açık biçimde yeniden tanımlanması gerekliliğini ortaya koymaktadır. Eğitim sürecinde ana dal ve yan dal asistanlarının rotasyon, olgu paylaşımı ve eğitici erişimi konularında dengeli biçimde planlanması ve görev tanımlarının sağlık mevzuatında karşılığını bulacak şekilde yasal zemine oturtulması gereklidir.

SONUÇ

TUKMOS, uzmanlık eğitiminin niteliğini artırmak ve ulusal düzeyde standartlaştırmak amacıyla önemli bir çerçeve sunmaktadır. Ancak bu sistemin sahada etkili olabilmesi için sadece hazırlanması değil, aynı zamanda uygulanmasının düzenli olarak izlenmesi ve denetlenmesi gerekmektedir. Müfredatların tüm eğitim kurumlarında eksiksiz ve tutarlı şekilde hayata geçirilmesi, eğiticilerin bu sürece aktif olarak katılması ve asistanların belirlenen öğrenme hedeflerine ulaşması için bir denetim mekanizmasına ihtiyaç vardır. Bu denetim; yapılandırılmış geri bildirimler, kurumlar arası değerlendirmeler ve ölçme-değerlendirme verileriyle desteklenmelidir. Ayrıca, sorunların çözümü için dış rotasyonlarda standardizasyon, aktif asistan katılımı ve ölçme-değerlendirme süreçlerinin güçlendirilmesi, bilimsel araştırma kültürünün desteklenmesi gerekmektedir. TUKMOS'un zorunlu hale gelmesi, hem eğiticilerin hem de asistanların bu sisteme yönelik farkındalık ve sorumluluklarının artmasını gerektirirken, etkin bir uygulama ve kalite güvencesi için sistematik izleme ve dış değerlendirme süreçlerinin kurulması kritik önem taşımaktadır.

Dipnotlar

Yazarlık Katkıları

Konsept: M.D., M.K., Ö.C., Dizayn: M.D., M.K., Ö.C., Veri Toplama veya İşleme: M.D., M.K., Analiz veya Yorumlama: M.D., M.K., Ö.C., Literatür Arama: M.D., Yazan: M.D., M.K., Ö.C.

Çıkar Çatışması: Yazarlar bu makale ile ilgili olarak herhangi bir çıkar çatışması bildirmemiştir.

Finansal Destek: Çalışmamız için hiçbir kurum ya da kişiden finansal destek alınmamıştır.

KAYNAKLAR

1. T.C. Sağlık Bakanlığı. Tıpta Uzmanlık Kurulunca Kabul Edilen Müfredatlar. Erişim Linki: https://tuk.saglik.gov.tr/TR-82498/mufredatlar.html
2. Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. Med Teach. 2007;29:642-7.
3. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system--rationale and benefits. N Engl J Med. 2012;366:1051-6.
4. Frank JR, Snell L, Sherbino J. CanMEDS 2015 Physician Competency Framework. Ottawa: Royal College of Physicians and Surgeons of Canada. 2015.
5. Accreditation Council for Graduate Medical Education–International (ACGME-I). Chicago: ACGME-I. Erişim Linki: https://www.acgme-i.org/
6. T.C. Resmi Gazete. Tıpta ve Diş Hekimliğinde Uzmanlık Eğitimi Yönetmeliği. 2009;No:27292. Erişim Linki: https://dosyamerkez.saglik.gov.tr/Eklenti/13291/0/tuey2011pdf.pdf
7. T.C. Sağlık Bakanlığı. Tıpta Uzmanlık Kurulu. Müfredat Arşivi. Erişim Linki: https://tuk.saglik.gov.tr/TR-87123/mufredat-arsivi.html
8. T.C. Sağlık Bakanlığı. Tıpta Uzmanlık Kurulu. V2.5 Dokümanları. Erişim Linki: https://tuk.saglik.gov.tr/TR-87040/v-25-dokumanlari.html
9. T.C. Resmi Gazete. Tıpta ve Diş Hekimliğinde Uzmanlık Eğitimi Yönetmeliği. 2022; No. 31942. Erişim Linki: https://www.resmigazete.gov.tr/eskiler/2022/09/20220903-2.htm
10. Schneiderhan J, Guetterman TC, Dobson ML. Curriculum development: a how to primer. Fam Med Community Health. 2019;7:e000046.
11. T.C. Sağlık Bakanlığı. Tıpta Uzmanlık Kurulu. TUK Kararları. Uzmanlık Dallarının Eğitim Süreleri Hakkında (Karar No: 1087). Erişim Linki: https://tuk.saglik.gov.tr/TR-31567/uzmanlik-dallarinin-egitim-sureleri-hakkinda-karar-no-1087.html
12. T.C. Resmi Gazete. 6225 sayılı Kanun hakkında; Tıpta Uzmanlık Eğitim Sürelerine Dair Kural Değişikliği. 2011; No. 27916. Erişim Linki: https://resmigazete.gov.tr/eskiler/2011/04/20110426-1.htm
13. T.C. Sağlık Bakanlığı. Tıpta Uzmanlık Kurulu. TUK Kararları. Erişim Linki: https://tuk.saglik.gov.tr/TR-23634/tuk-kararlari.html?Sayfa=1
14. Yıldırır A, Turan S. Self evaluation of cardiology residents on the clinical competencies defined in the core curriculum. Tıp Eğitimi Dünyası. 2021;20:44-56.
15. Ceyhun Peker AG, Deryal GN, Sarıca Çevik H, Ungan M. Empowering medical residency training: a comparative analysis

for understanding outpatient clinic demand and training needs. BMC Med Educ. 2025;25:454.

16. Anderson LW, Krathwohl DR. A Taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. 2001. 1st ed. Publisher: New York: Longman.
17. Sen V, Eren H, Kazaz IO, et al. Easily accessible, up-to-date and standardised training model in urology: E-learning residency training programme (ERTP). Int J Clin Pract. 2021;75:e13683.
18. Kolb DA. Experiential learning: experience as the source of learning and development. 2nd ed. New Jersey: Pearson Education; 2015.
19. Kocal S, Yılmaz F, Eyinç MB, Öztoprak H, Şirin İ. Investigation of emergency medicine residents' skills in using external pacemakers. Emergency Care Journal. 2025;21:13169.
20. Altıntop M, Uzun K, Karakullukçu S, Ardıç C. Evaluation of simulation-based educational approaches in family medicine specialization education. Konuralp Medical Journal. 2022;14:438-44.
21. Knowles MS, Holton EF, Swanson RA. The adult learner: the definitive classic in adult education and human resource development. 8th ed. London: Routledge; 2020.
22. Onat L, Hancı V, Özbilgin Ş, et al. Anaesthesiology and reanimation education in Turkey: residents' views. Turk J Anaesthesiol Reanim. 2019;47:496-502.
23. Lägervik M, Thörne K, Fristedt S, Henricson M, Hedberg B. Residents' and supervisors' experiences when using a feedback-model in post-graduate medical education. BMC Med Educ. 2022;22:891.
24. Atılgan SBÇ, Taşdelen Teker G, Sezer B, Yeşiltepe M, Odabaşı O. Hacettepe üniversitesi'nde tıpta uzmanlık öğrencilerinin klinik eğitim ortamlarına ilişkin algılarının değerlendirilmesi. STED. 2020;29:44-53.
25. Karademir M, Ağadayı E, Karahan S, Karademir D. Evaluation of medical specialization education students' perceptions of hospital education environment and affecting factors. Cumhuriyet Med J. 2022;44:348-55.
26. T.C. Sağlık Bakanlığı. Tıpta Uzmanlık Kurulu. TUK Kararları. Uzmanlık Eğitimi Rotasyonları (Karar No: 82). Erişim Linki: https://tuk.saglik.gov.tr/TR-89541/uzmanlik-egitimi-rotasyonlari.html
27. İşevi C, Pirzirenli MG, Çelik B, Bilgin Büyükkarabacak Y. Knowledge, attitudes and beliefs of thoracic surgery residents towards theoretical and practical medical life in Turkey. Curr Thorac Surg. 2024;9:132-40.
28. Arıkan H, Ürkmez S, Yardaş G, Başar H, Özkoçak Turan I, Demirkıran O. A current overview of intensive care subspeciality education in Turkey: what do educational staff, subspecialty residents and specialists think? Turk J Intensive Care. 2023;21:288-96.
29. Aktaş EÖ. Specialisation in forensic medicine advances and changes. Bull Leg Med. 2015;20.
30. Çamlar SA, Şahin H. Do pediatric nephrologists know their educational competencies? J Curr Pediatr. 2024;22:79-92.
31. Aybek H. Specialization education curriculum of medical biochemistry department, Pamukkale University Faculty of Medicine. Tıp Eğitimi Dünyası. 2018;52:70-3.
32. Holt KD, Miller RS, Nasca TJ. Residency programs' evaluations of the competencies: data provided to the ACGME about types of assessments used by programs. J Grad Med Educ. 2010;2:649-55.
33. Yıldırır A, Turan S. Assessment and evaluation of cardiology residency training in Türkiye: a national survey. Anatol J Cardiol. 2023;27:580-91.
34. Gayef A, Karlıkaya HC. Pulmonologists' competencies in bronchoscopy graduated from the same university clinic. Tıp Eğitimi Dünyası. 2022;21:31-41.
35. Aydoğan S, Keskin A. How much does the medical specialization exam measure basic medicine? Tıp Eğitimi Dünyası. 2022;21:42-9.
36. Eyigör H, Kara CO. Problems and suggestions about operating room education in otorhinolaryngology; trainees and trainers are talking. Tıp Eğitimi Dünyası. 2020;19:171-7.
37. Şenel S, Budakoğlu İİ. Evaluation of pediatrics residency students' competency perceptions towards UCEP 2020 basic medical practices. Tıp Eğitimi Dünyası. 2024;23:70-88.
38. Tatlı DG, Şenol Y. Investigation of the scientific study planning and implementation experiences of residents'. Tıp Eğitimi Dünyası. 2024;23:71-80
39. Altınordu Atcı A, Doğru Ş, Akkuş F. Awareness of gynecology and obstetricians in Turkey about perinatology specialist job definitions. Pamukkale Tıp Dergisi. 2022;15:682-92.
40. Acımış NM. Overview of burial services in family medicine. Klinik Tıp Aile Hekimliği Dergisi. 2018;10:19-21.
41. Karabağ İ, Şahan C, Alagüney ME, Yıldız NA. Opinions of employees on the duties, authorities, responsibilities and personal rights of occupational medicine specialists. Turk J Public Health. 2022;20:329-45.

DOI: 10.4274/forbes.galenos.2025.02350
Forbes J Med 2025;6(2):113-8

Original Article / Orijinal Araştırma

The Role of Biochemical and Hematological Parameters in Urethral Stricture Recurrence: Inflammation Indexes and De Ritis Rate

Biyokimyasal ve Hematolojik Parametrelerin Üretra Darlığı Rekürrensindeki Rolü: De Ritis Oranı ve Enflamasyon İndeksleri

Serkan ÖZCAN[1], Kürşad DÖNMEZ[2], Enis Mert YORULMAZ[1], Osman KÖSE[1], Sacit Nuri GÖRGEL[1], Yiğit AKIN[1]

[1]İzmir Katip Çelebi University Faculty of Medicine, Department of Urology, İzmir, Türkiye
[2]İzmir Katip Çelebi University Atatürk Training and Research Hospital, Department of Urology, İzmir, Türkiye

Cite as: Özcan S, Dönmez K, Yorulmaz EM, Köse O, Görgel SN, Akın Y. The role of biochemical and hematological parameters in urethral stricture recurrence: inflammation indexes and De Ritis rate. Forbes J Med. 2025;6(2):113-8

ABSTRACT

Objective: This study investigates the role of hematological inflammation markers in predicting urethral stricture recurrence. Specifically, it examines the prognostic significance of the systemic immune-inflammation index (SII) and De Ritis ratio [aspartate aminotransferase (AST)/alanine aminotransferase ALT)] in recurrence risk.

Methods: This retrospective cohort study included 51 patients who underwent direct visual internal urethrotomy between 2019 and 2023. Patients were divided into recurrence and non-recurrence groups. Clinical data, hematological, and biochemical parameters were analyzed, and the prognostic value of SII and De Ritis ratio (AST/ALT) was assessed using statistical methods.

Results: SII was identified as a significant predictor of recurrence risk (area under the curve=0.689, p=0.023). SII values exceeding 591.42 were associated with a fourfold increase in recurrence risk (odds ratio=4.0, 95% confidence interval: 1.36-11.74, p<0.05). In contrast, the De Ritis ratio (AST/ALT) was not correlated with recurrence (p=0.924). Additionally, stricture length >2 cm was significantly associated with higher recurrence risk (p=0.024).

Conclusion: SII appears to be a valuable biomarker for predicting urethral stricture recurrence. Integrating SII into clinical practice could facilitate early identification of high-risk patients and enable the development of personalized treatment strategies. For patients with high SII values, open urethroplasty may be preferable to repeated endoscopic procedures. However, prospective multicenter studies are needed to validate these findings in larger patient populations. Additionally, machine learning-based prediction models may enhance the accuracy of recurrence risk assessment.

Keywords: Urethral stricture, systemic immuno-inflammation index, De Ritis ratio

ÖZ

Amaç: Bu çalışmada, üretral darlık rekürrensinin öngörülmesinde hematolojik enflamasyon belirteçlerinin rolü incelenmiştir. Özellikle sistemik immün-enflamasyon indeksi (Sİİ) ve De Ritis oranının [aspartat aminotransferaz (AST)/alanin aminotransferaz (ALT)] rekürrens riski üzerindeki etkileri araştırılmıştır.

Yöntem: Bu retrospektif kohort çalışmasına, 2019-2024 yılları arasında doğrudan görüşlü iç üretrotomi uygulanan 51 hasta dahil edilmiştir. Hastalar rekürrens gelişenler ve gelişmeyenler olarak iki gruba ayrılmıştır. Klinik veriler, hematolojik ve biyokimyasal parametreler incelenmiş; Sİİ ve De Ritis oranının (AST/ALT) prognostik değeri istatistiksel analizlerle değerlendirilmiştir.

Received/Geliş: 04.03.2025
Accepted/Kabul: 15.04.2025
Epub: 17.07.2025
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Corresponding Author/ Sorumlu Yazar:
Serkan ÖZCAN, MD,
İzmir Katip Çelebi University Faculty of Medicine, Department of Urology, İzmir, Türkiye
drserkanozcan@hotmail.com
ORCID: 0000-0002-2459-139X

Bulgular: Rekürrens riski için Sİİ'nin anlamlı bir belirteç olduğu saptanmıştır (eğri altında kalan alan=0,689, p=0,023). 591,42'nin üzerindeki Sİİ değerleri, üretral darlık rekürrens riskini 4 kat artırmıştır (olasılık oranı=4,0, 95% güven aralığı: 1,36-11,74, p<0,05). Buna karşılık, De Ritis oranının (AST/ALT) rekürrens ile ilişkili olmadığı bulunmuştur (p=0,924). 2 cm'den uzun darlıkların rekürrens riski belirgin şekilde daha yüksektir (p=0,024).

Sonuç: Sİİ, üretral darlık rekürrensini öngörmede değerli bir biyomarker olarak öne çıkmaktadır. Klinik pratiğe entegrasyonu, yüksek riskli hastaların erken tespit edilmesine ve bireyselleştirilmiş tedavi stratejilerinin geliştirilmesine katkı sağlayabilir. Tekrarlayan endoskopik işlemler yerine açık üretroplasti, yüksek Sİİ değerleri olan hastalar için daha uygun bir seçenek olabilir. Ancak, bu bulguların daha geniş hasta gruplarında doğrulanması için ileriye dönük çok merkezli çalışmalara ihtiyaç vardır. Ayrıca, yapay zeka tabanlı tahmin modellerinin kullanılması, rekürrens öngörüsünü daha hassas hâle getirebilir.

Anahtar Kelimeler: Üretra darlığı, sistemik immün-enflamasyon indeksi, De Ritis oranı

INTRODUCTION

Urethral strictures (US) are medical conditions resulting from fibrotic changes in the urethral epithelium and corpus spongiosum. These changes can lead to lower urinary tract symptoms, posing a significant clinical concern for both patients and healthcare providers. Urethral lumen narrowing may occur due to trauma, infection, inflammation, or surgical interventions. The high recurrence rate of strictures substantially impacts patients' quality of life.[1]

Inflammation plays a crucial role in the pathogenesis of US. Chronic inflammatory processes can lead to stenosis by damaging the urethral mucosa, promoting connective tissue proliferation, and forming scar tissue. The complex interactions among various cellular components-such as neutrophils, lymphocytes, and platelets-are believed to contribute significantly to stricture recurrence.[2] However, the specific biological markers associated with this inflammatory process remain insufficiently understood.

US predominantly affect men, with a prevalence of approximately 0.6%. The associated healthcare costs exceed $200 million annually in direct expenditures.[3] Patients typically present with symptoms such as dysuria (painful urination), pollakiuria (frequent urination), and reduced urine flow. In advanced stages, US can lead to severe complications, including urinary retention and recurrent urinary tract infections.[4]

The management of US depends on various factors, including the length, location, and underlying cause of the stricture. One widely used treatment modality is endoscopic intervention, favored for its cost-effectiveness, minimally invasive nature, and reproducibility. Among these procedures, direct visual internal urethrotomy (DVIU) is particularly notable.[5] However, the high recurrence rates following DVIU highlight the need for new prognostic markers to improve patient outcomes.

Recent research has focused on hematological inflammation indices to enhance our understanding of US and improve their clinical management. Key biomarkers such as the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), systemic immune-inflammation index (SII), and De Ritis ratio [aspartate aminotransferase (AST)/alanine aminotransferase (ALT)] have emerged as potential indicators of inflammation's role in these pathophysiological processes.[6,7]

Studies have explored the significance of inflammation indices in predicting the prognosis of various urological diseases. Specifically, NLR and SII have been investigated in relation to cardiovascular diseases, cancer, and chronic inflammatory conditions.[8] However, their impact on urethral stricture recurrence risk remains insufficiently explored.[9]

Although these biomarkers are thought to correlate with stricture severity, recurrence risk, and treatment response, they must be considered alongside other clinical and radiological findings. This study aims to assess whether the De Ritis ratio (AST/ALT) and SII can predict urethral stricture recurrence by analyzing clinical preoperative data from patients undergoing single or multiple DVIU procedures. Our findings may provide valuable insights for developing personalized treatment strategies in urethral stricture management.

METHODS

This retrospective cohort study examines the clinical outcomes of 51 patients who underwent DVIU at the Urology Clinic of İzmir Katip Çelebi University Atatürk Training and Research Hospital, between 2019 and 2024. Among them, 21 patients underwent a single DVIU, while 30 patients required multiple DVIU procedures. Data were meticulously collected from medical records within the urology service and outpatient clinic. The study was approved by the İzmir Katip Çelebi University Health Research Ethics Committee (decision number: 0063, date: 13.02.2025) and was conducted in accordance with the Declaration of Helsinki.

The study included male patients aged 18 years or older who had previously received treatment for urethral stricture. However, the following exclusion criteria were applied: patients with incomplete medical records, those diagnosed with malignancies, individuals with uncontrolled diabetes mellitus (DM), and patients with hematological, liver, or kidney dysfunction. Additionally, patients who

had received blood transfusions or had undergone open urethral surgery were excluded from the study.

The collected patient data included both demographic and clinical information. Demographic variables included age, comorbidities, and smoking status, while clinical data covered stenosis recurrence, duration of recurrence, postoperative catheterization time, and follow-up period.

Additionally, a complete blood count and biochemical analyses were conducted as part of the preoperative anesthesia evaluation. These included neutrophil and lymphocyte counts, hematocrit (HTC) levels, albumin levels, NLR, De Ritis ratio (AST/ALT), and SII. The SII was calculated using the formula: SII = (neutrophil count × platelet count)/lymphocyte count. All laboratory tests were performed using standardized methods in the hospital's laboratory.

Preoperative evaluation of US was performed using retrograde urethrography. The frequency and duration of stricture recurrence were recorded based on follow-up uroflowmetry and cystoscopy results. Recurrence duration was defined as the time until either the first clinical signs appeared (including symptoms and uroflowmetry findings) after DVIU, or when a repeat DVIU became necessary. Patients were monitored through clinical evaluations and uroflowmetry every three months during the first year post-DVIU, and every six months in the subsequent two years.

Statistical Analysis

All analyses were carried out using IBM SPSS Statistics 27 (Statistical Package for the Social Sciences) software. Before conducting the statistical analyses, the distribution of the data was assessed, and non-parametric tests were selected for use when the data did not conform to a normal distribution. The Mann-Whitney U test was employed for group comparisons. For the analysis of categorical variables, the chi-square (χ^2) test and Fisher's exact test were utilized. Furthermore, the receiver operating characteristic analysis was conducted to evaluate the discriminatory power of continuous variables, with the area under the curve (AUC) calculated accordingly. Descriptive statistics are presented as mean ± standard deviation or mean±standard error, depending on the specific type of variable. A significance level of $p<0.05$ was established for all statistical tests.

RESULTS

A total of 51 patients were included in the study. The mean age was 63.7±2.5 years in the non-recurrence group and 64.6±2.7 years in the recurrence group. No significant difference in age was observed between the two groups ($p>0.05$). Smoking prevalence was 66.7% in the non-recurrence group and 76.7% in the recurrence group; however, this difference was not statistically significant. The mean time to recurrence was 24.6±8.4 months. The follow-up period was 15.3±1.9 months (range: 3-32 months) for non-recurrent cases, while it was 25±4.6 months (range: 1-120 months) in recurrent cases (Table 1).

The prevalence of hypertension (HT), DM, and coronary artery disease (CAD) did not differ significantly between the recurrence and non-recurrence groups. HT prevalence was 47.6% in the non-recurrence group and 53.3% in the recurrence group. DM incidence was 28.6% in the non-recurrence group and 20% in the recurrence group. CAD was observed in 4.8% of non-recurrent cases, while it affected 26.7% of recurrent cases (Table 1). The mean postoperative catheterization duration was 6.76±0.3 days in the non-recurrence group and 7.37±0.2 days in the recurrence group. Notably, patients with a stricture length >2 cm had a significantly higher risk of recurrence ($p=0.024$).

Of the non-recurrence group, 59.1% of patients had a stricture length <2 cm, whereas of the recurrence group,

Table 1. Demographic data (non-recurrence and recurrence groups)

	No recurrence (N) (Mean±SD)	**Recurrence (N) (Mean±SD)**	**p value**
Age (years)	63.7±2.5	64.6±2.7	p>0.05
Smoking (%)	66.7%	76.7%	
Stenosis recurrence time (months)	-	24.6±8.4	
Follow-up period (months)	15.3±1.9 (3-32)	25±4.6 (1-120)	
Hypertension (%)	10 (47.6%)	16 (53.3%)	
Diabetes (%)	6 (28.6%)	6 (20.0%)	
Coronary artery disease (%)	1 (4.8%)	8 (26.7%)	
Post-operative catheter duratiom (days)	6.76±0.3	7.37±0.2	
Stricture length <2 cm (%)	13 (59.1%)	9 (40.9%)	**p=0.024**
Stricture length >2 cm (%)	8 (27.6%)	21 (72.4%)	

SD: Standard deviation

72.4% had a stricture length >2 cm (Table 1). HTC levels were 40.8±0.69 in both groups and showed no significant association with recurrence (p>0.05). Additionally, AST and ALT levels were comparable between groups (p>0.05). No significant differences were found in neutrophil, platelet, lymphocyte, or monocyte levels (Table 2).

The De Ritis ratio (AST/ALT) was not a significant predictor of urethral stricture recurrence (AUC=0.508, p=0.924). In contrast, SII was a significant predictor of recurrence [AUC=0.689, 95% confidence interval (CI): 0.541-0.837, p=0.023]. The optimal SII cut-off was 591.42. Patients with SII >591.42 had a fourfold increased risk of recurrence (odds ratio=4.0, 95% CI: 1.36-11.74, p<0.05) (Figure 1).

No significant association was found between smoking, DM, HT, CAD, and stricture recurrence risk (p>0.05). However, stricture length >2 cm was significantly associated with recurrence (p=0.024).

DISCUSSION

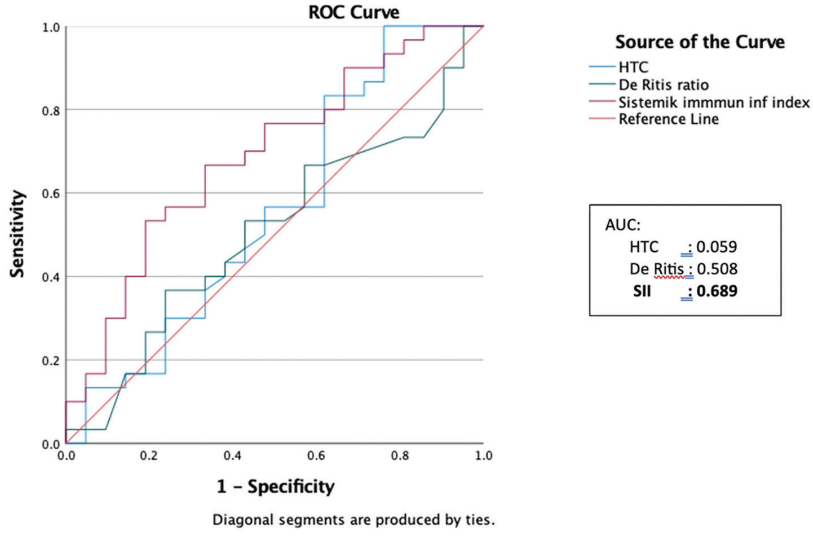
US is a chronic condition characterized by inflammation and spongiofibrosis, posing a significant clinical challenge in urology due to its high recurrence rates.[1] In recent years, interest in hematological inflammation indices has increased, particularly for predicting recurrence risk and optimizing treatment strategies.[7]

This study aimed to evaluate whether specific inflammatory markers could predict urethral stricture recurrence. Our findings indicate that SII is a strong predictor of recurrence risk (AUC=0.689, p=0.023). Notably, patients with SII >591 had a markedly higher recurrence

Table 2. Laboratory data (non-recurrence and recurrence groups)

	No recurrence (Mean±SD)	Recurrence (Mean±SD)	p value
HTC (%)	40.8±0.69	40.8±0.69	p>0.05
ALT (U/L)	17.4±1.0	17.4±1.0	
AST (U/L)	19.2±5.81	19.2±5.81	
Neutrophil (x10^9/L)	5.67±2.68	5.67±2.68	
Platelet (x10^9/L)	270±82.1	270±82.1	
Lymphocyte (x10^9/L)	1.99±0.73	1.99±0.73	
Monocyte (x10^9/L)	0.707±0.467	0.707±0.467	

SD: Standard deviation; HTC: Hematocrit, ALT: Alanine aminotransferase, AST: Aspartate aminotransferase



	Cut-Off	Sensitivity	Specificity	Area	OR	95% Confidence Interval	p
SII	591.42	%67	%66	0.689	4.0	1.36 – 11.74	0.023

Figure 1. Sensitivity, specificity and ROC curve analysis of SII, De Ritis and HTC

ROC: Receiver operating characteristic; AUC: Area under the curve; HTC: Hematocrit; OR: Odds ratio; SII: Systemic immune-inflammation index

risk. This result aligns with previous studies exploring the link between inflammation and fibrosis development.[7]

A large-scale study by Özsoy et al.[7] (703 patients) identified a clear association between SII and urethral stricture recurrence. Their findings suggest that SII may serve as a reliable biomarker for predicting recurrence.

Similarly, a study by Urkmez et al.[8] found a strong correlation between NLR and urethral stricture recurrence. Our results further reinforce the potential role of inflammatory biomarkers in predicting recurrence risk.

A study (208 patients) by Tokuc et al.[10] aimed to develop a preoperative method for predicting primary urethroplasty success. Their innovative approach integrates machine learning algorithms with inflammation indices such as NLR, PLR, SII, and pan-immune-inflammation values. These findings represent a notable advancement in clinical practice and provide promising opportunities for the development of personalized treatment strategies.

In a study conducted by Gul et al.[11] involving 303 patients, various risk factors for recurrent urethral stricture following internal urethrotomy were investigated, and a significant association was found between stricture recurrence and inflammation parameters. The study particularly highlighted that hematological inflammation indices might play a crucial role in predicting recurrent US. Our findings are consistent with this study and further support the impact of inflammation on urethral scar formation.

Our study did not establish a meaningful relationship between the De Ritis ratio (AST/ALT) and recurrence (p=0.924), suggesting that its role in urethral stricture pathophysiology may be limited. Conversely, our results highlight a clear association between SII and recurrence, reinforcing the idea that systemic inflammation plays a key role in urethral scar formation mechanisms. In the literature, the effects of inflammation on epithelial damage and fibrotic tissue development have been widely studied, and cytokines such as TGF-β and IL-6 are known to play a significant role in this process.[2,12]

When evaluating patient comorbidities, no statistically significant differences were detected between recurrence and non-recurrence groups regarding smoking, HT, diabetes, or CAD (p>0.05). However, patients with a stricture length >2 cm exhibited a substantially higher recurrence risk (p=0.024).

A systematic meta-analysis by Endo et al.[9] examined recurrence risk factors after internal urethrotomy. Their study emphasized the impact of stricture length, surgical technique, inflammation, and patient characteristics on recurrence rates. Our findings are in line with previous studies, further underscoring the importance of inflammatory and anatomical factors in recurrence prediction.

Another meta-analysis by Ma et al.[13] identified smoking as an independent risk factor for stricture recurrence after urethroplasty. Smoking has been shown to contribute to chronic inflammation and impair vascular endothelial function, thereby increasing the risk of stricture recurrence. In our study, smoking rates were higher in the recurrence group, but this difference did not reach statistical significance. Future research involving larger patient cohorts may provide more conclusive evidence regarding the influence of smoking on urethral stricture recurrence.[14]

Study Limitations

This study has certain limitations. Its retrospective design hinders the assessment of dynamic changes in inflammatory markers over time, while the relatively small, single-center sample may limit the generalizability of the findings. To validate these results and enhance their clinical applicability, future prospective, multicenter studies with larger patient populations are needed.

CONCLUSION

This study demonstrates that SII is a significant biomarker for predicting urethral stricture recurrence after DVIU. When SII levels exceed a specific threshold, the risk of recurrence increases substantially. In contrast, the De Ritis ratio (AST/ALT) was not found to be associated with urethral stricture recurrence, indicating that its prognostic value in this context is limited.

Our findings also suggest that patients with a stricture length greater than 2 cm face a significantly higher recurrence risk. This emphasizes that both inflammatory processes and anatomical factors contribute to stricture recurrence. Therefore, for patients with longer strictures, it may be beneficial to consider surgical treatment options at an earlier stage.

Integrating SII into clinical practice could help identify high-risk patients at an earlier stage, allowing for the development of personalized treatment strategies. In cases where SII values are particularly high, open urethroplasty techniques may be preferable over repeated endoscopic procedures.

Additionally, incorporating machine learning algorithms into recurrence prediction models may further enhance their accuracy and clinical utility. By leveraging advanced predictive tools, clinicians may be able to make more informed decisions and optimize patient management in urethral stricture treatment.

Ethics

Ethics Committee Approval: The study was approved by the İzmir Katip Çelebi University Health Research Ethics Committee (decision number: 0063, date: 13.02.2025).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: S.Ö., Y.A., Concept: S.Ö., S.N.G., Design: S.Ö., K.D., Data Collection or Processing: K.D., S.N.G., Analysis or Interpretation: E.M.Y., O.K., Y.A., Literature Search: E.M.Y., O.K., Writing: S.Ö., K.D.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Stormont TJ, Suman VJ, Oesterling JE. Newly diagnosed bulbar urethral strictures: etiology and outcome of various treatments. J Urol. 1993;150:1725-8.
2. Doersch KM, Barnett D, Chase A, Johnston D, Gabrielsen JS. The contribution of the immune system to genitourinary fibrosis. Exp Biol Med (Maywood). 2022;247:765-78.
3. Santucci RA, Joyce GF, Wise M. Male Urethral stricture disease. J Urol. 2007;177:1667-74.
4. Pansadoro V, Emiliozzi P. Iatrogenic prostatic urethral strictures: classification and endoscopic treatment. Urology. 1999;53:784-9.
5. Lacy JM, Cavallini M, Bylund JR, Strup SE, Preston DM. Trends in the management of male urethral stricture disease in the veteran population. Urology. 2014;84:1506-9.
6. Gül M, Altıntaş E, Kaynar M, Buğday MS, Göktaş S. The predictive value of platelet to lymphocyte and neutrophil to lymphocyte ratio in determining urethral stricture after transurethral resection of prostate. Turk J Urol. 2017;43:325-9.
7. Özsoy E, Kutluhan MA, Tokuç E, et al. Predictive value of systemic immune-inflammation index in recurrent urethral strictures. Urologia. 2022;90:510-5.
8. Urkmez A, Topaktas R, Ozsoy E, et al. Is neutrophil to lymphocyte ratio a predictive factor for recurrence of urethral stricture? Rev Assoc Med Bras. 2020;65:1448-53.
9. Endo D, Robayo J, García-Perdomo HA. Predictors of urethral stricture recurrence following internal urethrotomy: a systematic review. Urologia. 2025;92:32-8.
10. Tokuc E, Eksi M, Kayar R, et al. Inflammation indexes and machine-learning algorithm in predicting urethroplasty success. Investig Clin Urol. 2024;65:240-7.
11. Gul A, Ekici O, Zengin S, Barali D, Keskin T. Investigation of risk factors in the development of recurrent urethral stricture after internal urethrotomy. World J Clin Cases. 2024;12:2324-31.
12. Wessells H, Angermeier KW, Elliott S, et al. Male urethral stricture: American Urological Association Guideline. J Urol. 2017;197:182-90.
13. Ma YC, Lin L, Luo Z, Jin T. Smoking is an independent risk factor for stricture recurrence after the urethroplasty: a systematic review and meta-analysis. Int Braz J Urol. 2023;49:8-23.
14. Azhar RA, Bochner B, Catto J, et al. Enhanced recovery after urological surgery: a contemporary systematic review of outcomes, key elements, and research needs. Eur Urol. 2016;70:176-87.

DOI: 10.4274/forbes.galenos.2025.93824
Forbes J Med 2025;6(2):119-26

Original Article / Orijinal Araştırma

Imaging and the Value of the Pediatric Appendicitis Score for Diagnosis of Acute Appendicitis in the Pediatric Emergency Department

Çocuk Acil Servisinde Akut Apandisit Tanısı için Görüntüleme ve Pediatrik Apandisit Skorunun Değeri

Tuğçe NALBANT[1], Gamze GÖKALP[1], Yüksel BİCİLİOĞLU[1], Emine Burcu ÇIĞŞAR KUZU[2], Esin ERGÖNÜL[3]

[1]İzmir Katip Çelebi University Faculty of Medicine, Department of Pediatric Emergency, İzmir, Türkiye
[2]İzmir Katip Çelebi University Faculty of Medicine, Department of Pediatric Surgery, İzmir, Türkiye
[3]Dokuz Eylül University Faculty of Medicine, Department of Medical Education, İzmir, Türkiye

Cite as: Nalbant T, Gökalp G, Bicilioğlu Y, Çığşar Kuzu EB, Ergönül E. Imaging and the value of the pediatric appendicitis score for diagnosis of acute appendicitis in the pediatric emergency department. Forbes J Med 2025;6(2):119-26

ABSTRACT

Objective: Acute appendicitis (AA) is a common cause of acute abdominal pain in pediatric patients. The diagnosis of AA in children can sometimes be difficult to make accurately. For this reason, ultrasonography (US) and computed tomography (CT) are also commonly used methods for diagnosis. Scoring systems such as the pediatric appendicitis score (PAS) have gained attention because they integrate multiple clinical and laboratory parameters to help predict AA. The aim of this study was to evaluate the efficacy of various clinical, laboratory, and imaging parameters in diagnosing appendicitis in children and to determine which factors are most predictive for accurate diagnosis.

Methods: The study included 153 patients who presented to the pediatric emergency department with abdominal pain and were operated on for AA. PAS was calculated for each case. Patients were divided into two groups as "appendicitis" and "non-appendicitis" according to pathology results.

Results: The positive predictive value and negative predictive value of US in the diagnosis of appendicitis were found to be 60% and 87%, respectively, while CT showed higher accuracy rates with 86% sensitivity and 97% specificity. Logistic regression modeling to investigate which parameters influence the histomorphologic diagnosis of appendicitis revealed that younger age increased the likelihood of appendicitis by 4.29-fold, high white blood cell increased it by 5.44-fold, high PAS increased it by 7.87-fold, and diagnostic US increased it by 7.91-fold.

Conclusion: The study suggests that the combination of laboratory tests and imaging modalities can improve the accuracy of appendicitis diagnosis, especially in children, and prevent unnecessary surgery.

Keywords: Acute appendicitis, pediatric appendicitis score, clinical decision

ÖZ

Amaç: Akut apandisit (AA), çocuk hastalarda akut karın ağrısının yaygın bir nedenidir. Çocuklarda AA tanısını doğru koymak bazen zor olabilir. Bu nedenle ultrasonografi (US) ve bilgisayarlı tomografi (BT) de tanı için yaygın olarak kullanılan yöntemlerdir. Pediatrik apandisit skoru (PAS) gibi skorlama sistemleri, AA'yı öngörmeye yardımcı olmak için birden fazla klinik ve laboratuvar parametresini entegre ettiği için dikkat çekmiştir. Bu çalışmanın amacı, çocuklarda apandisit tanısı koymada çeşitli klinik, laboratuvar ve görüntüleme parametrelerinin etkinliğini değerlendirmek ve hangi faktörlerin doğru tanı için en öngörücü olduğunu belirlemektir.

Received/Geliş: 10.03.2025
Accepted/Kabul: 25.04.2025
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Corresponding Author/ Sorumlu Yazar:
Tuğçe NALBANT, MD,
İzmir Katip Çelebi University Faculty of Medicine, Department of Pediatric Emergency, İzmir, Türkiye
✉ tugcelik83@gmail.com
ORCID: 0000-0002-7408-2382

Presented in: This study was presented as a poster at the "20. Ulusal Çocuk Acil Tıp ve Yoğun Bakım Kongresi ve 16. Çocuk Acil Tıp ve Yoğun Bakım Hemşireliği Kongresi, 27-30 Kasım 2024, Cornelia Diamond Otel, Antalya, Türkiye"

Yöntem: Çalışmaya çocuk acil servisine karın ağrısı ile başvuran ve AA ön tanısı ile ameliyat edilen 153 hasta dahil edildi. Fizik muayene bulguları ve laboratuvar parametreleri kullanılarak her olgu için PAS hesaplandı. Hastalar patoloji sonuçlarına göre "apandisit" ve "apandisit olmayan" olarak iki gruba ayrıldı. Bu iki grup laboratuvar parametreleri, görüntüleme bulguları ve PAS açısından karşılaştırıldı.

Bulgular: Apandisit tanısında US'nin pozitif prediktif değeri ve negatif prediktif değeri sırasıyla %60 ve %87 olarak bulunurken, BT %86 duyarlılık ve %97 özgüllük ile daha yüksek doğruluk oranları gösterdi. Apandisitin histomorfolojik tanısını hangi parametrelerin etkilediğini araştırmak için yapılan lojistik regresyon modellemesi, genç yaşın apandisit olasılığını 4,29 kat, yüksek beyaz kan hücresi'nin 5,44 kat, yüksek PAS'nin 7,87 kat ve tanısal US'nin 7,91 kat artırdığını ortaya koydu.

Sonuç: Bu çalışma, laboratuvar testleri ve görüntüleme yöntemlerinin kombinasyonunun özellikle çocuklarda apandisit tanısının doğruluğunu artırabileceğini ve gereksiz cerrahiyi önleyebileceğini göstermektedir.

Anahtar Kelimeler: Akut apandisit, pediatrik apandisit skoru, klinik karar

INTRODUCTION

Acute onset abdominal pain is one of the most common causes of admission to the pediatric emergency department. When evaluating abdominal pain according to its etiology, it is necessary to first rule out the causes of acute abdomen because of the high mortality and morbidity associated with it. In childhood, especially in a certain age group, the most common and important cause of acute abdomen presenting to the emergency department is acute appendicitis (AA).[1-3] AA is most commonly diagnosed in the second decade of life. The average lifetime risk of developing AA is 8%.[4,5] In the United States of America, 80,000 children are treated for AA every year, and this rate was reported to be 4/1000 in children under 14 years, of age.[6]

In addition to the high number of cases presenting to emergency departments with abdominal pain, the difficulty in diagnosing AA in pediatric patients poses a diagnostic challenge. In approximately one-third of pediatric cases, AA does not present with typical symptoms. This leads to delayed diagnosis, the emergence of complications, and an increase in mortality rates.[7] For this reason, additional imaging methods and scoring systems that facilitate diagnosis are becoming increasingly important.[8,9] A detailed history and focused physical examination are of paramount value. Initial diagnostic tests should evaluate the serum levels of inflammatory biomarkers in these patients. Among the imaging methods, ultrasonography (US) should be the initial examination because it is non-invasive, it is accessible, and its cost is low compared to other methods. However, the inherent conditions, such as dependence on the individual and being affected by adipose tissue and intestinal gases, make it difficult for this examination to reach an accurate result.[10,11] If there is difficulty in diagnosis, pelvic-abdominal computed tomography (CT) is used in many clinics, including ours, for the diagnosis of AA in children due to its high sensitivity, and specificity.[12,13] Although magnetic resonance imaging (MRI) is known to produce results comparable to those of CT, unfortunately, patients cannot undergo MRI due to difficult access in many emergency departments, as CT scans continue to be performed. Although CT scanning has saved many patients from unnecessary surgeries, it has also exposed many patients to unnecessary radiation. This situation is worrisome regarding the cancer risk that pediatric patients will face.[14,15] Therefore, the importance of scoring systems that combine multiple parameters has increased. Some of these are scoring systems such as the pediatric appendicitis score (PAS), the Alvarado score, and the Refined Low-Risk Appendicitis Score.[8,16,17]

In this study we aimed to determine the parameters most important for accurately diagnosing appendicitis in childhood.

METHODS

This study was planned as an observational, cross-sectional, retrospective study. The study was conducted in a tertiary pediatric emergency department between January 1, 2019, and December 31, 2021. The study data were accessed from the hospital automation system. The study was approved by the İzmir Katip Çelebi University Non-Interventional Clinical Research Ethics Committee (decision number: 0478, date: 26.10.2023). All patients who presented to the emergency department with complaints of abdominal pain, whose examinations were completed in the emergency department, and who were diagnosed with appendicitis, and operated on, were included in the study. Those who were administered analgesic drugs before admission or in the emergency department, those who had some tests performed in a different laboratory, those who were not operated on, those who could not undergo a histomorphologic examination of the specimen obtained after the operation, those who were younger than two years of age, those with abdominal pain due to non-abdominal causes, pregnant women, those who could not establish a verbal relationship due to mental retardation or other mental illnesses, those diagnosed with Familial Mediterranean Fever, those who were immunosuppressed, and those who had undergone abdominal surgery were excluded from the study. In the study, besides age and gender, the day and time of admission was obtained. Hemogram parameters, serum

C-reactive protein (CRP) level, and procalcitonin level were determined as acute phase reactants in each patient. The nephelometric method was used to measure CRP (BN™ II System Siemens, CRP reagent Ireland), and a level of 0.0 to 0.8 mg/dL was thought to be normal. A complete blood count was evaluated with an electronic cell counter (COULTER LH 780 Hematology Blood Analyzer from Beckman Coulter). PAS was calculated for each case using physical examination findings and laboratory parameters.[6] PAS includes eight parameters and was scored on a scale of 10 points (Table 1). According to this scoring system, 1-3 were considered negative, 4-7 were considered suspicious, and 8 and above were considered positive.[16] In addition, abdominal ultrasounds evaluation was performed in all patients. All of the patients were examined using a system called SonoAce X8 from Samsung Medison in Seoul, Korea. This system had a 3 to 7 MHz convex transducer and a 5 to 12 MHz linear transducer. According to the ultrasound, the appendix had "AA" if it had a diameter of more than 6 mm from front to back, fluid around the appendix, an image of low-contrast inflammation, fluid inside the appendix, and the ability to compress the appendix.[10,18] Abdominal CT examination was performed, additionally, in cases where the diagnosis could not be made with these evaluations. CT scanning was performed using a Siemens SOMATOM Definition Flash dual-source 128 multi-detector scanner (Siemens Medical Solutions, Forchheim, Germany): tube voltage, 100 kVp; tube current, 87 and 190 mA; slice thickness, 6 mm; and 40 mL Xenetix® (Guerbet, Gorinchem, The Netherlands) at a rate of 4 mL/s. Post-contrast scanning was performed 60 seconds after intravenous injection of Xenetix 350. A low-dose technique was applied, and the effective dose was calculated for each scan with a size-specific dose estimate. If the appendix could not be seen on the CT scan, its outer diameter was less than 6 mm, or peri-appendiceal strands were missing, it was thought that there was "no AA".[12] Postoperatively, the specimens were evaluated by blinded pathologists independent of the study group. The presence of polymorphonuclear leukocytes, lymphocytes, or plasma cells in appendicitis biopsy was considered histopathologically diagnostic for "AA". The absence or minimal presence of acute inflammatory cells or a normal appearance of the appendix was interpreted histopathologically as "no AA." The patients were divided into two groups as "those with appendicitis" and "those without appendicitis" according to the pathology results. We compared these two groups in terms of laboratory parameters, imaging findings, and PAS.

Statistical Analysis

We used SPSS (Statistical Package for Social Sciences version 17.0) while evaluating the study's findings. The Kolmogorov-Smirnov test was used to determine whether the distribution of continuous numerical variables was close to normal. Frequency indicators of numerical data were presented as mean ± standard deviation and median [interquartile range (IQR)], whereas discrete data, as percentages. A t-test was used to compare the relationships between continuous data. A chi-square test was used to compare discrete data, and Fisher's exact test was used when the chi-squared assumption (χ^2) was not met. In addition, the relationships of continuous data were analyzed by Pearson's correlation analysis. Logistic regression analysis was used to predict which cases would be histopathologically diagnosed as appendicitis. In addition, a receiver operating characteristic (ROC) curve was drawn to find the cut-off value of PAS. A p value less than 0.05 was considered significant.

RESULTS

A total of 153 patients were included in the study. Of these, 99 (64.7%) were male and 45 (29.41%) presented between 08:00 and 16:00. The median age was 13 years (IQR=11-15). We took a detailed history, physical examination, blood tests, abdominal ultrasound, and abdominal CT imaging. According to the results of these tests, the median white

Table 1. Pediatric appendicitis score and right lower quadrant criteria

Sign/symptoms	**Point**
Nausea/emesis	1
Anorexia	1
Migration of pain to RLQ	1
Low grade fever (>38.0°)	1
RLQ tenderness on light palpation	2
Cough/percussion/heel tapping tenderness at RLQ	2
Leukocytosis (>10,000/mm^3)	1
Left shift (>75% neutrophilia)	1
Total	**10**

RLQ: Right lower quadrant

blood cell (WBC) count was 14500/mm^3 (IQR=11600-19050), the median neutrophil count was 11800/mm^3 (IQR=8200-16650), the median serum CRP level was 19.8 mg/dL (IQR=4.5-75.6), and the median serum procalcitonin level was 0.09 mg/dL (IQR=0.02-0.53) (Table 2).

According to the abdominal ultrasound results, 68 (44.4%) patients had normal results. The mean PAS obtained from these data was 6.8±1.05 (minimum-maximum=3-9). Abdominal CT imaging was performed in 21 undiagnosed cases, and pathology was detected in 15 (71.42%). As a result, surgeries were performed on all cases. We took the specimen from the operation for pathologic examination and evaluated the results. Accordingly, histomorphologic findings in 137 (89.5%) cases were consistent with appendicitis (Table 3).

The comparison of age, sex, WBC, autonomic nervous system (ANS), CRP, and procolcitonin results of patients with and without appendicitis according to histomorphologic findings is shown in Table 2.

Of the 137 people whose appendicitis was confirmed by histomorphology, 83 (60.58%) had it also detected on abdominal ultrasound. In the abdominal ultrasound examinations of 16 patients, in whom appendicitis was not identified via histomorphology, appendicitis was found on the ultrasound in only 2 cases (12.5%). This difference was statistically significant (p=0.01, X^2=11.5) (Table 3). Based on these results, the PPV of abdominal ultrasound for the diagnosis of appendicitis was 0.60, and the negative predictive value (NPV) was 0.87 (Table 4).

Table 2. Distribution of age, gender, and laboratory data of cases by groups

	Total (n=153)	AA (+) (n=137)	AA (-) (n=16)	p
Age, median (IQR)	13 (11-15)	13 (11-15)	14.5 (11.25-17.00)	0.087
Sex, n (%)				
Male	99(64.7%)	90 (90.9%)	9 (9.1)	0.45
Female	54 (35.3%)	47 (87%)	7 (13)	
WBC (/mm^3) median (IQR)	14.5 (11.6-19.05)	14.8 (11.45-18.95)	14.1 (11.70-21.60)	0.87
ANC (/mm^3) median (IQR)	11.8 (8.2-16.65)	11.8 (8.25-16.3)	9.9 (7.4250-19.35)	0.53
CRP (mg/dL), median (IQR)	19.8 (4.5-75.6)	22 (4.35-76.4)	7.95 (7.20-21.725)	0.30
PRC (mg/dL), median (IQR)	0.09 (0.02-0.53)	0.08 (0.0125-1.28)	0.27 (0.23-0.275)	0.25

IQR: Interquartile range, AA: Acute appendicitis, WBC: White blood cell, ANC: Absolute neutrophil count, CRP: C-reactive protein, PCR: Procalcitonin (gender comparison X^2, t-test used for other comparisons)

Table 3. Distribution of imaging and scoring systems according to the presence or absence of appendicitis based on the histopathological result of the operation material of the cases

	Op AA (-) n=16 (10.5%)	Op AA (+) n=137 (89.5%)	Total	p	X^2
US AA (-)	14 (20.6%)	54 (79.4%)	68 (100%)	0.01	11.5
US AA (+)	2 (2.4%)	83 (97.6%)	85 (100%)		
CT AA (-)	4 (66.7%)	2 (33.3%)	6 (100%)	0.015	5.9
CT AA (+)	2 (13.3%)	13 (86.7%)	15 (100%)		
PAS <7	11 (21.15%)	41 (78.85%)	52 (100%)	0.02	0.6
PAS >7	5 (4.95%)	96 (95.05%)	101 (100%)		

Op: Operation, US: Ultrasound, CT: Computed tomography, AA: Acute appendicitis, PAS: Pediatric appendicitis score

Table 4. Validity test results of imaging and scoring systems at different cutoff points

	Sensitivity	Specificity	PPV	NPV
CT	0.86	0.97	0.86	0.66
US	0.97	0.20	0.60	0.87
PAS=5	0.94	0.31		
PAS=6	0.7	0.68		
PAS=7	0.95	0.21	0.70	0.68

PPV: Positive predictive value, NPV: Negative predictive value, US: Ultrasound, CT: Computed tomography, PAS: Pediatric appendicitis score

In the abdominal CT tests that were done on 15 patients who had histomorphologic signs of appendicitis, 13 (86.66%) showed signs of appendicitis. In the abdominal CT evaluations, appendicitis was detected in only 2 out of 6 patients (33.3%) in whom it was not initially detected by histomorphology. This difference was statistically significant (p=0.015, χ^2=5.9) (Table 3). Based on these results, the PPV of abdominal CT for the diagnosis of appendicitis was 0.86, and the NPV was 0.66 (Table 4).

Based on the histomorphology results, 96 (70.07%) of the 137 patients who were diagnosed with appendicitis had PAS 7 or higher. Based on the histomorphology results of 16 cases, no signs of appendicitis were found. PAS 7 or higher was found in only 5 cases, or 31.25%. This difference was statistically significant (p=0.02, χ^2=0.6) (Table 3). To determine if someone has appendicitis, the PPV for a PAS of seven or more was found to be 0.70, and the NPV was 0.68 (Table 4). No significant association was found when examining the effect of presentation during working hours or on-call hours on the histomorphologic diagnosis of appendicitis (p=0.18 and χ^2=1.7).

We used logistic regression modeling to determine which parameters changed the histomorphologic diagnosis of appendicitis. It was found that younger age increased the odds of appendicitis 4.29 times, higher WBC counts increased them 5.44 times, higher PAS increased them 7.87 times, and the use of diagnostic ultrasound increased them 7.91 times. Lab tests were compared to each other using PAS, and a weak positive correlation was found between WBC and PAS (r=0.43, p<0.01) and between ANS and PAS (r=0.42, p<0.01).

We investigated the relationship between the results of abdominal ultrasound and PAS. The mean PAS score of patients with a diagnostic abdominal ultrasound result was 6.94±0.83, while the mean PAS score of patients with a normal abdominal ultrasound result was 6.75±1.24 (p=0.26 and T=1.1).

We analyzed the relationship between the results of abdominal CT and the PAS score. The mean PAS score was 7.06±1.03 in patients with a diagnostic result on abdominal CT, while it was 6.0±0.89 in patients with a normal abdominal CT (p=0.03, T=2.2). Regarding the correlation between histomorphologic results of surgery and PAS, the mean PAS was 6.94±1.02 in cases with a diagnostic result for appendicitis, while the mean PAS was 6.12±1.02 in cases with normal results (p=0.003 and T=3).

As a result of our analysis , we found the sensitivity and specificity of abdominal CT to be 0.86 and 0.97, respectively, and the sensitivity and specificity of abdominal ultrasound to be 0.97 and 0.20, respectively. According to the ROC curve, we drew to determine the ideal cutoff of the PAS score, the sensitivity was 0.70, and the specificity was 0.68 when PAS=6. The area under the curve was 0.721 (intermediate level) (Figure 1).

DISCUSSION

The aim of this study was to determine how effective the investigations are when performed in patients presenting to the pediatric emergency department with acute abdominal pain , especially in the diagnosis of AA. We analyzed the role of ultrasound in the diagnosis of appendicitis. Ultrasound was able to confirm the diagnosis of appendicitis 97% of the time; but it wasn't always enough to make a final diagnosis because of its limitations. In such cases, a PAS value of 7 or higher has been found to have a diagnostic sensitivity of 70%. However, the combination of young age, high PAS score, high WBC count, and diagnostic ultrasound has been shown to be extremely likely to indicate appendicitis. These findings emphasize the importance of using multiple tests and parameters together in the diagnosis of appendicitis, especially in children presenting with acute abdominal pain. Such results may make important contributions to clinical practice, as the combined use of multiple diagnostic tools may increase the chances of making the correct diagnosis and help prevent unnecessary surgeries. However, it should be kept in mind that each test alone may not be sufficient, and the specific situation of each patient should be taken into account.

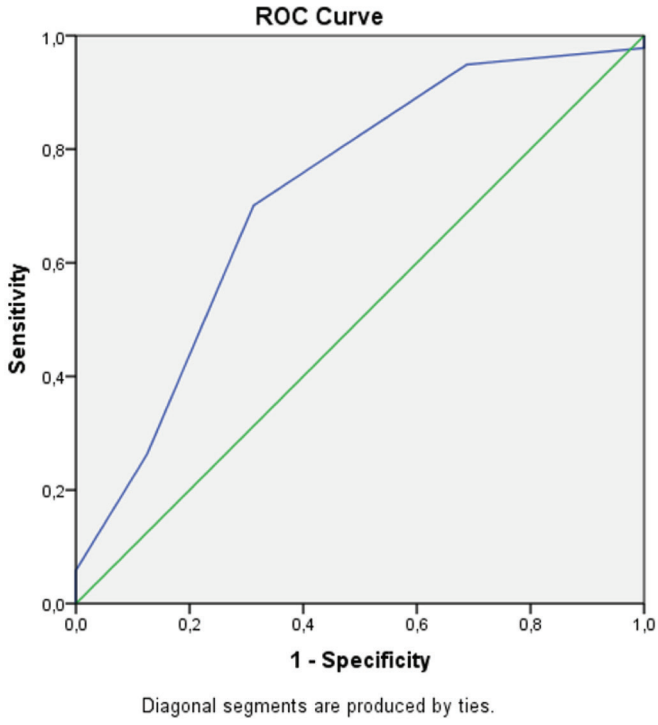


Figure 1. ROC curve for the pediatric appendicitis score
ROC: Receiver operating characteristic

In the study by Taşar et al.[6], the mean age of 220 patients diagnosed with AA was found to be 13 years. This finding is consistent with the studies in the literature on appendicitis in childhood, because these studies generally emphasize that appendicitis cases are more common in children in the second decade.[4,5] Similarly, in our study, the median age of the cases diagnosed with AA was found to be 13 years. This is in parallel with literature findings and suggests that appendicitis cases in childhood are more common in this age group.

In terms of gender, Sayed et al.[12] reported that 60% of the 140 cases diagnosed with AA were male. This finding is also compatible with the general trend in the literature because, in many studies, it has been reported that boys have 7-9 times higher risk of appendicitis than girls.[5] A similar result was obtained in our study. The number of male and female cases was found to be 99 and 54, respectively, and it was observed that boys were more common.

These findings support the idea that appendicitis cases are more common, especially in boys, and the average age is more concentrated in individuals in the second decade. Furthermore, this gender and age distribution suggests that they are important factors for the diagnosis and management of appendicitis in clinical practice.

In this study, we examined the validity of various laboratory tests and imaging methods to determine the diagnostic accuracy of AA. In a study conducted by Bal et al.[13] on 96 cases of AA, mean WBC, ANC, and CRP of the patients with histopathologic AA were 15.1, 12.1, and 14 mg/dL, respectively. In cases without histopathologic AA, the mean WBC was 11300, ANS was 8200, and CRP was 4.8 mg/dL. These findings indicate elevated acute phase reactants. Although the increase in these values is non-specific in inflammatory conditions such as AA, studies have found that inflammatory marker levels in such tests are significantly elevated.[19-24] Similarly, in our study, WBC, ANS, and CRP values were found to be significantly higher in the group with histopathologically positive results after appendectomy. This is an important finding that increases the accuracy of the diagnosis of AA.

The validity of ultrasound in the diagnosis of AA is also frequently discussed in the literature. On the other hand, a study by Taşar et al.[6] found that ultrasound had a 67.9% sensitivity, a 78.4% specificity, a 64.7% PPV, and an 80.7% NPV. The study by Bal et al.[13] also found that ultrasound had a sensitivity of 59%, a specificity of 75%, a PPV of 77%, and a NPV of 57%. Sayed et al.[12] found the sensitivity of ultrasound to be 55.6%, specificity to be 76.9%, PPV to be 76.9%, and NPV to be 68%. The literature reports that the sensitivity of ultrasound in diagnosing AA varies between 60% and 93.1%.[18,25-27] In our study, the sensitivity of ultrasound was found to be 97%, specificity was found to be 20%, PPV was found to be 60%, and NPV was found to be 87%. Although this high sensitivity indicates that ultrasound plays an important role in the diagnosis of appendicitis, the low specificity suggests that it may sometimes lead to false positive results.

CT, another imaging modality, has a very high sensitivity and specificity in the diagnosis of AA. Studies have shown that the sensitivity and specificity of CT vary between 94% and 100% and 93% and 100%, respectively.[14,28-32] However, the use of CT in children may lead to long-term health risks due to ionizing radiation. Therefore, it is emphasized that minimizing radiation use should be our focus. In our study, the sensitivity, specificity, positive predictive value, and NPV of CT were 86%, 97%, 86%, and 66%, respectively.

These results were found to be quite high, consistent with the literature. The findings of this study show that laboratory tests and imaging methods are important tools in the diagnosis of AA. In particular, high WBC, ANS, and CRP values are important parameters supporting the diagnosis of appendicitis. The tests, ultrasound with its high sensitivity and CT with its high accuracy, are important for provideing clinical decision support. However, considering the radiation risks of CT, it is concluded that alternative methods should be used to prevent its unnecessary use.

PAS is a practical, reproducible, and low-cost scoring system that can be easily applied in emergency departments. Many studies in the literature have shown that the PAS score is effectively used in the diagnosis of AA.[33,34]

The study by Schneider et al.[8] looked at 588 people who were suspected of having AA. They found that PAS ≥6 had 82% sensitivity and 65% specificity. Similarly, in two other studies in which PAS values of 7 and above were accepted, sensitivity ranged between 97.6% and 100%, and specificity ranged between 92% and 96%.[17,30] In the study by Bhatt et al.[16], PAS ≥8 values were considered significant, with selectivity for this value being 95.1 and PPV found to be 85.2%. In the study by Taşar et al.[6], sensitivity was calculated as 58.0%, specificity as 94.9%, PPV as 87.0%, and NPV as 79.5% in cases with a PAS above 8.6 In the study by Sayed et al.[12], when they took 5 as the cut-off point for PAS, sensitivity was 95%, specificity 84%, PPV 82%, and NPV 82%.

In our study, the sensitivity and specificity were 94% and 31% when the PAS score was 5 and above, 70.1% and 68% when PAS ≥6, 95% and 21% when PAS ≥7. These results are consistent with other studies in the literature and show that the sensitivity and specificity of the PAS score vary at different cut-off points. With its high sensitivity, the PAS score may be effective as a clinical decision support tool at an early stage in the diagnosis of appendicitis. However, the specificity of the PAS score may vary significantly depending

on the cut-off point used. Therefore, the PAS score alone should not be used to confirm the diagnosis and should be evaluated in conjunction with other diagnostic tests (e.g., ultrasound). In our study, PAS scores of 7 or higher had higher sensitivity and lower specificity than lower scores. This suggests that a certain threshold value should be used for the PAS score in diagnosing appendicitis. Generally, the PAS score may be an important adjunctive tool in the diagnosis of AA, but it is recommended to integrate this score into an algorithm that evaluates it along with other clinical findings, before incorporating it into clinical practice.

In the study by Bal et al.[13], as in our study, a logistic regression analysis was performed based on the positive histopathological result of appendectomy. In their study, they concluded that a diagnostic ultrasound and a PAS greater than 7 were significant. Similarly, in our study, ultrasound, PAS, and even higher WBC count, and younger age increased the likelihood of diagnosis.

Study Limitations

The retrospective nature of this study inevitably introduces the risk of selection bias and data loss. Although all eligible patients were included based on explicitly defined inclusion criteria, the accuracy and completeness of the data relied heavily on the precision of entries in the hospital information system. Moreover, US outcomes may have varied depending on the operator's level of expertise, a factor not controlled for in the study design.

Finally, while the PAS was assessed, it was not applied in a standardized manner across all clinicians, which may have contributed to inter-observer variability in scoring.

Future multicenter, prospectively designed studies are warranted to enhance the validity and generalizability of the findings.

CONCLUSION

We have shown that the presence of a diagnostic ultrasound in combination with high PAS is useful in the diagnosis of AA. The contribution of this study to the literature is to draw attention to the increased frequency of performing abdominal CT in recent years. The protection of children from ionizing radiation should be the primary duty of all health professionals working in this field. We believe that the design of this prospective study with a radiation-free examination, such as MRI instead of CT, will fill a gap in the literature, if the appropriate funding is provided.

Ethics

Ethics Committee Approval: The study was approved by the İzmir Katip Çelebi University Non-Interventional Clinical Research Ethics Committee (decision number: 0478, date: 26.10.2023).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: E.B.Ç.K., Concept: T.N., Y.B., E.B.Ç.K., Design: T.N., G.G., E.E., Data Collection or Processing: T.N., Y.B., E.B.Ç.K., Analysis or Interpretation: G.G., E.E., Literature Search: G.G., Y.B., E.E., Writing: T.N.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Kharbanda AB, Stevenson MD, Macias CG, et al. Interrater reliability of clinical findings in children with possible appendicitis. Pediatrics. 2012;129:695-700.
2. Escribá A, Gamell AM, Fernández Y, Quintillá JM, Cubells CL. Prospective validation of two systems of classification for the diagnosis of acute appendicitis. Pediatr Emerg Care. 2011;27:165-9.
3. Myers AL, Williams RF, Giles K, et al. Hospital cost analysis of a prospective, randomized trial of early vs interval appendectomy for perforated appendicitis in children. J Am Coll Surg. 2012;214:427-34.
4. Graham JM, Pokorny WJ, Harberg FJ. Acute appendicitis in preschool age children. Am J Surg. 1980;139:247-50.
5. Lee JH, Park YS, Choi JS. The epidemiology of appendicitis and appendectomy in South Korea: National Registry data. J Epidemiol. 2010;20:97-105.
6. Taşar S, Taşar MA, Ayyıldız NK, Güder L, Arıkan Fİ, Dallar YB. The importance of the pediatric appendicitis score and ultrasonographic findings for the diagnosis of acute appendicitis in the pediatric emergency department. Turkish J Pediatr Dis. 2015;9:184-8.
7. Blakely ML, Williams R, Dassinger MS, et al. Early vs interval appendectomy for children with perforated appendicitis. Arch Surg. 2011;146:660-5.
8. Schneider C, Kharbanda A, Bachur R. Evaluating appendicitis scoring systems using a prospective pediatric cohort. Ann Emerg Med. 2007;49:778-84.
9. Shera AH, Nizami FA, Malik AA, Naikoo ZA, Wani MA. Clinical scoring system for diagnosis of acute appendicitis in children. Indian J Pediatr. 2011;78:287-90.
10. Palabıyık F, Kayhan A, Cimilli T, Toksoy N, Bayramoğlu S, Aksoy S. The comparison of plain film and ultrasound findings of appendicitis in children. Marmara Medical Journal. 2008;21:203-9.
11. Goldin AB, Khanna P, Thapa M, McBroom JA, Garrison MM, Parisi MT. Revised ultrasound criteria for appendicitis in children improve diagnostic accuracy. Pediatr Radiol. 2011;41:993-9.
12. Sayed AO, Zeidan NS, Fahmy DM, Ibrahim HA. Diagnostic reliability of pediatric appendicitis score, ultrasound and low-dose computed tomography scan in children with suspected acute appendicitis. Ther Clin Risk Manag. 2017;13:847-54.

13. Bal A, Anil M, Nartürk M, et al. Importance of clinical decision making by experienced pediatric surgeons when children are suspected of having acute appendicitis: the reality in a high-volume pediatric emergency department. Pediatr Emerg Care. 2017;33:38-42.
14. Doria AS, Moineddin R, Kellenberger CJ, et al. US or CT for diagnosis of appendicitis in children and adults? A meta-analysis. Radiology. 2006;241:83-94.
15. Brenner DJ, Hall EJ. Computed tomography--an increasing source of radiation exposure. N Engl J Med. 2007;357:2277-84.
16. Bhatt M, Joseph L, Ducharme FM, Dougherty G, McGillivray D. Prospective validation of the pediatric appendicitis score in a Canadian pediatric emergency department. Acad Emerg Med. 2009;16:591-6.
17. Goldman RD, Carter S, Stephens D, Antoon R, Mounstephen W, Langer JC. Prospective validation of the pediatric appendicitis score. J Pediatr. 2008;153:278-82.
18. Erbay G, Karadeli E, Koç Z. The role of ultrasound and laboratory findings for diagnosis of appendicitis in pediatric patients. Cukurova Med J. 2012;37:84-9.
19. Rothrock SG, Pagane J. Acute appendicitis in children: emergency department diagnosis and management. Ann Emerg Med. 2000;36:39-51.
20. Bundy DG, Byerley JS, Liles EA, Perrin EM, Katznelson J, Rice HE. Does this child have appendicitis? JAMA. 2007;298:438-51.
21. Kwok MY, Kim MK, Gorelick MH. Evidence-based approach to the diagnosis of appendicitis in children. Pediatr Emerg Care. 2004;20:690-8.
22. Sack U, Biereder B, Elouahidi T, Bauer K, Keller T, Tröbs RB. Diagnostic value of blood inflammatory markers for detection of acute appendicitis in children. BMC Surg. 2006;6:15.
23. Williams R, Mackway-Jones K. Towards evidence based emergency medicine: best BETs from the Manchester Royal Infirmary. White cell count and diagnosing appendicitis in children. Emerg Med J. 2002;19:428-9.
24. Fawkner-Corbett D, Hayward G, Alkhmees M, Bruel AVD, Ordóñez-Mena JM, Holtman GA. Diagnostic accuracy of blood tests of inflammation in paediatric appendicitis: a systematic review and meta-analysis. BMJ Open. 2022;12:e056854.
25. Gökçe AH, Aren A, Gökçe FS, Dursun N, Barut AY. Reliability of ultrasonography for diagnosing acute appendicitis. Turk J Trauma Emerg Surg. 2011;17:19-22.
26. İnan M, Tülay SH, Besim H, Karakaya J. The value of ultrasonography and its' comparison with Alvarado scoring system in acute appendicitis. Turk J Surg. 2011;27:149-53.
27. Toprak H, Kilincaslan H, Ahmad IC, et al. Integration of ultrasound findings with Alvarado score in children with suspected appendicitis. Pediatr Int. 2014;56:95-9.
28. Peña BMG, Mandl KD, Kraus SJ, et al. Ultrasonography and limited computed tomography in the diagnosis and management of appendicitis in children. JAMA. 1999;282:1041-6.
29. Bachur RG, Dayan PS, Bajaj L, et al. The effect of abdominal pain duration on the accuracy of diagnostic imaging for pediatric appendicitis. Ann Emerg Med. 2012;60:582-90.
30. Nordin AB, Sales S, Nielsen JW, Adler B, Bates DG, Kenney B. Standardized ultrasound templates for diagnosing appendicitis reduce annual imaging costs. J Surg Res. 2018;221:77-83.
31. Lowe LH, Penney MW, Stein SM, et al. Unenhanced limited CT of the abdomen in the diagnosis of appendicitis in children. Am J Roentgenol. 2001;176:31-5.
32. Sivit CJ, Applegate KE, Stallion A, et al. Imaging evaluation of suspected appendicitis in a pediatric population. Am J Roentgenol. 2000;175:977-80.
33. Samuel M. Pediatric appendicitis score. J Pediatr Surg. 2002;37:877-81.
34. Demir Ş, Mert M, Yasin YK, Kahya MO, Demirtaş O. Importance of pediatric appendicitis scoring system and ultrasonography in the diagnosis of acute appendicitis in children. Forbes J Med. 2023;4:259-64.

DOI: 10.4274/forbes.galenos.2025.57855
Forbes J Med 2025;6(2):127-37

Original Article / Orijinal Araştırma

Artificial Intelligence-Based Prediction of Bloodstream Infections Using Standard Hematological and Biochemical Markers

Standart Hematolojik ve Biyokimyasal Belirteçler Kullanılarak Kan Dolaşımı Enfeksiyonlarının Yapay Zeka Tabanlı Tahmini

Ferhat DEMİRCİ[1], Murat AKŞİT[1], Aylin DEMİRCİ[2]

[1]University of Health Sciences Türkiye, İzmir Tepecik Training and Research Hospital, Clinic of Medical Biochemistry, İzmir, Türkiye
[2]University of Health Sciences Türkiye, İzmir Tepecik Training and Research Hospital, Clinic of Family Medicine, İzmir, Türkiye

Cite as: Demirci F, Akşit M, Demirci A. Artificial intelligence-based prediction of bloodstream infections using standard hematological and biochemical markers. Forbes J Med. 2025;6(2):127-37

ABSTRACT

Objective: Bloodstream infections (BSIs) require rapid identification to initiate timely antimicrobial therapy, yet blood culture-the current diagnostic gold standard-suffers from delayed results and limited sensitivity. This study aimed to develop an interpretable machine learning (ML) model using routine laboratory parameters to predict blood culture positivity.

Methods: A total of 1,972 adult patients who underwent complete blood count, C-reactive protein, procalcitonin (PCT), and blood culture testing at a tertiary hospital were retrospectively included. Three models-random forest, H_2O automated ML, and an ensemble model-were developed and evaluated using standard classification metrics [area under the curve (AUC)-receiver operating characteristic (ROC), sensitivity, specificity, F1 score]. SHapley Additive exPlanations (SHAP) analysis was employed to enhance interpretability.

Results: The ensemble model yielded the best performance, achieving an AUC-ROC of 0.95, sensitivity of 0.78, specificity of 0.97, and F1 score of 0.84. External validation on an independent cohort confirmed the model's generalizability (AUC-ROC: 0.85). SHAP analysis revealed that age and PCT were the most influential features with both statistical and clinical relevance. Basophil count, while ranked highest by SHAP, showed low sensitivity, highlighting the difference between algorithmic weight and bedside utility.

Conclusion: These findings support the integration of routine, readily available laboratory data into an explainable AI framework to accurately predict culture positivity. The model's strong performance and interpretability suggest its potential application in clinical decision support systems to improve diagnostic stewardship, reduce unnecessary cultures, and optimize resource use in suspected BSI cases.

Keywords: Sepsis, blood culture, machine learning, procalcitonin, C-reactive protein

Received/Geliş: 12.04.2025
Accepted/Kabul: 06.05.2025
Epub: 17.07.2027
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Corresponding Author/ Sorumlu Yazar:
Ferhat DEMİRCİ, MD,
University of Health Sciences Türkiye, İzmir Tepecik Training and Research Hospital, Clinic of Medical Biochemistry, İzmir, Türkiye
drdemirci05@gmail.com
ORCID: 0000-0002-5999-3399

ÖZ

Amaç: Kan dolaşımı enfeksiyonlarında (KDE) erken tanı, zamanında antimikrobiyal tedavi başlatılması açısından kritik öneme sahiptir. Ancak, mevcut altın standart tanı yöntemi olan kan kültürü, gecikmeli sonuç vermesi ve düşük pozitiflik oranı nedeniyle sınırlıdır. Bu çalışmanın amacı, rutin laboratuvar verilerini kullanarak kan kültürü pozitifliğini öngörebilecek yorumlanabilir bir makine öğrenimi (ML) modeli geliştirmektir.

Yöntem: Üçüncü basamak bir hastanede tam kan sayımı, C-reaktif protein, prokalsitonin (PCT) ve kan kültürü testi yapılan toplam 1.972 yetişkin hasta retrospektif olarak çalışmaya dahil edilmiştir. Rastgele orman, H_2O otomatik ML ve bir ensemble (birleşik) model olmak üzere üç farklı model geliştirilmiş ve

AUC-ROC, duyarlılık, özgüllük ve F1 skoru gibi sınıflandırma ölçütleriyle değerlendirilmiştir. Modelin yorumlanabilirliğini artırmak amacıyla SHapley Additive exPlanations (SHAP) analizi uygulanmıştır.

Bulgular: Ensemble model en iyi performansı göstermiş; [alıcı işletim karakteristiği eğrisi (AUC)-eğri altındaki alan (ROC)]: 0,95, duyarlılık: 0,78, özgüllük: 0,97 ve F1 skoru: 0,84 olarak bulunmuştur. Bağımsız bir doğrulama veri seti üzerinde yapılan analiz, modelin genellenebilirliğini doğrulamıştır (AUC-ROC: 0,85). SHAP analizine göre yaş ve PCT hem istatistiksel hem de klinik açıdan en etkili değişkenler olarak öne çıkmıştır. Basofil sayısı ise algoritmik olarak yüksek önem taşımasına rağmen düşük duyarlılığı nedeniyle klinik faydası sınırlı bulunmuştur.

Sonuç: Bu sonuçlar, rutin laboratuvar verilerinin açıklanabilir yapay zeka çerçevesinde kullanılarak kan kültürü pozitifliğinin yüksek doğrulukla öngörülebileceğini göstermektedir. Modelin güçlü performansı ve yorumlanabilirliği, tanı yönetimini iyileştirmek, gereksiz kültürleri azaltmak ve şüpheli KDE olgularında kaynak kullanımını optimize etmek için klinik karar destek sistemlerinde potansiyel bir uygulama olduğunu göstermektedir.

Anahtar Kelimeler: Sepsis, kan kültürü, makine öğrenimi, prokalsitonin, C-reaktif protein

INTRODUCTION

Bloodstream infections and sepsis remain leading causes of morbidity and mortality, especially in critically ill and immunocompromised patients. Early identification of bacteremia is essential to initiate timely antimicrobial therapy, which can significantly reduce adverse outcomes. However, blood culture-the current gold standard diagnostic method-is limited by low positivity rates and delayed results, often requiring 24-72 hours.[1,2]

To bridge this diagnostic delay, clinicians frequently rely on nonspecific biomarkers such as complete blood count (CBC), C-reactive protein (CRP), and procalcitonin (PCT). While these markers offer some insight, their standalone predictive value remains suboptimal. Studies have shown that PCT outperforms CRP in specificity for bacterial infections, but both lack adequate sensitivity to reliably predict positive cultures.[3,4] Moreover, traditional clinical assessment is often inaccurate and inconsistent in estimating bacteremia risk.[5]

Recent developments in artificial intelligence (AI) and machine learning (ML) have opened new avenues for early bacteremia prediction using routine clinical data. Several studies have demonstrated that ML algorithms can improve predictive accuracy by combining laboratory values, vital signs, and demographic information.[6,7] These models have yielded area under the curve (AUC) values of up to 0.84, indicating high potential in differentiating between true infections and false alarms.[8]

However, many existing models are limited by dataset specificity, exclusion of key demographic factors (such as age and sex), and lack of explainability, which hinders clinical adoption.[9] Additionally, hematological markers like neutrophil-to-lymphocyte ratio, band count, and platelet levels (which are routinely available and cost-effective) are often underutilized in current models, despite evidence supporting their role in predicting bacteremia.[10]

The present study aims to address these limitations by developing a ML-based model that integrates hemogram parameters, CRP, PCT, age, and gender to predict blood culture positivity. This approach not only leverages data readily available at the point of care but also contributes to diagnostic stewardship by reducing unnecessary testing and improving the timing of antimicrobial interventions.

METHODS

Study Population/Subjects

This study was conducted at University of Health Sciences Türkiye, İzmir Tepecik Training and Research Hospital. Patients who presented to this center and its affiliated hospital (AH) between January 1, 2024, and March 31, 2025, and underwent first-time blood culture, CBC, PCT, and CRP tests were included. The baseline characteristics of the study population are shown in Table 1. Patients with incomplete test results, sub-parameters missing, or contaminating agents detected in blood cultures were excluded.

Hemogram samples were analyzed in both hospitals using Sysmex XN-1000 (Kobe, Japan) hematology analyzers; CRP tests were analyzed using Beckman Coulter AU-5800 in the main hospital and Beckman Coulter AU680 (California, USA) in the AH; and PCT tests were analyzed using Siemens Advia Centaur XPT (chemiluminescence immune assay, Erlangen, Germany) at the main hospital and Beckman Coulter DXI-800 (chemiluminescence immune assay, California, USA) at the AH.

Venous blood samples were collected under aseptic conditions into automated blood culture bottles from the Biomerieux BacT/Alert 3D (France) brand. Translated with DeepL.com (free version). The bottles were placed in the corresponding brand-specific incubator system for continuous monitoring. Bottles that flagged positive for microbial growth were subcultured onto appropriate culture media. Following incubation, colony morphology was assessed, and species identification was performed by a clinical microbiologist using Gram staining, biochemical assays, and/or automated identification systems.

The reagents and calibrators were provided by Sysmex for hemogram analyses and by Beckman Coulter for CRP and PCT measurements. Using the respective automated analyzers, all analyses were carried out in accordance with

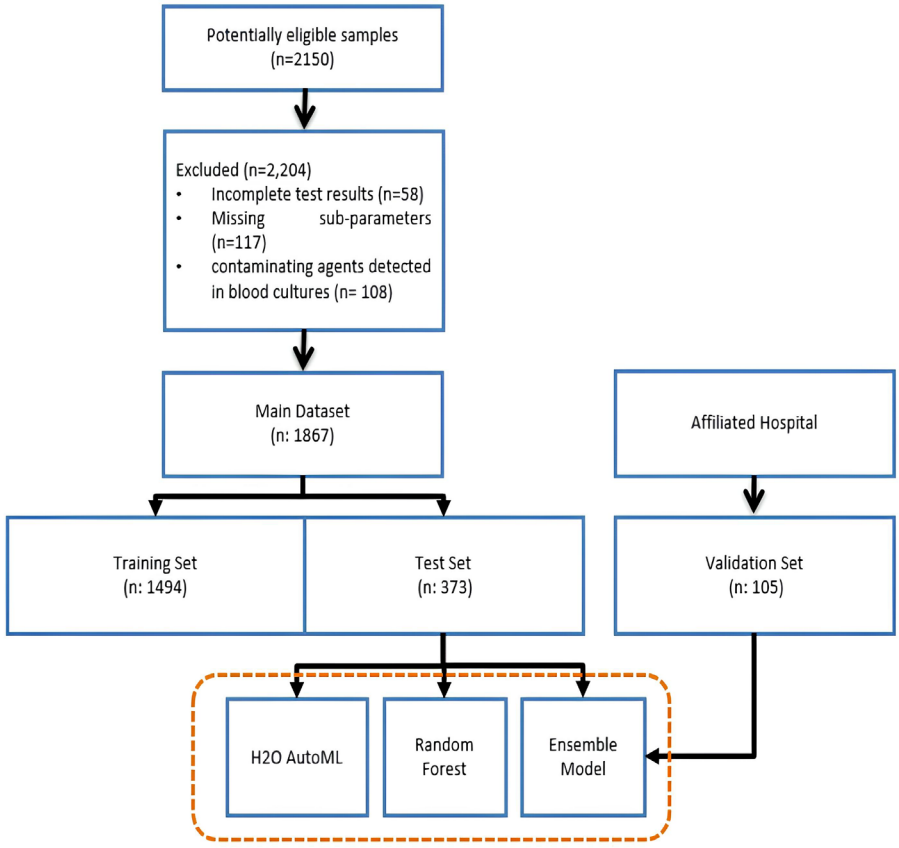


Figure 1. The Standards for Reporting Diagnostic Accuracy diagram

AutoML: Automated machine learning

the manufacturers' instructions. Routine maintenance and calibration procedures were performed regularly to ensure analytical accuracy and reliability.

Internal quality control for the hemogram was ensured using materials supplied by the manufacturer (Sysmex), whereas quality control for CRP and PCT assays was performed using materials obtained from Bio-Rad (California, USA).

Study Design

Before starting this retrospective study, ethical approval was obtained from the Ethics Committee of University of Health Sciences Türkiye, İzmir Tepecik Training and Research Hospital (decision number: 2025/03-27, date: 10.04.2025). Patient identity information was anonymized, and a dataset containing age, sex, CRP, PCT, CBC, and blood culture results from 2345 patients (2150 from the main building and 195 from the AH) was transferred to Microsoft Excel 2021 (USA).

After applying exclusion criteria, the final dataset included 1972 patients (1867 from the main hospital and 105 from the AH). This dataset was then transferred to Python software (version 3.11, USA) for ML analysis.

Following data cleaning, the dataset was randomly partitioned into training and testing sets in an 80:20 ratio using stratified sampling based on the binary outcome variable, ensuring preservation of the original class distribution. The Standards for Reporting Diagnostic Accuracy diagram illustrating the patient flow throughout the study is presented in Figure 1.

Data Preprocessing and Training of Machine Learning Algorithms

For data preprocessing, patient results were transferred to Microsoft Excel. Cases with missing values were excluded from the dataset. Bacterial culture results were evaluated and converted into a binary classification. Patient samples in which bacterial species were identified by a clinical microbiologist were classified as "growth present = 1," whereas samples with no growth were classified as "no growth = 0." Samples reported as contamination were excluded from the study. Additionally, sex was encoded as

a binary variable, with male = 0 and female = 1. The cleaned dataset was transferred to Python for ML analysis.

The models were trained using the 15 most important predictive parameters, which included:

- **Demographic variables:** Age, sex
- **Biochemical variables:** PCT and CRP
- **Hematologic variables:** Hemogram leucocyte variables (total white blood cell, neutrophil, lymphocyte, monocyte, eosinophile, basophile count) and hemoglobin.

Following model training, performance evaluation was conducted using the test and validation datasets.

The cleaned dataset was imported into the Python programming environment for ML analysis. Model development was conducted using Python within the PyCharm integrated development environment (IDE). PyCharm is a widely adopted and robust IDE for Python, offering advanced functionalities such as intelligent code completion, comprehensive debugging tools, and integrated testing frameworks. These features facilitate efficient data preprocessing, model training, and algorithm optimization, while providing seamless integration with widely used ML libraries, including scikit-learn, thereby supporting streamlined and scalable project workflows.[11]

A total of three AI ML algorithms were evaluated in this study: random forest (RF), H_2O automated ML (AutoML) (version 3.46), and an ensemble ML method. Model development was carried out in a Python 3.11 environment using H_2O AutoML.[12] To overcome the limitations inherent in manual model development-particularly when the primary expertise of the user is not in data science-AutoML tools have emerged as a practical solution. AutoML tools automate key steps such as feature engineering, model building, and hyperparameter optimization, which traditionally require extensive domain expertise. Despite the clear advantages and the growing interest in ML applications, few studies have applied AutoML tools within the clinical laboratory context.[13] The best-performing model within the H_2O AutoML framework was selected based on AUC-receiver operating characteristic (ROC) and logloss values.

Computational Environment and Libraries

In the development of classification models using ML and deep learning techniques, a variety of open-source Python libraries were employed for data preprocessing, model training, evaluation, and visualization. All procedures were conducted within the Python 3.11 programming environment. The libraries utilized are categorized as follows:

Data Processing and Analysis

- Pandas (v1.5): For creating and manipulating data frames
- Numpy (v1.23): For numerical operations and vectorized calculations

ML Model Development

- Scikit-learn (v1.2): For implementing ML algorithms and performance evaluation
- H_2O (.frame, .model) (v3.46.0.6): H_2OAutoML

Model Evaluation and Visualization

- Matplotlib (v3.6): For data visualization and plotting
- Sklearn (.metrics, .ensemble) (v1.2): For performance metrics such as confusion matrix, ROC-AUC, and precision-recall (PR)-AUC
- Shap (v0.47): For SHapley Additive exPlanations (SHAP) analysis and feature importance visualization

Following model training, performance evaluation was conducted using the designated test dataset.

Performance Evaluation

Scikit-learn, Pandas, NumPy, SciPy, StatsModels, and Matplotlib/Seaborn-among Python's most robust libraries for ML and statistical analysis-were employed in this project. The modeling process underwent a comprehensive evaluation, including hyperparameter tuning and model selection through internal cross-validation. Model performance was assessed using multiple evaluation metrics. The following criteria were used for classification:

1. Classification performance metrics

- AUC-ROC
- AUC-PR
- Confusion matrix analysis
- Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), positive likelihood ratio (PLR), negative likelihood ratio (NLR) F1 score, odds ratio

2. Model interpretability metrics

- Feature importance analysis
- SHAP graphs

3. Validation results of the predictive models were analyzed to ensure a comprehensive assessment. This structured and multifaceted evaluation approach provides a robust framework for predicting treatment modality outcomes based on laboratory-derived data.

RESULTS

Dataset Description and Data Pre-Processing

The dataset used in this study included a total of 1,972 records, consisting of 1,494 entries in the training set, 373 in the test set, and an additional 105 records in the validation set. All datasets contained hemogram parameters alongside demographic data, allowing for comprehensive baseline characterization.

Baseline demographic characteristics of the study population are presented in Table 1. The mean age was 46.08±30.13 years in the training set, 44.47±30.51 years in the test set, and significantly higher in the validation set at 65.68±16 years (p<0.001). When stratified by sex, no significant differences were observed in mean age between male and female participants within each subset (all p>0.05). The reason the mean age was significantly higher in the AH compared to the main building is that the data here were obtained from patients mainly hospitalized in the palliative care ward.

Regarding sex distribution, males comprised 53.4% of the training set, 52.3% of the test set, and 61% of the validation set, while females made up 46.6%, 47.7%, and 39%, respectively. These differences were not statistically significant (p=0.277), suggesting a relatively balanced gender distribution across the subsets.

Descriptive statistics for hemogram and related variables are presented in Table 2. Among all measured biomarkers, basophil count (BASO) were the only variable showing a statistically significant difference between the datasets (p<0.001). Other parameters, including white blood cell count, neutrophils, lymphocytes, monocytes, eosinophils, hemoglobin, CRP, and PCT, did not show significant variation across the training, test, and validation cohorts (all p>0.05). This indicates general homogeneity in these biomarkers across subsets, enhancing the comparability of model training and validation.

"The performance of RF, H_2O AutoML, and ensemble models was comparatively evaluated based on their predictive capabilities, classification metrics, and interpretability. Classification metrics such as F1 score, sensitivity, specificity, and AUC-ROC were used to assess the models' ability to discriminate between classes."

Table 1. The baseline characteristics of the study population

Characteristics	Train set (n=1494) Value±SD	Test set (n=373) Value±SD	Validation set (n=105) Value±SD	p value
Age (years)	46.08±30.13	44.47±30.51	65.68±16	<0.001
Male	46.03±30.07	43.05±31.81	65.69±15.51	
Female	46.15±30.22	46.02±29.03	65.66±16.95	
Sex				0.277
Male	798 (53.4%)	195 (52.3)	64 (61%)	
Female	696 (46.6%)	178 (47.7%)	41 (39%)	
Blood culture positivity rate	363 (24.3 %)	91 (24.4 %)	36 (34.29 %)	

SD: Standart deviation

Table 2. Descriptive statistics of the hemogram and related variables

Variable	Unit	Train set Value±SD	Test set Value±SD	Validation set Value±SD	p value
White blood cell	($\times10^9$/L)	13.34±15.14	12.63±7.94	10.28±5.39	0.075
Neutrophil	($\times10^9$/L)	8.9±6.6	8.99±6.79	8.06±5.15	0.420
Lymphocyte	($\times10^9$/L)	3.35±12.18	2.74±3.69	1.30±0.76	0.154
Monocyte	($\times10^9$/L)	1.09±2.17	1.09±0.77	0.77±0.42	0.257
Eosinophil	($\times10^9$/L)	0.28±0.68	0.27±0.32	0.13±0.22	0.060
Basophil	($\times10^9$/L)	0.12±0.18	0.11±0.09	0.04±0.03	**<0.001**
Hemoglobin	(g/dL)	10.97±2.6	11.07±2.59	10.64±2.28	0.327
C-reactive protein	(mg/L)	78.23±91.44	81.97±95.94	94.03±78.09	0.217
Procalcitonin	(ng/mL)	3.48±14.05	3.36±10.68	3.76±9.14	0.794

SD: Standart deviation

Comparison of Classification Performance Metrics

The analysis began with the RF model prior to evaluating other models. To determine the optimal classification threshold, multiple cut-off values (0.3, 0.5, and 0.7) were evaluated based on their corresponding F1 scores. Among these, a threshold of 0.3 yielded the highest F1 score, indicating a better balance between precision and recall. Consequently, the analysis proceeded using this threshold for subsequent model evaluation.

Following the initial modeling phase, the H_2O AutoML framework was employed to systematically explore a wide range of algorithms and hyperparameter configurations. Among the candidate models generated, a gradient boosting machine (GBM) emerged as the most performing, striking an optimal trade-off between discrimination and calibration metrics-specifically AUC-ROC and log loss. The selected model (ID: GBM_grid_1_AutoML_9_20250326_201624_model_8) achieved an AUC-ROC of 0.942 and a log loss of 0.283, indicating both high classification accuracy and well-calibrated probabilistic outputs.

In the current dataset, the RF model demonstrated superior sensitivity, whereas the H_2O AutoML framework yielded higher specificity. Given the complementary strengths of these models, an ensemble approach combining both was hypothesized to offer enhanced overall performance. Accordingly, further performance analyses were conducted using the ensemble model.

Table 3 presents a comprehensive comparison of the three models-RF (threshold=0.30), H_2O AutoML, and the ensemble model-based on various diagnostic and predictive performance metrics. Among these, the Ensemble Model demonstrated the most balanced and robust performance across nearly all evaluated criteria. In terms of sensitivity, the RF model achieved the highest value (0.80), indicating its superior ability to correctly identify positive cases. However, the H_2O AutoML model excelled in specificity (0.98) and PPV (0.90), highlighting its strength in correctly identifying negative cases and reducing false positives. The ensemble model, which was developed by combining the strengths of the two approaches, achieved a high sensitivity (0.78) close to RF, and a specificity (0.97) comparable to AutoML, reflecting its effectiveness in maintaining a strong trade-off among both metrics.

Furthermore, the ensemble model yielded the highest odds ratio (121.59) and F1 score (0.84) among the three, indicating a superior overall discriminatory power and a well-balanced PR relationship. Its PLR of 27.50 and NLR of 0.23 also suggest a high diagnostic utility. These results support the rationale for using an ensemble strategy, as it effectively leverages the complementary advantages of the individual models.

Figure 2a illustrates the ROC and PR curves of the three ML models evaluated in this study: RF, H_2O AutoML, and the ensemble model. All models demonstrated excellent discriminative performance, with identical AUC-ROC and AUC-PR values of 0.95/0.89, indicating a strong ability to

Table 3. Diagnostic and predictive performance metrics for machine learning models

	Random forest (Th=0.30)	H_2O automated ML	Ensemble model	Validation set
Sensitivity	0.80 (0.71-0.87)	0.63 (0.52-0.72)	0.78 (0.68-0.85)	0.78 (0.68-0.85)
Specificity	0.94 (0.91-0.96)	0.98 (0.95-0.99)	0.97 (0.95-0.99)	0.93 (0.90-0.96)
Positive predictive value	0.81 (0.73-0.89)	0.90 (0.83-0.97)	0.90 (0.83-0.97)	0.85 (0.78-0.92)
Negative predictive value	0.94 (0.90-0.96)	0.89 (0.86-0.93)	0.93 (0.90-0.969	0.89 (0.86-0.92)
Positive likelihood ratio	13.31 (8.30-21.33)	29.44 (13.13-66.00)	27.50 (13.77-54.92)	10.73 (5.64-18.95)
Negative likelihood ratio	0.21 (0.14-0.32)	0.38 (0.29-0.50)	0.23 (0.15-0.33)	0.24 (0.15-0.33)
Odds ratio	63.22 (31.02-128.80)	77.12 (30.93-192.26)	121.59 (51.42-287.46)	44.8 (21.97-93.01)
F1 score	0.81 (0.73-0.87)	0.74 (0.66-0.81)	0.84 (0.77-0.89)	0.81 (0.74-0.86)
Matthews correlation coefficient	0.745 (0.660-0.822)	0.694 (0.603-0.775)	0.790 (0.714-0.861)	0.721 (0.669-0.787)

Th: Threshold, ML: Machine learning

distinguish between positive and negative cases. Despite the equal AUC values, the slight variations in curve shapes across models reflect differences in threshold behavior and confidence calibration.

Figure 2b provides the confusion matrices for the respective models, further highlighting their classification behaviors.

The RF model (threshold=0.30) achieved a relatively higher sensitivity, correctly identifying 73 out of 91 actual positive cases, albeit at the expense of slightly more false positives (17 cases). In contrast, the H_2O AutoML model prioritized specificity, yielding only 6 false positives while missing true positives (34 false negatives). The ensemble model balanced these trade-offs effectively, reducing false negatives compared to AutoML (20 cases) while maintaining a high specificity (274 true negatives).

These findings confirm that although overall discriminative capacity was similar across models (as reflected in the AUC metrics), the ensemble model provided the most favorable balance between sensitivity and specificity-an important consideration in scenarios requiring both reliable detection and minimization of false alarms.

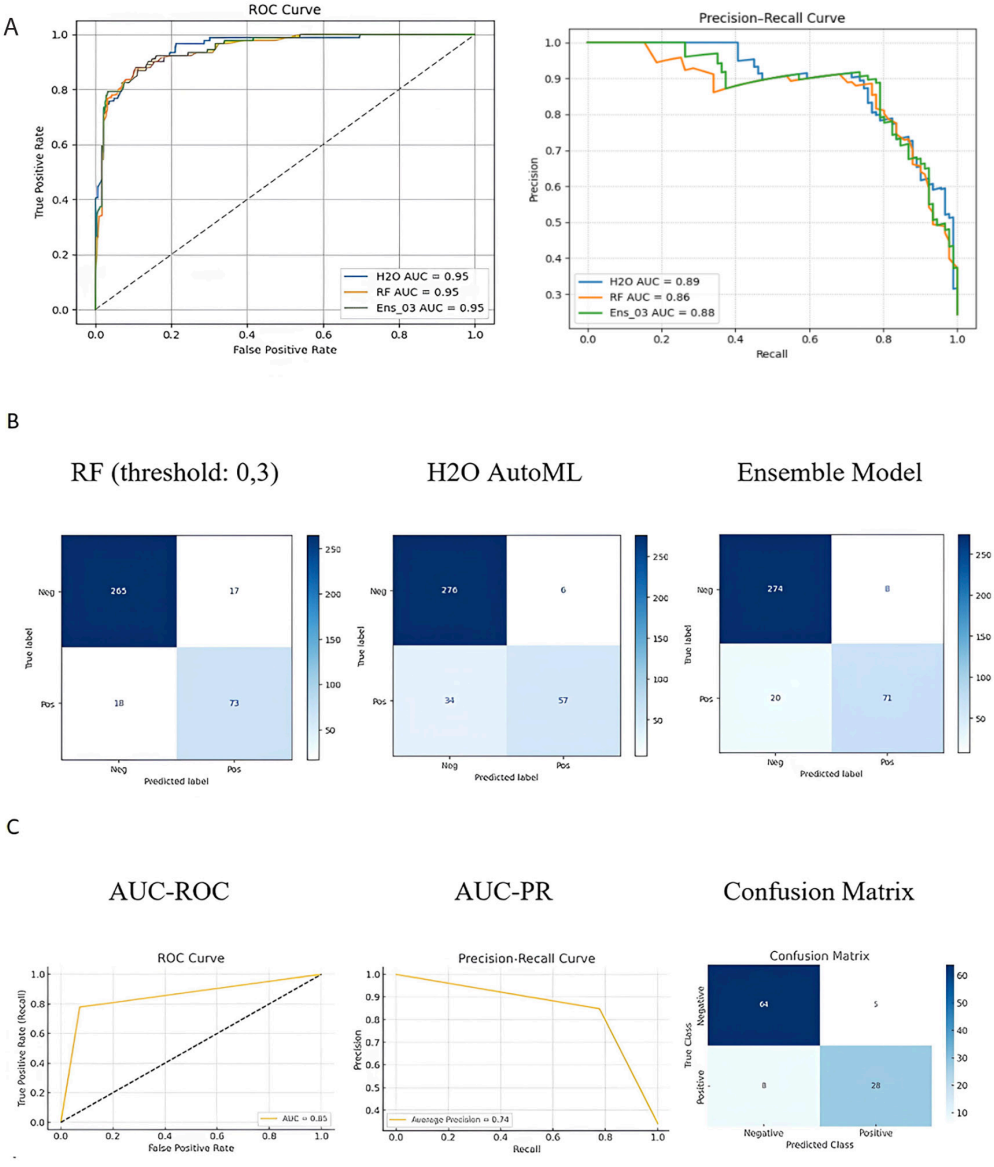


Figure 2. Model performance outputs. (A) AUC-ROC and AUC-PR plots for machine learning models. (B) Confusion matrixes for machine learning models. (C) Performance metrics of validation set

ROC: Receiver operating characteristic, AUC: Area under the curve, RF: Random forest, AutoML: Automated machine learning, PR: Precision-recall

Validation Results of the Models

In accordance with the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) recommendations, external validation was conducted to ensure the generalizability and robustness of our models. The performance outcomes of the validation set are summarized in Table 3 and Figure 2c.

The ensemble model developed in this study was further evaluated on an independent validation set to assess its generalizability beyond the original test data, in accordance with IFCC guidelines that recommend external validation for diagnostic algorithms. The model demonstrated strong classification performance, achieving a sensitivity of 0.78 and a specificity of 0.93, indicating a balanced ability to detect both positive and negative cases. The area under the ROC curve (AUC-ROC) was 0.85, while the area under the PR curve (AUC-PR) reached 0.74, reflecting solid discriminative power even in the presence of class imbalance. Additional metrics such as a PPV of 0.85, a NPV of 0.89, a PLR of 10.73, and an odds ratio of 44.8, further emphasize the clinical relevance of the model's predictions. The F1 score of 0.81 and Matthews correlation coefficient (MCC) of 0.721 confirm the model's robustness and diagnostic accuracy. These results are consistent with the findings obtained from the test set, further supporting the model's reliability across different data sources.

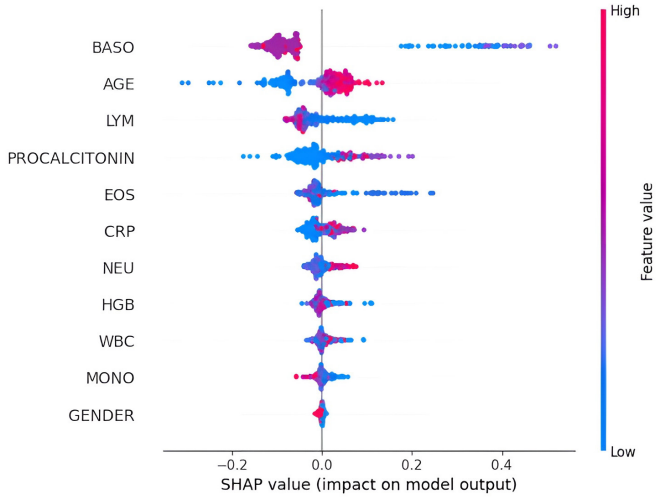


Figure 3. SHAP Summery plot of feature contributions

BASO: Basophil count, LYM: Lymphocyte, EOS: Eosinophil, CRP: C-reactive protein, NEU: Neutrophil, HGB: Hemoglobin, WBC: White blood cell, MONO: Monocyte, SHAP: SHapley Additive exPlanations

Interpretability and Threshold-Based Diagnostic Performance of Key Variables

To improve the interpretability of the model, SHAP analysis was applied to evaluate the contribution of individual features to model predictions. As illustrated in Figure 3, the most impactful variable was "BASO", followed by "AGE", "LYM", and "PROCALCITONIN". Although "BASO" ranked highest, its SHAP value distribution was narrow and centered near zero, indicating frequent but limited directional impact. In contrast, AGE and "PROCALCITONIN" displayed broader SHAP distributions, suggesting stronger influence on model output when elevated. Features such as GENDER, MONO, and WBC ranked lower in importance, showing minimal effect.

Complementing the SHAP results, ROC-based threshold analysis (Table 4) revealed that age (threshold=46.0) provided the highest sensitivity (0.92), with moderate specificity (0.55). PCT (threshold=0.26) showed a sensitivity of 0.78 and specificity of 0.68, indicating balanced diagnostic value. In contrast, basophil demonstrated high

Table 4. Receiver operating characteristic-derived diagnostic thresholds of selected variables

Feature	Threshold	Sensitivity	Specificity
Basophil	0.4	0.03	0.98
Gender	2.0	0.0	1.0
C-reactive protein	99.6	0.69	0.75
Eosinophil	4.5	0.0	1.0
Hemoglobin	6.7	1.0	0.03
Lymphocyte	60.0	0.0	1.0
Monocyte	5.0	0.01	1.0
Neutrophil	6.9	0.65	0.5
Procalcitonin	0.26	0.78	0.68
White blood cell	22.6	0.09	0.93
Age	46.0	0.92	0.55

specificity (0.98) but low sensitivity (0.03), while gender, eosinophil, lymphocyte , and monocyte achieved perfect or near-perfect specificity but negligible sensitivity.

DISCUSSION

The demographic characteristics and baseline laboratory findings of our cohort provide critical context for interpreting model behavior and clinical performance. Our median age of 46 years, male predominance, and blood culture positivity rate of 24.3% are generally consistent with prior studies on similar hospital populations.[1-3] This rate is slightly higher than in some multicenter analyses that report rates ranging from 6.6% to 12%, likely due to different inclusion criteria or local epidemiology.[14-16]

Although the mean age in the validation set was significantly higher than in the training and test sets ($p<0.001$), the model demonstrated robust performance across multiple metrics. The validation results, including high sensitivity (0.78), specificity (0.93), PPV (0.85), and an F1 score of 0.81, indicate effective generalizability without signs of overfitting. Moreover, the positive and NLRs (10.73 and 0.24, respectively) and a strong odds ratio (44.8) support the model's diagnostic strength. Nonetheless, the notable age discrepancy suggests a potential distributional shift, which could impact external validity. Therefore, monitoring model performance across different age groups in future applications is recommended to ensure consistent generalizability.

Notably, serum PCT and CRP levels were significantly elevated among culture-positive patients, which aligns with previous findings. Jeong et al.[17] reported median PCT and CRP levels of 3.2 ng/mL and 132 mg/dL, respectively, in patients with bacteremia, significantly higher than in non-bacteremia groups (0.4 ng/mL and 82.2 mg/dL). Nasimfar et al.[18] similarly observed that septic children had markedly elevated PCT (3.42 ng/mL) and CRP (55.18 mg/L) levels compared to controls. These results support the early diagnostic potential of these biomarkers. Liaudat et al.[19] further demonstrated that PCT outperformed CRP and white blood cell count in predicting culture positivity using principal component analysis.

The ensemble ML model developed in our study achieved excellent predictive metrics (AUC-ROC: 0.95, F1 score: 0.84, MCC: 0.79), outperforming most previously reported models. In contrast, other ML approaches using CatBoost or RF have typically yielded AUCs between 0.75 and 0.85.[6,7,20,21] Crucially, our model maintained high performance in external validation (AUC: 0.85; sensitivity: 0.78; specificity: 0.93), which strengthens its generalizability.

Interpretability remains a cornerstone of clinical ML implementation. Our approach integrated SHAP analysis for global feature importance with ROC-derived thresholds for clinical usability. Age and PCT stood out as dual anchors of model strength-ranking high in SHAP impact and exhibiting clear diagnostic cut-offs. This is consistent with findings by Galli et al.[22], who confirmed that PCT provides greater specificity and sensitivity than CRP in critically ill patients, and by Morgan et al.[23], who demonstrated that PCT-guided strategies improve antibiotic stewardship in febrile neutropenia.

BASO was the top-ranked SHAP feature, yet exhibited poor clinical sensitivity. This highlights the distinction between algorithmic influence and practical diagnostic utility, an important concept in applied ML.[5,13] Features such as eosinophils and sex had high specificity but low sensitivity, indicating value in reducing false positives.

The combined use of SHAP analysis and ROC-derived thresholds offers a powerful dual approach for interpreting model behavior in both algorithmic and clinical domains. While BASO emerged as the top-ranking feature in SHAP importance, its limited diagnostic utility, particularly due to low sensitivity, suggests it plays a secondary role in actual decision-making. Conversely, age and PCT were notable not only for their strong SHAP contributions but also for yielding clinically meaningful threshold values, reinforcing their role as primary diagnostic indicators within the model.

Interestingly, some variables, such as gender and eosinophils, demonstrated high specificity yet poor sensitivity. This suggests they are better suited for ruling out false positives rather than identifying true disease states, underscoring that statistical importance does not always equate to clinical utility. These findings illustrate the nuanced roles that features may play: a variable may be statistically dominant in shaping predictions (via SHAP) but lack practical impact at the bedside (via threshold behavior), or vice versa.[4,24]

Clinically, the implementation of our model could have a significant impact. ML tools have been shown to reduce unnecessary blood cultures and antibiotic use while maintaining diagnostic accuracy. For instance, Boerman et al.[25] and Martin et al.[20] report, up to 60% reduction in unnecessary cultures using ML models in ED and PICU settings. Similar benefits were observed in studies using real-time prediction tools based on vital signs or lab panels.[26]

Nonetheless, our study has limitations. The retrospective single-center design introduces potential bias, though external validation provides partial mitigation. The exclusion of clinical signs, comorbidities, and vital data-known to enhance predictive performance in other models-limits our model's clinical depth.[9,27] In addition,

the model does not distinguish true bacteremia from contamination, an issue commonly encountered in blood culture interpretation.[17,28]

Despite these limitations, our model offers a promising path forward. It uses only routine laboratory values and demographic features to deliver high performance with interpretable logic. Prospective multicenter studies and real-world deployment in clinical decision support systems (CDSS) are warranted to evaluate the true impact on diagnostic stewardship.

CONCLUSION

Our findings demonstrate that an interpretable ensemble ML model, leveraging routine hematologic and inflammatory parameters, can accurately predict blood culture positivity. By integrating SHAP-based model interpretability with clinically meaningful thresholds, the model provides both algorithmic transparency and bedside usability. Key predictors such as age and PCT consistently contributed to diagnostic performance across statistical and clinical domains. Given its external validity and reliance on readily available data, this model has the potential to be deployed in CDSS to improve diagnostic stewardship and reduce unnecessary blood cultures. Future prospective, multicenter validation studies are warranted to confirm these benefits and facilitate clinical implementation.

Ethics

Ethics Committee Approval: Ethical approval was obtained from the University of Health Sciences Türkiye, İzmir Tepecik Training and Research Hospital Ethics Committee before initiating the study (decision number: 2025/03-27, date 10.04.2025).

Informed Consent: Retrospective study.

Acknowledgments

The authors would like to express their sincere gratitude to Prof. Dr. Ayfer Çolak for her invaluable academic guidance throughout the study. We also thank Prof. Dr. Savaş Yakan, Chief Physician of the hospital, for his administrative support and facilitation of the research process.

Footnotes

Authorship Contributions

Surgical and Medical Practices: F.D., Concept: F.D., Design: F.D., A.D., Data Collection or Processing: F.D., M.A., Analysis or Interpretation: F.D., M.A., Literature Search: F.D., M.A., A.D., Writing: F.D., A.D.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Mahmoud E, Al Dhoayan M, Bosaeed M, Al Johani S, Arabi YM. Developing machine-learning prediction algorithm for bacteremia in admitted patients. Infect Drug Resist. 2021;14:757-65.
2. Shapiro NI, Wolfe RE, Wright SB, Moore R, Bates DW. Who needs a blood culture? A prospectively derived and validated prediction rule. J Emerg Med. 2008;35:255-64.
3. Raita Y, Goto T, Faridi MK, Brown DFM, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care. 2019;23:64.
4. Vijayakumar S, Nair SN, S AC, et al. AI Enhanced explainable early prediction of blood culture positivity in neutropenic patients using clinical and hematologic parameters. Comput Biol Med. 2025;189:109979.
5. Clark M. Prediction of clinical risks by analysis of preclinical and clinical adverse events. J Biomed Inform. 2015;54:167-73.
6. Campagner A, Agnello L, Carobene A, et al. Complete blood count and monocyte distribution width-based machine learning algorithms for sepsis detection: multicentric development and external validation study. J Med Internet Res. 2025;27:e55492.
7. Cheng M, Zhao X, Ding X, Gao J, Xiong S, Ren Y. Prediction of blood culture outcome using hybrid neural network model based on electronic health records. BMC Med Inform Decis Mak. 2020;20:121.
8. Zhang J, Liu W, Xiao W, Liu Y, Hua T, Yang M. Machine learning-derived blood culture classification with both predictive and prognostic values in the intensive care unit: a retrospective cohort study. Intensive Crit Care Nurs. 2024;80:103549.
9. Buchan K, Filannino M, Uzuner Ö. Automatic prediction of coronary artery disease from clinical narratives. J Biomed Inform. 2017;72:23-32.
10. van der Geest PJ, Mohseni M, Linssen J, Duran S, de Jonge R, Groeneveld ABJ. The intensive care infection score - a novel marker for the prediction of infection and its severity. Crit Care. 2016;20:180.
11. JetBrains Team. Learn IDE features. February 23, 2025. Available from: https://www.jetbrains.com/help/pycharm/feature-trainer.html
12. Fryda T, LeDell E, Gill N, Aiello S. H_2O: R Interface for the "H_2O" Scalable Machine Learning Platform. 2024. Available from: https://docs.h2o.ai/h2o/latest-stable/h2o-r/docs/index.html
13. Topcu Dİ, Bayraktar N. Searching for the urine osmolality surrogate: an automated machine learning approach. Clin Chem Lab Med. 2022;60:1911-20.
14. Zhang XJ, Zhao H, Zhang D, et al. Blood cell count-derived inflammation indices as predictors of the osteoporotic risk of postmenopausal women. Eur Rev Med Pharmacol Sci. 2024;28:2207-16.
15. Chang YH, Hsiao CT, Chang YC, et al. Machine learning of cell population data, complete blood count, and differential count parameters for early prediction of bacteremia among adult patients with suspected bacterial infections and blood culture sampling in emergency departments. J Microbiol Immunol Infect. 2023;56:782-92.

16. Parente DM, Cunha CB, Mylonakis E, Timbrook TT. The clinical utility of methicillin-resistant staphylococcus aureus (MRSA) nasal screening to rule out MRSA pneumonia: a diagnostic meta-analysis with antimicrobial stewardship implications. Clin Infect Dis. 2018;67:1-7.
17. Jeong S, Park Y, Cho Y, Kim HS. Diagnostic utilities of procalcitonin and C-reactive protein for the prediction of bacteremia determined by blood culture. Clinica Chim Acta. 2012;413:1731-6.
18. Nasimfar A, Sadeghi E, Karamyyar M, Manesh L. Comparison of serum procalcitonin level with erythrocytes sedimentation rate, C-reactive protein, white blood cell count, and blood culture in the diagnosis of bacterial infections in patients hospitalized in Motahhari hospital of Urmia (2016). J Adv Pharm Technol Res. 2018;9:147-52.
19. Liaudat S, Dayer E, Praz G, Bille J, Troillet N. Usefulness of procalcitonin serum level for the diagnosis of bacteremia. Eur J Clin Microbiol Infect Dis. 2001;20:524-7.
20. Martin B, Payan M, Greer C, et al. 411: Machine learning to improve blood culture stewardship in the picu. Crit Care Med. 2025;53(1).
21. Mooney C, Eogan M, Ní Áinle F, et al. Predicting bacteraemia in maternity patients using full blood count parameters: a supervised machine learning algorithm approach. Int J Lab Hematol. 2021;43:609-15.
22. Galli F, Bindo F, Motos A, et al. Procalcitonin and C-reactive protein to rule out early bacterial coinfection in COVID-19 critically ill patients. Intensive Care Med. 2023;49:934-45.
23. Morgan JE, Phillips B. PAnTher Cub: procalcitonin-guided antibiotic therapy for febrile neutropenia in children and young people with cancer - a single-arm pilot study. BMJ Paediatr Open. 2022;6:001339.
24. Versbraegen N, Fouché A, Nachtegael C, et al. Using game theory and decision decomposition to effectively discern and characterise bi-locus diseases. Artif Intell Med. 2019;99:101690.
25. Boerman AW, Schinkel M, Meijerink L, et al. Using machine learning to predict blood culture outcomes in the emergency department: a single-centre, retrospective, observational study. BMJ Open. 2022;12:053332.
26. Zoabi Y, Kehat O, Lahav D, Weiss-Meilik A, Adler A, Shomron N. Predicting bloodstream infection outcome using machine learning. Sci Rep. 2021;11:20101.
27. Baghdadi JD, Brook RH, Uslan DZ, et al. Association of a care bundle for early sepsis management with mortality among patients with hospital-onset or community-onset sepsis. JAMA Intern Med. 2020;180:707-16.
28. Kristóf K, Pongrácz J. Interpretation of blood microbiology results - function of the clinical microbiologist. EJIFCC. 2016;27:147-55.

DOI: 10.4274/forbes.galenos.2025.97759
Forbes J Med 2025;6(2):138-46

Original Article / Orijinal Araştırma

Retrospective Analysis of the Relationship of Serum Magnesium, Vitamin D and Non-HDL Cholesterol Values with Hepatosteatosis, Fibrosis-4 Index and Aspartate Aminotransferase to Platelet Ratio Index

Serum Magnezyum, D Vitamini ve Non-HDL Kolesterol Düzeylerinin Hepatosteatoz, Fibrozis-4 İndeksi ve Aspartat Aminotransferaz/ Trombosit Oranı İndeksi ile İlişkisinin Retrospektif Analizi

İsmail SELİMOĞLU[1], Hüseyin DÖNGELLİ[1], Mehmet Emin ARAYICI[2], Özlem GÜRSOY DORUK[3], Canan ALTAY[4], Tevfik DEMİR[5]

[1]Dokuz Eylül University Hospital, Department of Internal Medicine, İzmir, Türkiye
[2]Dokuz Eylül University Hospital, Department of Biostatistics and Medical Informatics, İzmir, Türkiye
[3]Dokuz Eylül University Hospital, Department of Biochemistry, İzmir, Türkiye
[4]Dokuz Eylül University Hospital, Department of Radiology, İzmir, Türkiye
[5]Dokuz Eylül University Hospital, Department of Internal Medicine, Division of Endocrinology, İzmir, Türkiye

Cite as: Selimoğlu İ, Döngelli H, Arayıcı ME, Gürsoy Doruk Ö, Altay C, Demir T. Retrospective analysis of the relationship of serum magnesium, vitamin D and non-HDL cholesterol values with hepatosteatosis, fibrosis-4 index and aspartate aminotransferase to platelet ratio index. Forbes J Med.2025;6(2):138-46

ABSTRACT

Objective: Non-alcoholic fatty liver disease (NAFLD) is a common chronic liver condition with an increasing global prevalence. Magnesium and vitamin D are essential for various physiological functions. However, their potential relationships with NAFLD and liver fibrosis remain unclear. This study aimed to assess the correlation between serum magnesium and vitamin D, and liver fibrosis indices, including the aspartate aminotransferase to platelet ratio index (APRI) and the fibrosis-4 (FIB-4) index.

Methods: In this retrospective study, 414 patients underwent abdominal ultrasound for hepatic steatosis assessment. Data on demographics, laboratory parameters, and imaging findings were recorded. Patients were categorized by hepatosteatosis presence, and severity. Biochemical fibrosis scores and their relationships with NAFLD were evaluated.

Results: Serum magnesium levels were inversely correlated with the FIB-4 index (r=-0.101, p=0.045). Magnesium levels were higher in low-risk groups (FIB-4 <1.45) (p=0.006). Vitamin D was inversely associated with hepatic steatosis severity (r=-0.107, p=0.031) and with APRI scores in non-diabetic/non-hyperlipidemic patients (r=-0.383, p=0.044). Low magnesium levels were linked to increased hepatic steatosis in prediabetic patients (p=0.013). Non-high-density lipoprotein cholesterol is positively correlated with steatosis severity (p<0.001).

Conclusion: Magnesium and vitamin D may have protective roles against hepatic steatosis and fibrosis. Their inverse correlations with fibrosis indices suggest potential antifibrotic effects, and magnesium could play a role in NAFLD pathogenesis, particularly in prediabetic patients. These findings support further investigation of magnesium and vitamin D in fibrosis scoring and NAFLD management.

Keywords: Non-alcoholic fatty liver disease, magnesium, vitamin D, HOMA-IR index, fibrosis-4 index

Received/Geliş: 23.04.2025
Accepted/Kabul: 02.06.2025
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Corresponding Author/ Sorumlu Yazar:
İsmail SELİMOĞLU, MD,
Dokuz Eylül University Hospital, Department of Internal Medicine, İzmir, Türkiye
ismailselm@gmail.com
ORCID: 0000-0003-2803-3456

ÖZ

Amaç: Alkole bağlı olmayan yağlı karaciğer hastalığı (NAFLD), küresel yaygınlığı giderek artan yaygın bir kronik karaciğer hastalığıdır. Magnezyum ve D vitamini birçok fizyolojik işlev için gereklidir. Ancak bu elementlerin NAFLD ve karaciğer fibrozisi ile olası ilişkileri hala net değildir. Bu çalışmanın amacı, serum magnezyum ve D vitamini düzeyleri ile aspartat aminotransferaz/trombosit oranı (APRI) ve fibrosis-4 (FIB-4) indeksi gibi karaciğer fibrozis göstergeleri arasındaki ilişkiyi değerlendirmektir.

Yöntem: Bu retrospektif çalışmada, hepatik steatoz değerlendirmesi için karın ultrasonu yapılan 414 hastanın verileri incelendi. Demografik veriler, laboratuvar bulguları ve görüntüleme sonuçları kaydedildi. Hastalar, karaciğerde yağlanma varlığı ve şiddetine göre sınıflandırıldı. Biyokimyasal fibrozis skorları ile NAFLD arasındaki ilişkiler değerlendirildi.

Bulgular: Serum magnezyum düzeyleri, FIB-4 indeksi ile ters yönde ilişkiliydi (r=-0,101, p=0,045) ve düşük risk grubunda (FIB-4 <1,45) daha yüksekti (p=0,006). D vitamini, hepatik steatoz şiddetiyle (r=-0,107, p=0,031) ve diyabetik olmayan/hiperlipidemik olmayan hastalarda APRI skorlarıyla ters ilişki gösterdi (r=-0,383, p=0,044). Prediyabetik hastalarda düşük magnezyum düzeyleri, artmış karaciğer yağlanmasıyla ilişkiliydi (p=0,013). Non-yüksek yoğunluklu lipoprotein kolesterol, steatoz şiddetiyle pozitif korelasyon gösterdi (p<0,001).

Sonuç: Magnezyum ve D vitamini, karaciğer yağlanması ve fibrozisine karşı koruyucu etkilere sahip olabilir. Fibrozis indeksleriyle olan ters ilişkileri, potansiyel anti-fibrotik etkilerini düşündürmektedir. Özellikle prediyabetik hastalarda magnezyum, NAFLD patogenezinde rol oynayabilir. Bu bulgular, magnezyum ve D vitamininin fibrozis değerlendirmesi ve NAFLD yönetimindeki potansiyel rollerinin daha fazla araştırılmasını desteklemektedir.

Anahtar Kelimeler: Alkole bağlı olmayan yağlı karaciğer hastalığı, magnezyum, D vitamini, HOMA-IR indeksi, fibrozis-4 indeksi

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is the most common chronic liver disease worldwide, with a prevalence of approximately 25%, and its incidence is increasing. It is characterized by excessive fat accumulation in the liver in the absence of significant alcohol consumption, where significant alcohol consumption is defined as less than 21 units (30 g/day) per week in men and 14 units (20 g/day) per week in women. Several risk factors contribute to the development of NAFLD, including advanced age, male sex, ethnicity, obesity, hypertriglyceridemia, low levels of high-density lipoprotein (HDL) cholesterol, hypertension, type 2 diabetes mellitus (T2DM), and various genetic predispositions.[1]

NAFLD is now recognized as a hepatic manifestation of insulin resistance (IR) and a component of metabolic syndrome. Increased IR leads to increased lipid influx into the liver, which, combined with inflammatory cytokines and oxidative stress factors, triggers inflammation and subsequent fibrosis.[1-4]

Currently, NAFLD is most commonly detected using ultrasonography (USG) during routine examinations or screening because of its feasibility and accessibility. This assessment is based on differences in echogenicity between the liver, kidney, and portal vein walls.[5] However, it does not provide histological evidence of disease activity. Therefore, several scoring systems have been developed that incorporate biochemical parameters to estimate histological activity, without the need for liver biopsy. Among these, the fibrosis-4 (FIB-4) index and the aspartate aminotransferase to platelet ratio index (APRI) are widely used in clinical practice.[6-8] Although these scores are not fully sufficient to differentiate simple steatosis from steatohepatitis, one study has reported that the APRI score correlates with the severity of NAFLD as determined by radiological findings.[9] Magnesium (Mg^{2+}) is the most abundant intracellular divalent cation and plays a crucial role in many physiological processes. Due to its involvement in enzymatic reactions, Mg^{2+} has been reported to play a protective role in hepatic inflammatory processes and hepatocyte damage, which are fundamental histological features of NAFLD.[10,11]

Vitamin D-related receptors are known to affect metabolic pathways in skeletal muscle, liver, adipose tissue, and pancreas.[12,13] Through these mechanisms, vitamin D deficiency contributes to IR. In addition, inadequate anti-inflammatory responses have been associated with increased hepatocyte inflammation, the development of hepatic steatosis or steatohepatitis, and an increased risk of metabolic syndrome.[14,15]

Based on these data, we aimed to investigate the relationship between biochemical parameters and NAFLD, as well as the aspartate APRI and the FIB-4 index, to identify thresholds that may indicate a protective role against steatosis and fibrosis.

METHODS

Patients and Data Collection

A total of 691 patient records were reviewed. Among them, the following cases were excluded based on predefined exclusion criteria: 38 patients with liver cirrhosis; 65 patients with a history of past or chronic hepatitis; 25 patients who had received pioglitazone or liraglutide therapy; 67 patients with malignancy; 14 patients with a history of inflammatory bowel disease; 9 patients with autoimmune hepatitis; 2 patients with celiac disease; 10 patients with a history of chronic alcohol use; 43 patients using medications known to induce hepatic steatosis (anti-tumor necrosis factor therapy, corticosteroids, chemotherapeutic agents, etc.); and 4 patients diagnosed with Cushing's syndrome.

As a result, 414 patients (230 females and 184 males) were included in this study.

Inclusion Criteria

The study included male and female patients aged 18 years and older who presented to our outpatient clinic between March 2012 and September 2022. Participants were required to have undergone abdominopelvic USG and to have relevant laboratory parameters available for evaluation. Individuals were only included if they did not meet any of the exclusion criteria listed below at the time of their hospital visit.

Exclusion Criteria

Patients with any of the following conditions during the data collection period were excluded from the study: Age under 18, alcohol consumption, presence of autoimmune hepatobiliary disease, history of viral hepatitis carriage or active hepatitis, diagnosis of cirrhosis of the liver or inflammatory bowel disease, presence of active malignancy or history of malignancy, or Cushing's syndrome. Excessive alcohol consumption was defined as an average intake of more than 30 grams per day for men and more than 20 grams per day for women.

Hepatosteatosis Assessment

Ultrasound assessment of hepatic steatosis was performed using a LOGIQ S7 Expert ultrasound machine. A convex probe was used for B-mode imaging, scanning from the superior to inferior through the subcostal and intercostal regions. IR, non-HDL/HDL cholesterol ratio, and biochemical fibrosis parameters, including the APRI and FIB-4 scores, were calculated. The presence and severity of hepatic steatosis were assessed using USG. Hepatic steatosis was graded according to increased echogenicity, as follows:[16]

- **Grade 1:** Mild diffuse increase in echogenicity with normal visualization of the diaphragm and intrahepatic vessel walls.
- **Grade 2:** Moderate increase in echogenicity with slight opacity of the diaphragm and intrahepatic vessel walls.
- **Grade 3:** Severe enhancement of echogenicity with marked blurring of the diaphragm and intrahepatic vessel walls, and loss of visualization of the posterior segment of the right hepatic lobe.

Indices

The non-HDL/HDL cholesterol ratio was calculated by subtracting HDL cholesterol from total cholesterol to obtain non-HDL cholesterol, which was then divided by HDL cholesterol. The aspartate APRI, one of the indices used to assess liver fibrosis, was calculated by dividing the AST elevation ratio by the platelet count and multiplying the result by 100. The AST elevation ratio was determined by dividing the patient's AST level by the upper laboratory limit (35 U/L for women and 50 U/L for men). The FIB-4 index was calculated using the formula: (age × AST) / (platelet count × √ALT).

Patients were also categorised according to

- Vitamin D levels: ≥20 ng/mL vs. <20 ng/mL
- Magnesium levels: ≤0.75 mmol/L vs. >0.75 mmol/L
- Non-HDL cholesterol levels: ≤150 mg/dL vs. >150 mg/dL

Ethics

Approval for this retrospective study was obtained from Dokuz Eylül University Non-Interventional Research Ethics Committee (decision number: 2023/37-17, date: 23.11.2022).

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics 24.0. Descriptive statistics are presented as mean ± standard deviation (median, minimum, and maximum) and percentages. Shapiro-Wilk and Kolmogorov-Smirnov tests were used to assess the normality of the data distribution. As the data did not follow a normal distribution, the Mann-Whitney U test was used for pairwise comparisons of numerical variables. The Kruskal-Wallis test was used for comparisons involving more than two groups. Subgroup analyses in the parametric tests were interpreted using the Bonferroni correction. Categorical variables were analyzed using the chi-square test, and correlations between two numerical variables were assessed using the Spearman correlation test. A p value <0.05 or a 95% confidence interval that does not include the null value was considered statistically significant.

RESULTS

Demographic and Clinical Characteristics of the Patients

The mean age of all patients was 53.2±17.1 years (minimum: 18, maximum: 91 years). Descriptive statistics of the patients' biochemical parameters are presented in Table 1. On ultrasonographic evaluation, 166 patients (40.1%) did not have hepatic steatosis, whereas 248 patients (59.9%) had varying degrees of steatosis (Table 1).

The correlations between biochemical parameters and hepatic steatosis when patients were evaluated as a whole, without stratification by comorbidities or sex, are summarized in Table 2. Non-HDL cholesterol levels and non-HDL/HDL cholesterol ratios were inversely correlated with the degree of hepatic steatosis ($p<0.001$ and $p<0.001$, respectively). In non-diabetic patients, non-HDL cholesterol and the non-HDL/HDL ratio were significantly

higher in those with steatosis (Table 3). In the entire study population, a positive correlation was observed between non-HDL cholesterol levels and the APRI scores (r=0.126, p=0.012). In patients without diabetes and hyperlipidemia, non-HDL cholesterol levels showed a positive correlation with the steatosis grade (r=0.415, p=0.025), whereas the non-HDL/HDL ratio did not show a significant correlation (r=0.289, p=0.128).

Factors Related to Hepatosteatosis

When patients were evaluated as a whole, without stratification by comorbidities or sex, the correlations between biochemical parameters and hepatic steatosis were summarized in Table 2. Non-HDL cholesterol levels and non-HDL/HDL cholesterol ratios were inversely correlated with the degree of hepatic steatosis (p<0.001 and p<0.001, respectively).

Table 1. Clinical characteristics of study group

	No hepatosteatosis (n=166)	**Hepatosteatosis (n=268)**	**Total (n=414)**	**p**
Age mean ± SD	53.6±20.3	53.0±14.6	53.2±17.1	0.774
Gender n (%) Female Male	 104 (62.7) 62 (37.3)	 126 (50.8) 122 (49.2)	 230 (55.6) 184 (44.4)	 **0.017**
Comorbidities n (%) None Dyslipidemia DM	 19 (11.4) 136 (81.9) 83 (50.0)	 10 (3.7) 220 (88.7) 187 (75.4)	 29 (7.0) 356 (86.0) 270 (65.2)	 **0.002** 0.051 **<0.001**
Hepatosteatosis n (%) Grade 0 Grade 1 Grade 2 Grade 3	 166 (100.0) 0 (0) 0 (0) 0 (0)	 0 (0) 94 (37.9) 123 (49.6) 31 (12.5)	 166 (40.1) 94 (22.7) 123 (29.7) 31 (7.5)	 N/A
Fibrosis indicis median (IQR) FIB-4 APRI	 0.93 (0.86) 0.20 (0.11)	 0.87 (0.61) 0.20 (0.13)	 0.88 (0.69) 0.20 (0.12)	 0.217 0.979
Laboratory findings median (IQR) Creatinine mg/dL BUN mg/dL AST U/L ALT U/L GGT U/L ALP U/L LDL mg/dL HDL mg/dL VLDL mg/dL Triglyceride mg/dL T. cholesterol mg/dL Albumin g/dL Calcium mg/dL Magnesium mmol/L Vitamin D ng/mL Ferritine ng/mL CRP mg/dL FBG mg/dL HbA1c% Lymphocyte 10^9/L Thrombocyte 10^9/L HOMA-IR	 0.71 (0.32) 13.8 (10.2) 19 (9) 16 (12) 18 (16) 68 (28) 120 (53) 52 (22) 21 (12) 105 (59) 199 (64) 4.2 (0.6) 9.5 (0.7) 0.84 (0.11) 18 (15) 35 (72) 2.1 (6.2) 91 (17) 5.7 (0.7) 1.9 (0.9) 246 (77) 2.0 (2.5)	 0.80 (0.20) 14 (7) 21 (10) 22 (17) 27 (21) 76 (27) 139 (53) 48 (18) 29 (18) 144 (92) 217 (72) 4.3 (0.5) 9.6 (0.6) 0.84 (0.10) 17 (13) 46 (69) 4.5 (6.4) 97 (29) 6.0 (1.1) 2.2 (0.9) 263 (94) 3.5 (3.1)	 0.79 (0.31) 10 (9) 20 (10) 19 (16) 22 (20) 72 (28) 131 (55) 50 (18) 25 (17) 124 (87) 207 (71) 4.3 (0.5) 9.6 (0.7) 0.84 (0.10) 17 (13) 40 (69) 3.6 (7.0) 94 (26) 5.9 (0.9) 2.1 (0.9) 257 (89) 2.9 (2.8)	 0.562 0.652 **0.003** **<0.001** **<0.001** **<0.001** **<0.001** **0.032** **<0.001** **<0.001** **<0.001** **0.017** **0.007** 0.842 0.237 0.235 **<0.001** **<0.001** **<0.001** **<0.001** **0.026** **0.001**
Vitamin D deficiency n (%)	10 (6.0)	23 (8.6)	33 (7.9)	0.360
Hypomagnesemia n (%)	29 (17.5)	35 (13.1)	64 (15.5)	0.354

ALP: Alkaline phosphatase, ALT: Alanine transferase, APRI: Aspartate transferase to platelets ratio index, AST: Aspartate transferase, BUN: Blood urea nitrogen, CRP: C-reactive protein, DM: Diabetes mellitus, FBG: Fasting blood glucose, FIB-4: Fibrosis 4 score, GGT: Gamma-glutamyl transferase, HbA1c: Hemoglobin A1c, HDL: High-density lipoprotein, HOMA-IR: Homeostasis model assessment of insulin resistance, LDL: Low-density lipoprotein, N/A: Not applicable, IQR: Interquartile range, VLDL: Very low-density lipoprotein, IQR: Interquartile range, SD: Standard deviation

In non-diabetic patients, non-HDL cholesterol and non-HDL/HDL ratio were significantly higher in those with steatosis (Tables 2 and 3). In the entire study population, a positive correlation was observed between non-HDL cholesterol levels and APRI scores (r=0.126, p=0.012). In patients without diabetes and hyperlipidaemia, non-HDL cholesterol levels showed a positive correlation with the steatosis grade (r=0.415, p=0.025), whereas the non-HDL/HDL ratio did not show a significant correlation (r=0.289, p=0.128).

Serum vitamin D levels were inversely correlated with the degree of steatosis in the overall population and in patients with both hyperlipidemia and T2DM/preDM. Non-HDL cholesterol levels and non-HDL/HDL ratio were positively correlated with steatosis grade in the overall population, independent of comorbidities. In addition, these parameters were positively correlated with steatosis grades in patients with both hyperlipidemia and diabetes/prediabetes (Table 2).

Fibrosis Indices

A significant negative correlation was observed between serum magnesium levels and the FIB-4 index (r=-0.101, p=0.045). Furthermore, when the FIB-4 score was stratified into two groups based on a cutoff value of 1.45, magnesium levels were significantly higher in the low-risk FIB-4 group (Table 4). In addition, in patients with hyperlipidemia, magnesium levels were significantly increased in the FIB-4 <1.45 group (p<0.05) (Table 4).

Subgroup Analysis

When patients were divided into two groups based on a serum magnesium threshold of 0.75 mmol/L, no correlation between vitamin D levels and hepatic steatosis severity was observed in the group with higher magnesium levels. However, in patients with lower magnesium levels, there was a significant inverse correlation between serum vitamin D levels and hepatic steatosis severity according to Spearman's rho test (p<0.01; r=0.385). Furthermore, chi-squared analysis showed that hepatic steatosis was more common in patients with both serum magnesium levels ≤0.75 mmol/L and vitamin D levels <20 ng/mL (Table 5).

No significant correlation was found between vitamin D levels and another fibrosis index, APRI, in the whole population. However, in patients without diabetes and hyperlipidaemia, there was a negative correlation between vitamin D levels and APRI (r=-0.383, p=0.044). We also analysed the diabetic/prediabetic group separately. After excluding patients receiving any diabetes treatment and/or those with HbA1c ≥6.5% and fasting blood glucose ≥126 mg/dL, 45 prediabetic patients remained.

Table 2. Relationship between hepatosteatosis degree and biochemical parameters

	r	p*
All patients		
Magnesium	-0.200	0.678
Vitamin D	-0.107	0.031
Non-HDL cholesterol	0.277	<0.001
Non-HDL/HDL cholesterol ratio	0.323	<0.001
Patients with DM and hyperlipidemia		
Magnesium	-0.030	0.620
Vitamin D	-0.128	0.048
Non-HDL/HDL ratio	0.313	<0.001
Non-HDL cholesterol	0.274	<0.001

*Determined by Spearman's rho correlation test. r: Correlation coefficient.
HDL: High-density lipoprotein, DM: Diabetes mellitus

Table 3. Comparison of non-high-density lipoprotein cholesterol and non-high-density lipoprotein/high-density lipoprotein cholesterol ratios with hepatosteatosis in non-type 2 diabetes mellitus/pre-diabetes mellitus patients

	Hepatosteatosis status (n)	Median (25%-75%)	p*
Non-HDL (mg/dL)	Absent (83)	148 (112-179)	**0.007**
	Present (61)	163 (141.5-192.5)	
Non-HDL/HDL ratio	Absent (83)	2.63 (1.97-3.64)	**0.023**
	Present (61)	3.18 (2.40-3.74)	

*Calculated using Mann-Whitney U test.
HDL: High-density lipoprotein, T2DM: Type 2 diabetes mellitus

Table 4. Comparison of magnesium levels in patients with fibrosis-4 scores below and above 1.45			
	Sample size (n)	**Magnesium (mmol/L) Median (25-75%)**	**p**
Total			
FIB-4 <1.45	316	0.85 (0.792-0.90)	**0.006**
FIB-4 ≥1.45	80	0.82 (0.762-0.887)	
Patient with hyperlipidemia			
FIB-4 <1.45	264	0.845 (0.79-0.89)	**0.026**
FIB-4 ≥1.45	74	0.820 (0.76-0.88)	
*p values were calculated using the Mann-Whitney U test. FIB-4: Fibrosis-4			

Table 5. Association between vitamin D levels and the presence of hepatic steatosis in patients with serum magnesium levels ≤0.75 mmol/L				
		Present n (%)	**Absent n (%)**	**p***
Magnesium ≤0.75 mmol/L	Vitamin D <20 ng/mL	23 (69.7)	10 (30.3)	0.013
	Vitamin D ≥20 ng/mL	12 (38.7)	19 (61.3)	
*Chi-square test				

In this subgroup of exclusively prediabetic patients, we observed a significant inverse correlation between serum Mg^{+2} levels and HOMA-IR (Spearman's rho, p<0.01, r=-0.396). In prediabetic but non-diabetic patients, serum magnesium levels were significantly higher in those with HOMA-IR <2 compared to those with HOMA-IR ≥2 (median 0.90 vs. 0.87 mmol/L; p=0.008), suggesting a potential inverse association between IR and magnesium levels.

DISCUSSION

Although the relationship between magnesium and the inflammatory response remains unclear, adequate serum and intracellular magnesium levels are thought to have a protective role against inflammation.[17] In a study by Rayssiguier et al.[18], in rats, hepatic magnesium depletion was associated with increased collagen accumulation in the liver. In addition, data suggest that both serum and intracellular magnesium levels are significantly lower in cirrhotic patients.[19] Furthermore, studies have shown that each 100 mg increase in magnesium intake is associated with a 49% reduction in the risk of death from all causes of liver disease, and increased magnesium intake is associated with a reduced risk of advanced decompensated liver disease.[20,21] Although no similar study has been reported in the literature, our study found a weak inverse correlation between serum magnesium levels and the FIB-4 index, which is considered a predictive index of liver fibrosis, in the entire population evaluated (p<0.05, r=-0.101). In addition, when patients were grouped based on a FIB-4 threshold of 1.45, serum magnesium levels were significantly lower in the higher FIB-4 group (p<0.01). Due to the absence of patients with cirrhosis in our population, the APRI and FIB-4 scores were generally low. Nevertheless, the negative correlation between serum magnesium and the FIB-4 index suggests that magnesium may be directly or indirectly involved in antifibrotic and anti-inflammatory pathways.

Another study investigated the relationship between serum and intracellular magnesium levels, IR, and severity of hepatic steatosis in patients with metabolic syndrome. Patients with grades 2 and 3 steatosis, as assessed by ultrasound, were found to have lower magnesium levels than those without steatosis. In the same study, individuals with a HOMA-IR above 2.7 also had lower serum magnesium levels compared to those with a HOMA-IR <2.7.[22] It should be noted that maintaining a high HOMA-IR threshold may have contributed to this finding. In an obesity study of patients with liver biopsy-confirmed NASH, patients with NASH had lower magnesium levels than those without NASH. Furthermore, low magnesium concentrations were found to be associated with high HOMA-IR.[23] A meta-analysis of randomized controlled trials showed that magnesium supplementation improved IR in patients with type 2 diabetes, further supporting these findings.[24] As IR is a fundamental mechanism of NAFLD, it plays a crucial role in the disease's progression. Our study also provided data consistent with the literature. In the prediabetic patient population, a significant inverse correlation was found between serum magnesium and HOMA-IR levels. When patients were divided into two groups based on a HOMA-IR threshold of 2.00, those with a higher HOMA-IR had lower magnesium levels. This suggests that even in the absence of diabetes, serum magnesium levels may be affected under conditions of high IR. For example, a study comparing 50 patients with glucose intolerance or

diabetes with 50 healthy controls found no statistically significant difference in serum magnesium levels between the groups.[25]

Vitamin D regulates several genes that are widely expressed in the liver, some of which are involved in glucose and lipid metabolism.[26] A study by Chung et al.[27] of 6,055 healthy individuals diagnosed with steatosis by ultrasound found an inverse correlation between vitamin D levels and the severity of hepatic steatosis. Individuals with vitamin D levels >20 ng/mL were found to have a lower prevalence of steatosis. Similarly, a meta-analysis of 15 studies involving 20,096 individuals showed that patients with NAFLD of Western origin tended to have lower vitamin D levels. In the general population, low vitamin D levels have been associated with an increased risk of NAFLD.[28] A cohort study of 10,960 cases also found an inverse association between ultrasonographic steatosis severity and vitamin D levels, and between NAFLD fibrosis scores and vitamin D levels. Another study of biopsy-confirmed cases of liver fibrosis reported an inverse association between vitamin D levels and the severity of fibrosis. However, a separate biopsy-based study found no association between vitamin D levels and the severity of hepatic steatosis.[29-31]

In a study that examined 614 patients by biopsy, without stratifying them by comorbidities such as hyperlipidemia and diabetes, patients with advanced steatosis had significantly lower vitamin D levels. There was also an inverse association between vitamin D levels and steatosis severity. However, the same study found no association between serum magnesium levels and the presence or severity of steatosis.[32] Notably, these studies examined vitamin D and magnesium levels separately, rather than assessing their combined effects.

Consistent with the literature, we found an inverse association between vitamin D levels and the severity of hepatic steatosis in the whole population. Furthermore, when we analyzed patients with both hyperlipidemia and diabetes/prediabetes separately from those without these comorbidities, we found a significant inverse association between the steatosis severity and vitamin D levels in these patients. However, no such correlation was observed in the patients without these comorbidities.

Studies in the literature have separately evaluated the relationships between serum magnesium levels, vitamin D levels, and NAFLD. However, in our analysis, when patients were divided into two groups based on a threshold serum magnesium level of 0.75 mmol/L, a significant inverse association between vitamin D levels and the degree of hepatic steatosis was observed in those within the lower magnesium group. Furthermore, when the group with low magnesium levels was subdivided based on a vitamin D threshold of 20 ng/mL, hepatic steatosis was more prevalent in patients with both low vitamin D and low magnesium levels (p=0.013). No significant association was found in the higher magnesium group. Based on the available literature, if low serum magnesium levels are accepted as being associated with NAFLD, it can be suggested that in patients with hypomagnesemia or those who are unable to maintain optimal magnesium levels for any reason. Ensuring adequate vitamin D levels or preventing vitamin D deficiency may have a protective effect against the occurrence and severity of NAFLD.

Vitamin D and APRI levels were examined in a study of 3,972 patients categorized into diabetes, pre-diabetes, and normal glucose tolerance groups. Despite a significant p value for correlation, no significant difference in mean APRI scores was found between different categories of vitamin D levels.[33] Another study of 58 HCV cases divided into two groups based on vitamin D levels found that APRI scores were higher in the low vitamin D group, with an inverse correlation between vitamin D levels and APRI.[34] Similarly, a study of 185 patients with primary biliary cholangitis and 141 healthy controls found an inverse association between vitamin D levels and APRI.[35]

In our study, no association was found between vitamin D and APRI when the entire population was analyzed. However, in patients without hyperlipidemia and diabetes/pre-diabetes comorbidities, a significant inverse association was found between vitamin D levels and APRI. This suggests that vitamin D levels may not be sufficient to predict fibrosis in patients with metabolic comorbidities. However, in patients without metabolic comorbidities, incorporating vitamin D levels into the APRI score may help predict liver fibrosis.

In a study that followed 147 patients without initial NAFLD for seven years, 19% developed NAFLD. However, none of the patients with a baseline non-HDL level of <130 mg/dL developed NAFLD. The incidence of NAFLD increased with increasing non-HDL-C levels, leading to the conclusion that non-HDL-C levels may be a strong predictor of NAFLD development.[36] Similarly, in a study of 2,717 cases with a mean follow-up of 1.6 years, 9.1% (264 cases) developed NAFLD, and the non-HDL/HDL ratio was significantly higher in those who developed NAFLD than in those who did not. Statistical analysis identified cut-off values for the non-HDL/HDL cholesterol ratio of 2.4 in women and 2.3 in men.[37] Another clinical study of 265 cases found that non-HDL/HDL cholesterol levels were positively associated with the degree of hepatosteatosis and the FIB-4 index.[38]

Our analyses showed that both non-HDL cholesterol and non-HDL/HDL cholesterol ratio were positively correlated with the degree of steatosis. When patients were

divided into groups based on the presence or absence of hepatosteatosis, the parameters of non-HDL levels and non-HDL/HDL ratios showed significant differences between groups, with higher values observed in those with steatosis.

Although recent studies have begun to highlight the superiority of the non-HDL/HDL cholesterol ratio over non-HDL cholesterol alone, our findings suggest that non-HDL cholesterol levels may be more indicative than the non-HDL/HDL cholesterol ratio. For example, in individuals with neither diabetes/prediabetes nor hyperlipidemia, no association was found between the non-HDL/HDL cholesterol ratio and the degree of steatosis, whereas in the same group, non-HDL cholesterol levels were correlated with the degree of steatosis. Another indication of the functional superiority of non-HDL cholesterol over the non-HDL/HDL ratio is its relationship with the APRI score. In the whole population, there was a significant positive correlation between the APRI and non-HDL cholesterol levels. The correlation between biochemical fibrosis parameters and radiological steatosis has also been investigated. In a study by Sahin et al.[9] in which 205 organ transplant donors were evaluated, a significant correlation was found between the severity of steatosis detected by CT and the APRI score. However, in our study, these parameters did not provide any data regarding the presence or severity of hepatosteatosis.

Study Limitations

This study has several limitations that should be acknowledged. First, the retrospective and cross-sectional design precludes the establishment of causal relationships between serum magnesium, vitamin D levels, and hepatic fibrosis or steatosis. Second, the study was conducted at a single center, which may limit the generalizability of the findings to broader populations. Third, some laboratory and clinical parameters had missing data, potentially affecting the completeness and accuracy of the analyses. Additionally, hepatic steatosis was assessed using abdominal ultrasonography, which, while non-invasive and widely available, is less sensitive and specific compared to histopathological evaluation. Lastly, liver biopsy, the gold standard for diagnosing and staging non-alcoholic fatty liver disease and fibrosis, was not performed, which limits the precision in determining the severity of hepatic pathology.

CONCLUSION

In light of this information and the data obtained, serum vitamin D levels and non-HDL/HDL cholesterol ratios could be included in the modification of the APRI score, which is currently used to predict liver fibrosis. Magnesium and vitamin D appear to exert protective effects against hepatic steatosis and fibrosis. The observed inverse association with fibrosis indices may indicate potential anti-fibrotic properties. Moreover, magnesium may contribute to the pathophysiological mechanisms underlying NAFLD, particularly in individuals with prediabetes. These findings underscore the need for further research on the roles of magnesium and vitamin D in fibrosis assessment and clinical management of NAFLD.

Ethics

Ethics Committee Approval: Approval for this retrospective study was obtained from Dokuz Eylül University Non-Interventional Research Ethics Committee (decision number: 2023/37-17, date: 23.11.2022).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

Concept: İ.S., H.D., M.E.A., Ö.G.D., C.A., T.D., Design: İ.S., H.D., M.E.A., Ö.G.D., C.A., T.D., Data Collection or Processing: İ.S., H.D., Analysis or Interpretation: İ.S., H.D., M.E.A., T.D., Literature Search: İ.S., H.D., Writing: İ.S., H.D.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

Note: This study was derived from the medical specialty thesis of İsmail Selimoğlu, which was completed in 2022.

REFERENCES

1. Vernon G, Baranova A, Younossi ZM. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. Aliment Pharmacol Ther. 2011;34:274-85.
2. Rinella ME. Nonalcoholic fatty liver disease: a systematic review. JAMA. 2015;313:2263-73.
3. Kotronen A, Yki-Järvinen H. Fatty liver: a novel component of the metabolic syndrome. Arterioscler Thromb Vasc Biol. 2008;28:27-38.
4. Jou J, Choi SS, Diehl AM. Mechanisms of disease progression in nonalcoholic fatty liver disease. Semin Liver Dis. 2008;28:370-9.
5. Saverymuttu SH, Joseph AE, Maxwell JD. Ultrasound scanning in the detection of hepatic fibrosis and steatosis. Br Med J (Clin Res Ed). 1986;292:13-5.
6. Kruger FC, Daniels CR, Kidd M, et al. APRI: a simple bedside marker for advanced fibrosis that can avoid liver biopsy in patients with NAFLD/NASH. S Afr Med J. 2011;101:477-80.
7. Loaeza-del-Castillo A, Paz-Pineda F, Oviedo-Cárdenas E, Sánchez-Avila F, Vargas-Voráckova F. AST to platelet ratio index (APRI) for the noninvasive evaluation of liver fibrosis. Ann Hepatol. 2008;7:350-7.
8. Maselli DB, Bazerbachi F, Matar R, et al. Sa1563 Fib-4 index correlates with measures of nash activity in subjects with nash and low burden of fibrosis. Gastroenterology. 2020;158:S-351.

9. Sahin T, Arikan BT, Serin A, et al. The utility of noninvasive scoring systems for prediction of hepatic steatosis in liver transplantation donor candidates. Transplant Proc. 2019;51:2383-6.
10. Lu L, Hao K, Hong Y, et al. Magnesium isoglycyrrhizinate reduces hepatic lipotoxicity through regulating metabolic abnormalities. Int J Mol Sci. 2021;22:5884.
11. Fengler VH, Macheiner T, Goessler W, Ratzer M, Haybaeck J, Sargsyan K. Hepatic response of magnesium-restricted wild type mice. Metabolites. 2021;11:762.
12. Benetti E, Mastrocola R, Chiazza F, et al. Effects of vitamin D on insulin resistance and myosteatosis in diet-induced obese mice. PLoS One. 2018;13:0189707.
13. Pittas AG, Joseph NA, Greenberg AS. Adipocytokines and insulin resistance. J Clin Endocrinol Metab. 2004;89:447-52.
14. Dawson-Hughes B, Staten MA, Knowler WC, et al. Intratrial exposure to vitamin D and new-onset diabetes among adults with prediabetes: a secondary analysis from the vitamin D and type 2 diabetes (D2d) Study. Diabetes Care. 2020;43:2916-22.
15. Barchetta I, De Bernardinis M, Capoccia D, et al. Hypovitaminosis D is independently associated with metabolic syndrome in obese patients. PLoS One. 2013;8:68689.
16. Sonsuz A, Baysal B. Karaciğer yağlanması ve non alkolik steatohepatit. Güncel Gastroenteroloji. 2011;15:98-106.
17. Liu M, Yang H, Mao Y. Magnesium and liver disease. Ann Transl Med. 2019;7:578.
18. Rayssiguier Y, Chevalier F, Bonnet M, Kopp J, Durlach J. Influence of magnesium deficiency on liver collagen after carbon tetrachloride or ethanol administration to rats. J Nutr. 1985;115:1656-62.
19. Romani AM. Magnesium homeostasis and alcohol consumption. Magnes Res. 2008;21:197-204.
20. Lu L, Chen C, Li Y, et al. Magnesium intake is inversely associated with risk of non-alcoholic fatty liver disease among American adults. Eur J Nutr. 2022;61:1245-54.
21. Li W, Zhu X, Song Y, et al. Intakes of magnesium, calcium and risk of fatty liver disease and prediabetes. Public Health Nutr. 2018;21:2088-95.
22. Lima Mde L, Cruz T, Rodrigues LE, et al. Serum and intracellular magnesium deficiency in patients with metabolic syndrome--evidences for its relation to insulin resistance. Diabetes Res Clin Pract. 2009;83:257-62.
23. Rodríguez-Hernández H, Gonzalez JL, Rodríguez-Morán M, Guerrero-Romero F. Hypomagnesemia, insulin resistance, and non-alcoholic steatohepatitis in obese subjects. Arch Med Res. 2005;36:362-6.
24. Song Y, He K, Levitan EB, Manson JE, Liu S. Effects of oral magnesium supplementation on glycaemic control in type 2 diabetes: a meta-analysis of randomized double-blind controlled trials. Diabet Med. 2006;23:1050-6.
25. Rahim MA, Mitra P. Serum magnesium level of newly detected patients with glucose intolerance and its comparison with serum magnesium level of age and sex matched healthy volunteers. J Med (Bangladesh). 2018;19:95-9.
26. Black LJ, Jacoby P, She Ping-Delfos WC, et al. Low serum 25-hydroxyvitamin D concentrations associate with non-alcoholic fatty liver disease in adolescents independent of adiposity. J Gastroenterol Hepatol. 2014;29:1215-22.
27. Chung GE, Kim D, Kwak MS, et al. The serum vitamin D level is inversely correlated with nonalcoholic fatty liver disease. Clin Mol Hepatol. 2016;22:146-51.
28. Liu T, Xu L, Chen FH, Zhou YB. Association of serum vitamin D level and nonalcoholic fatty liver disease: a meta-analysis. Eur J Gastroenterol Hepatol. 2020;32:140-7.
29. Kim HS, Rotundo L, Kothari N, Kim SH, Pyrsopoulos N. Vitamin D is associated with severity and mortality of non-alcoholic fatty liver disease: a US population-based study. J Clin Transl Hepatol. 2017;5:185-92.
30. Luger M, Kruschitz R, Kienbacher C, et al. Prevalence of liver fibrosis and its association with non-invasive fibrosis and metabolic markers in morbidly obese patients with vitamin D deficiency. Obes Surg. 2016;26:2425-32.
31. Patel YA, Henao R, Moylan CA, et al. Vitamin D is not associated with severity in NAFLD: results of a paired clinical and gene expression profile analysis. Am J Gastroenterol. 2016;111:1591-8.
32. Bertol FS, Araujo B, Jorge BB, Rinaldi N, De Carli LA, Tovo CV. Role of micronutrients in staging of nonalcoholic fatty liver disease: a retrospective cross-sectional study. World J Gastrointest Surg. 2020;12:269-76.
33. Corbin KD, Pittas AG, Desouza C, et al. Indices of hepatic steatosis and fibrosis in prediabetes and association with diabetes development in the vitamin D and type 2 diabetes study. J Diabetes Complications. 2023;37:108475.
34. de Almeida JP, Liberatti LS, Barros FE, et al. Profile of oxidative stress markers is dependent on vitamin D levels in patients with chronic hepatitis C. Nutrition. 2016;32:362-7.
35. Wang Z, Peng C, Wang P, et al. Serum vitamin D level is related to disease progression in primary biliary cholangitis. Scand J Gastroenterol. 2020;55:1333-40.
36. Zelber-Sagi S, Salomone F, Yeshua H, et al. Non-high-density lipoprotein cholesterol independently predicts new onset of non-alcoholic fatty liver disease. Liver Int. 2014;34:128-35.
37. Wang K, Shan S, Zheng H, Zhao X, Chen C, Liu C. Non-HDL-cholesterol to HDL-cholesterol ratio is a better predictor of new-onset non-alcoholic fatty liver disease than non-HDL-cholesterol: a cohort study. Lipids Health Dis. 2018;17:196.
38. Tutunchi H, Naeini F, Ebrahimi-Mameghani M, Mobasseri M, Naghshi S, Ostadrahimi A. The association of the steatosis severity, NAFLD fibrosis score and FIB-4 index with atherogenic dyslipidaemia in adult patients with NAFLD: a cross-sectional study. Int J Clin Pract. 2021;75:14131.

DOI: 10.4274/forbes.galenos.2025.02360
Forbes J Med 2025;6(2):147-55

Original Article / Orijinal Araştırma

Multi-CNN Deep Feature Fusion and Stacking Ensemble Classifier for Breast Ultrasound Lesion Classification

Meme Ultrasonu Lezyon Sınıflandırması için Multi-CNN Derin Özellik Füzyonu ve Yığınlama Topluluk Sınıflandırıcısı

Kemal PANÇ[1], Sümeyye SEKMEN[2]

[1]Karakoçan State Hospital, Clinic of Radiology, Elazığ, Türkiye
[2]Patnos State Hospital, Clinic of Radiology, Ağrı, Türkiye

Cite as: Panç K, Sekmen S. Multi-CNN deep feature fusion and stacking ensemble classifier for breast ultrasound lesion classification. Forbes Med J. 2025;6(2):147-55

ABSTRACT

Objective: To develop and validate a robust machine learning model for classifying breast ultrasound images into benign, malignant, and normal categories, aiming to enhance diagnostic accuracy using advanced feature extraction and ensemble learning techniques.

Methods: A dataset comprising 2233 images from five public datasets was utilized. After masking regions of interest, deep features were extracted using pre-trained VGG16, ResNet50V2, and EfficientNetB3 models, and concatenated. A multi-step feature selection process involving principal component analysis, recursive feature elimination with LightGBM, and partial least squares discriminant analysis was applied. A stacking ensemble classifier, integrating LightGBM, XGBoost, CatBoost, and random forest with a logistic regression meta-learner, was trained using 5-fold cross-validation on a 75% training set (balanced with synthetic minority oversampling technique), and evaluated on a 25% test set.

Results: The model achieved a macro average area under the curve-receiver operating characteristic (AUC-ROC) of 0.956 and an F1-score of 0.88 on the test set. Benign class results were AUC: 0.984, F1: 0.93, and normal class results were AUC: 0.969, F1: 0.92. The results for the malignant class were AUC: 0.916, F1 score: 0.79. Feature importance analysis showed that ResNet50V2 had the highest contribution to the model's performance.

Conclusion: The proposed approach, combining multi-convolutional neural network deep feature fusion, optimized feature selection, and ensemble stacking, shows significant potential for automated breast ultrasound classification, especially for benign and normal cases. While promising for clinical decision support, the model's lower sensitivity for malignant lesions necessitates further refinement.

Keywords: Breast ultrasound, deep learning, computer-aided diagnosis, ensemble learning, image classification

ÖZ

Amaç: Gelişmiş özellik çıkarma ve topluluk öğrenmesi teknikleri kullanarak tanısal doğruluğu artırmayı hedefleyen, meme ultrason görüntülerinin benign, malign ve normal kategorilerine sınıflandırılması için güçlü bir makine öğrenmesi modeli geliştirmek ve doğrulamaktır.

Yöntem: Beş halka açık veri setinden oluşan 2233 görüntülük bir veri seti kullanılmıştır. İlgili bölgeler maskelendikten sonra, önceden eğitilmiş VGG16, ResNet50V2 ve EfficientNetB3 modelleri kullanılarak derin özellikler çıkarılmış ve birleştirilmiştir. Temel bileşen analizi, LightGBM ile özyinelemeli özellik eleme ve kısmi en küçük kareler ayırt edici analizi içeren çok adımlı bir özellik seçme süreci uygulanmıştır. Lojistik regresyon meta-öğrenicisi ile LightGBM, XGBoost, CatBoost ve random forest'ı entegre eden bir yığınlama topluluk sınıflandırıcısı, %75'lik (sentetik azınlık aşırı örnekleme tekniği ile dengelenmiş) eğitim seti üzerinde *5-fold cross-validation* kullanılarak eğitilmiş ve %25'lik test seti üzerinde değerlendirilmiştir.

Received/Geliş: 19.04.2025
Accepted/Kabul: 19.06.2025
Epub: 17.07.2027
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Corresponding Author/ Sorumlu Yazar:
Kemal PANÇ, MD,
Karakoçan State Hospital, Clinic of Radiology, Elazığ, Türkiye
✉ kemal.panc@gmail.com
ORCID: 0000-0002-3951-7344

Bulgular: Model, test seti üzerinde makro ortalama işlem karakteristik eğrisi altındaki alan (EAA) değeri 0,956 ve F1 skoru 0,88 elde etmiştir. Benign sınıf sonuçları EAA: 0,984, F1: 0,93 ve normal sınıf sonuçları EAA: 0,969, F1: 0,92. Malign sınıf sonuçları EAA: 0,916, F1: 0,79. Özellik önem analizi ResNet50V2'nin en yüksek katkıyı sağladığını göstermiştir.

Sonuç: Çoklu evrişimli sinir ağları derin özellik birleştirme, optimize edilmiş özellik seçimi ve topluluk yığınlamayı birleştiren önerilen yaklaşım, özellikle benign ve normal olgular için otomatik meme ultrason sınıflandırması açısından önemli bir potansiyel göstermektedir. Klinik karar desteği için umut verici olmakla birlikte, modelin malign lezyonlar için daha düşük duyarlılığı, daha fazla iyileştirme gerektirmektedir.

Anahtar Kelimeler: Meme ultrasonu, derin öğrenme, bilgisayar destekli tanı, topluluk öğrenmesi, görüntü sınıflandırma

INTRODUCTION

Breast cancer remains a leading cause of cancer-related mortality in women worldwide.[1] Early detection is critical, as it not only improves survival rates but also leads to more effective treatment options. Ultrasonography has emerged as a key imaging modality in this context owing to its accessibility, low cost, lack of ionizing radiation, and ability to provide real-time visualization of breast tissue architecture.[1,2] However, despite its advantages, ultrasound imaging is inherently operator-dependent, and its image interpretation can be highly challenging, which may result in diagnostic variability.[1] To overcome these limitations, deep learning offers a novel solution for automated lesion classification.

Several studies have demonstrated that deep learning models, including convolutional neural networks (CNNs), can effectively identify and classify regions of interest within ultrasound images by learning hierarchical representations of features.[3-5] For instance, Cao et al.[3] compared multiple deep learning architectures for lesion detection and classification, underscoring the potential of CNNs to delineate lesion boundaries more consistently than manual methods. Similarly, Vigil et al.[4] introduced a dual-purpose deep learning model that concurrently detects and diagnoses breast lesions in ultrasound images, highlighting the benefits of integrated approaches for improving diagnostic consistency.

Advances in feature extraction through discriminative deep learning frameworks further enhance the performance of automated systems. Yu et al.[5] demonstrated that employing deep feature extraction from targeted regions in ultrasound images can lead to improved accuracy in differentiating between benign, malignant, and normal tissues. Moreover, incorporating attention mechanisms, as proposed by Kalafi et al.,[6] helps the model focus on the most diagnostically relevant parts of the image, thereby addressing the ambiguity inherent in ultrasound interpretation. Such strategies may ultimately reduce the incidence of unnecessary biopsies while ensuring high sensitivity in malignancy detection.

The integration of these models not only promises greater diagnostic consistency but also reduces operator variability and enhances clinical decision support systems. Furthermore, as shown by Yap et al.,[2] automated approaches based on CNNs offer scalable solutions that facilitate rapid and reliable lesion detection, potentially contributing to earlier intervention and improved patient outcomes.

This study aimed to develop and validate a machine learning model to classify breast ultrasound images as benign, malignant, or normal, thereby enhancing diagnostic accuracy and supporting radiologists. Our approach is distinguished from previous work through several key innovations: First, we systematically integrate deep features from three complementary CNN architectures (VGG16, ResNet50V2, EfficientNetB3) rather than relying on single architectures. Second, we implement a comprehensive multi-step feature optimization pipeline combining principal component analysis (PCA), recursive feature elimination (RFE) with LightGBM, and partial least squares discriminant analysis (PLS-DA), a more sophisticated approach than typically employed in breast ultrasound classification. Third, we utilize a stacking ensemble methodology that integrates four diverse base learners (LightGBM, XGBoost, CatBoost, random forest) with logistic regression meta-learning, going beyond simple voting or averaging approaches. Most importantly, our model is trained and validated on a robust, heterogeneous dataset created by systematically merging five publicly available collections, representing diverse imaging conditions, patient populations, and clinical settings, addressing the generalizability limitations inherent in single-dataset studies. This comprehensive approach was designed to achieve high accuracy and interpretability while improving classification robustness across diverse clinical scenarios, ultimately supporting radiologists in making accurate diagnoses and improving patient outcomes.

METHODS

The study was conducted using five publicly available and anonymized breast ultrasound datasets. Ethical approval for this specific analysis was waived as it involved secondary use of non-identifiable data.

Data Collection and Preprocessing

Our study utilized a comprehensive dataset created by merging five publicly available breast ultrasound

image collections: Breast Ultrasound Dataset from Universidad de Castilla-La Mancha, breast ultrasound lesion segmentation dataset[7], breast ultrasound images dataset (BUSI)[8], breast ultrasound images database[9], breast ultrasound classification dataset[10], and breast-lesions- ultrasonography dataset.[11] Details of the datasets are shown in Table 1. This approach allowed us to address the limitations of individual datasets while creating a more robust and diverse collection for training our classification model. These lesions were defined as separate cases In the presence of multiple masks belonging to an image containing more than one lesion. All data were classified as benign, malignant, and normal and organised them in separate directories according to their labels with corresponding masks.

Preprocessing was performed using Python 3.8 with OpenCV (version 4.5.5). Images were processed by applying corresponding masks to isolate regions of interest and increase focus on clinically relevant areas. The masked images were resized to a uniform size of 224x224 pixels and converted to red-green-blue colour space, standardized for compatibility with pre-trained deep learning models used in feature extraction. The experiments were conducted on an Apple M4 chip with 16 GB random-access memory, without a dedicated graphics processing unit.

Deep Feature Extraction

Three distinct widely-used CNN architectures, pre-trained on the ImageNet dataset, were selected as feature extractors: VGG16, ResNet50V2, and EfficientNetB3 implemented using TensorFlow 2.10.0. These models were chosen for their proven performance in medical imaging tasks and varying architectural complexities. Models were loaded without their final classification layers, allowing access to the rich, hierarchical feature representations learned during their original training.

Each image was preprocessed (e.g., normalized and scaled) for model compatibility. The CNNs then processed the masked ultrasound images, generating feature representations that captured patterns ranging from low-level textures to high-level semantic features. The outputs from all three models were concatenated into a composite feature vector per image, providing a comprehensive representation of lesion characteristics.

Feature Selection

To address the high dimensionality of the concatenated feature vectors, a multi-step feature selection process was implemented using Scikit-learn 1.0.2. Initially, the feature matrix was standardized using StandardScaler to ensure uniform scaling across features.

Then, PCA was applied to reduce dimensionality while preserving 95% of the variance. This step eliminated redundant and noisy features, transforming the high-dimensional feature vectors (200,704 dimensions) into a more manageable set of principal components, facilitating subsequent analysis.

Next, we used RFE with a LightGBM model (version 3.3.2) to select the top 50 features based on their importance scores. RFE iteratively removes the least significant features, ensuring that only the most informative features

Table 1. Composition of the breast ultrasound datasets

	BUS-UCLM[7]	Breast ultrasound dataset (BUSI)[8]	Breast ultrasound images database[9]	BUSC dataset[10]	Breast-lesions-USG[11]
Total images	683	780	232	250	256
Normal	419	133	0	0	4
Benign	174	437	109	100	154
Malignant	90	210	123	150	98
Number of patients	38	600	Not specified	Not specified	256
Number of radiologists	2	Not stated	1	Not stated	4
Ultrasound scanner	Siemens ACUSON S2000TM	GE LOGIQ E9 and LOGIQ E9 Agile ultrasound system	AixPlorer Ultimate, Supersonic Imagine ultrasound machine	Not stated	• Hitachi ARIETTA 70 • Esaote 6150 • Samsung RS85 • Philips Affiniti 70G and EPIQ 5G
Image file format	.png	.png	.bmp (image), .tif (mask)	.png	.png
Histopathological confirmation	Yes	Not stated	Yes	Not stated	Yes

BUS-UCLM: Breast ultrasound dataset from Universidad de Castilla-La Mancha, BUSC: Breast ultrasound classification dataset, USG: Ultrasonography

are retained. This step refined the feature set by focusing on those most relevant to the classification task.

Finally, we applied PLS-DA to the selected 50 features to project them into a lower-dimensional space optimized for class separation. Unlike PCA, which maximizes variance, PLS-DA prioritizes features that maximize the distinction between benign, malignant, and normal classes, enhancing the discriminative power of the feature set for the stacking classifier.

Following feature selection, we prepared the dataset for model training to ensure robust performance.

Data Splitting and Model Training

The dataset, comprising 2233 breast ultrasound images (1005 benign, 672 malignant, 556 normal), exhibited class imbalance, with benign images constituting 45.0%, malignant 30.1%, and normal 24.9% of the total. We split the dataset into training (75%) and test (25%) sets using stratified sampling to maintain these class proportions. This resulted in a training set of 1675 images (754 benign, 504 malignant, 417 normal) and a test set of 558 images (251 benign, 168 malignant, 139 normal). To address the class imbalance in the training set, where the normal and malignant classes were underrepresented compared to the benign class, we applied the synthetic minority oversampling technique (SMOTE). SMOTE generated synthetic samples for the minority classes (malignant and normal), balancing the training set while preserving the original data distribution in the test set for unbiased evaluation.

We developed a stacked ensemble classifier using four base models: LightGBM (version 3.3.2), XGBoost (version 1.6.2), CatBoost (version 1.0.6), and random forest (scikit-learn 1.0.2). These models were trained on selected features to leverage their unique decision boundaries. Then, their predictions were integrated using logistic regression as a meta-learner, with 5-fold cross-validation ensuring robustness during training. Model hyperparameters are shown in Table 2.

Model performance is assessed on the test set through multiple metrics. A confusion matrix evaluates classification accuracy per class, identifying potential misclassifications. Receiver operating characteristic (ROC) curves are plotted for each class, with area under the curve (AUC) scores calculated to quantify discriminatory ability. Model pipeline is summarized in Figure 1.

RESULTS

The dataset consisted of 2233 breast ultrasonography images, categorized as benign (1005), malignant (672), and normal (556). All images were successfully processed, with features extracted from the regions of interest defined by their corresponding masks. Feature extraction was performed using pre-trained VGG16, ResNet50V2, and EfficientNetB3 models, and the resulting features were concatenated to form a high-dimensional feature vector of 200,704 dimensions for each image. Dimensionality reduction was then applied using PCA to retain features explaining 95% of the variance, followed by exploration of class-discriminative dimensionality reduction via PLS-DA (Figure 2).

We evaluated the model's performance using metrics such as AUC-ROC, AUC-precision-recall (PR), F1-score, precision, and recall for each class, plus macro averages (Table 3). Notably, the system yielded high precision and recall for the benign and normal classes, with particularly strong performance in identifying normal cases, as evidenced by a recall of 1.000.

The model achieved high AUC scores: 0.984 for benign, 0.916 for malignant, and 0.969 for normal classes (Figure 3). The benign class achieved the highest AUC of 0.984, with a sharp curve near the top-left corner, indicating excellent sensitivity and specificity with minimal false positives. The normal class followed with an AUC of 0.969, reflecting strong performance consistent with its perfect recall. The malignant class had an AUC of 0.916, with a more gradual curve suggesting a higher balance between sensitivity and specificity due to the complexity of identifying malignant lesions. These high AUC scores highlight the model's effectiveness, especially for benign and normal cases. Predictive reliability was further assessed using positive predictive value (PPV) and negative predictive value (NPV), yielding strong results, for benign (PPV: 0.924, NPV: 0.949), malignant (PPV: 0.875, NPV: 0.899), and normal (PPV: 0.863, NPV: 1.000). The F1 scores, which balance precision and recall, were 0.93 for benign, 0.79 for malignant, and 0.92 for normal, with a macro average of 0.88.

Table 2. Hyperparameters of the base models

Model	Key parameters
LightGBM	num_leaves=63, max_depth=15, n_estimators=50
Random forest	max_depth=20, n_estimators=200
XGBoost	eval_metric='logloss'
CatBoost	verbose=0
Stacking	cv=5, final_estimator=LogisticRegression(max_iter=1000)

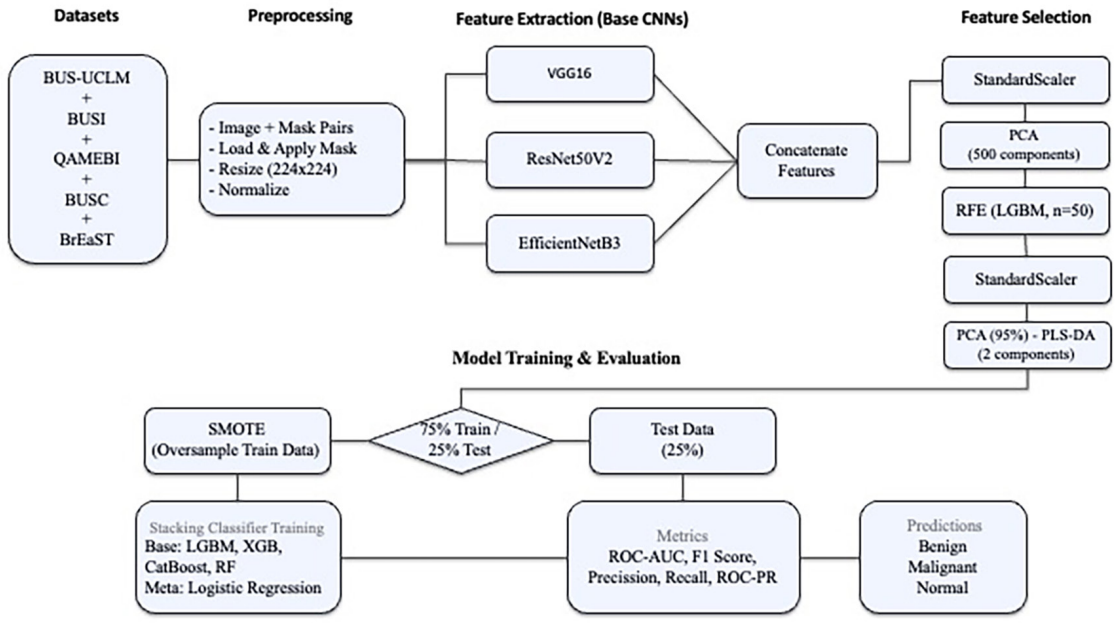


Figure 1. Summary of the model pipeline

BUS-UCLM: Breast ultrasound dataset from Universidad de Castilla-La Mancha, BUSI: Breast ultrasound images dataset, QAMEBI: Breast ultrasound images database, BUSC: Breast ultrasound classification, BrEaST: Breast-lesions-ultrasonography dataset, PCA: Principal component analysis, RFE: Recursive feature elimination, LGBM: Light gradient boosting machine, PLS-DA: Partial least squares discriminant analysis, SMOTE: Synthetic minority oversampling technique, XGB: eXtreme gradient boostin, RF: Random forest, ROC-AUC: Receiver operating characteristic-area under the curve, PR: Precision-recall

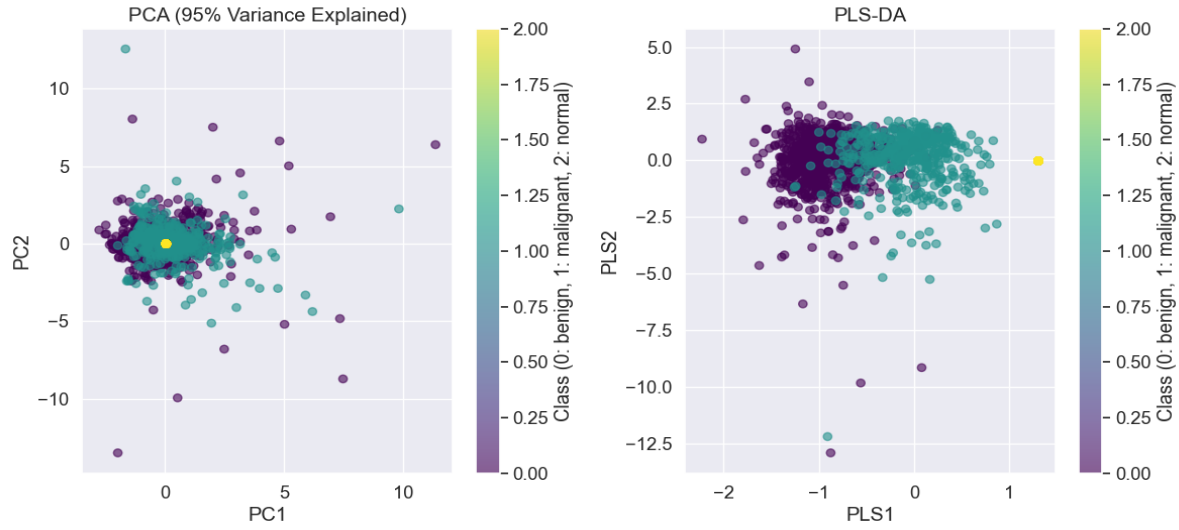


Figure 2. Comparison of dimensionality reduction techniques applied to the selected and scaled image features. (Left) PCA projection onto the first two principal components (PC1, PC2), capturing 95% of the total variance in the features used for classification. (Right) PLS projection onto the first two latent variables (PLS1, PLS2), derived specifically to maximize the separation between classes based on the same feature set

PCA: Principal component analysis, PLS-DA: Partial least squares discriminant analysis

Table 3. Performance metrics of the stacking classifier on the test set					
Class	**AUC-ROC**	**AUC-PR**	**F1-score**	**Precision**	**Recall**
Benign	0.984	0.98	0.93	0.92	0.93
Malignant	0.916	0.83	0.79	0.87	0.71
Normal	0.969	0.86	0.92	0.86	1.00
Macro Avg.	0.956	0.89	0.88	0.88	0.88
AUC-ROC: Area under the curve-receiver operating characteristic, PR: Precision-recall, Avg.: Average					

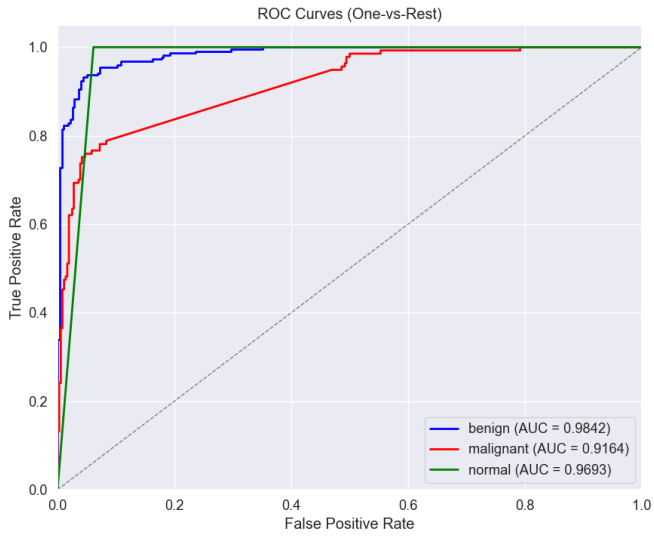


Figure 3. ROC curves for benign (AUC=0.9842), malignant (AUC=0.9164), and normal (AUC=0.9693) classes on the test set

ROC: Receiver operating characteristic, AUC: Area under the curve

The high F1 scores for benign and normal classes reflect the model's ability to achieve both high precision and recall for these classes, while the lower F1 score for the malignant class (0.79) indicates a challenge in balancing precision and recall.

The model's performance was further assessed using PR curves (Figure 4), which show the balance between precision and recall for each class. The PR curves show the model's ability to maintain high precision across varying recall levels. The benign class achieved the highest average precision (AP) score of 0.9803, reflecting the model's strong performance in correctly identifying benign cases with minimal false positives, as indicated by the curve maintaining high precision even at high recall values. The malignant class had an AP score of 0.8305, with the curve showing a more noticeable decline in precision as recall increases, suggesting a trade-off due to the complexity of distinguishing malignant lesions. The normal class, with an AP score of 0.8634, demonstrated a stable PR trade-off,

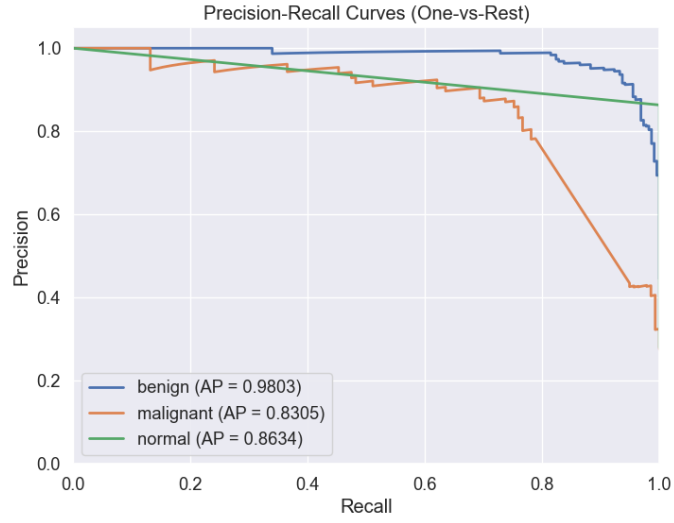


Figure 4. Precision-recall curves for benign (AP=0.9803), malignant (AP=0.8305), and normal (AP=0.8634) classes on the test set

AP: Average precision

reflecting the smaller sample size of normal cases in the dataset, though with a slightly steeper drop in precision at higher recall compared to the benign class.

A summary of the confusion matrix on the test set (251 benign, 168 malignant, 139 normal) reveals the following: 237 benign correctly predicted, with 15 misclassified as malignant and 0 as normal; 111 malignant correctly predicted, with 34 misclassified as benign and 23 as normal; and 139 normal correctly predicted, with 0 misclassified as benign or malignant. This corresponds to 0 false negatives for the benign and normal classes, and 57 false negatives in total for the malignant class (34 as benign, 23 as normal), representing 33.9% of malignant cases. The most frequent confusions occurred in malignant cases, where 34 were misclassified as benign and 23 as normal, highlighting a challenge in distinguishing malignant lesions from other classes.

The feature importance analysis of the 50 selected features reveals a clear contribution ranking among the models: ResNet50V2, VGG16, and EfficientNetB3 (Figure 5). ResNet50V2 leads with a total importance of 3,467 across

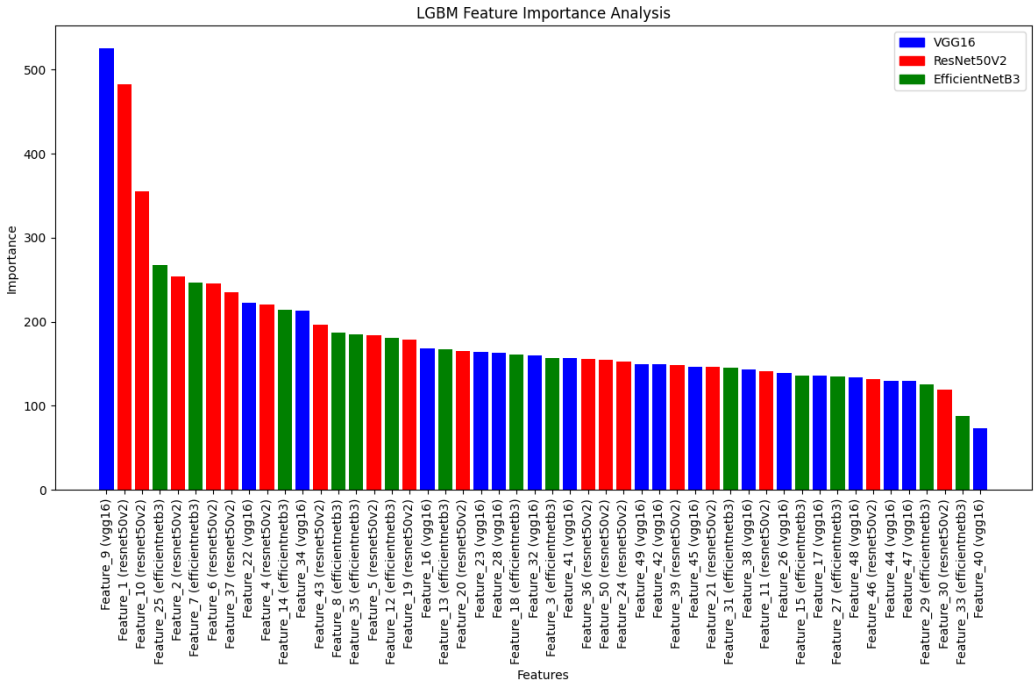


Figure 5. Feature importance analysis of the top 50 selected features from ResNet50V2, VGG16, and EfficientNetB3 models

LGBM: Light gradient boosting machine

18 features, yielding the highest average importance per feature, at 192.61. VGG16 follows closely in importance with a total score of 3,103, contributing 18 features and an average importance per feature of 172.39. EfficientNetB3 ranks third, with a total importance of 2,395 across 14 features, resulting in an average importance per feature of 171.07. These results highlight ResNet50V2's dominant influence in the LightGBM model.

DISCUSSION

Our study successfully developed and evaluated a machine learning model aimed at classifying breast ultrasound images into benign, malignant, and normal categories by leveraging deep features extracted from multiple pre-trained CNNs (VGG16; ResNet50V2; EfficientNetB3), employing sophisticated feature selection methods, and utilizing a stacking classifier. The model was trained and evaluated on a comprehensive dataset aggregated from five distinct public collections; this final dataset represents a heterogeneous population across multiple medical centers, enhancing the potential generalizability of our model for breast lesion classification in diverse clinical settings, a crucial aspect highlighted by studies like Gu et al.,[12] which demonstrated the value of large multi-center datasets. The findings presented in the results section indicate a robust overall performance, highlighted by a macro average AUC-ROC of 0.956 and an F1-score of 0.88 on the unseen test data. The model demonstrated particularly high efficacy in classifying benign lesions (AUC 0.984, F1 0.93) and normal tissue (AUC 0.969, F1 0.92), achieving perfect recall for the normal class. Despite the model's overall strong performance, classifying malignant lesions proved more challenging, as is evidenced by lower metrics (AUC: 0.916; F1-score: 0.79) compared to benign and normal classes.

The model's overall success can be attributed to several key factors inherent in the methodology. Firstly, the extraction of deep features using transfer learning from three powerful, pre-trained CNNs (VGG16; ResNet50V2; EfficientNetB3) provided rich, hierarchical representations of the ultrasound images. Concatenating features from these diverse architectures likely created a more comprehensive feature pool than relying on a single network, capturing a wider range of patterns relevant to classification. For example, Cao et al.[3] achieved 87.5% accuracy using DenseNet on 1043 images with binary classification, while Ellis et al.[13] reported 77.77% accuracy for ResNet50 and 73.80% for VGG-19 on 3-class classification with 571 images. However, direct performance comparison is inappropriate due to different dataset sizes (our dataset of 2,233 images vs. their smaller datasets); preprocessing methods, and validation approaches.

Feature importance analysis confirmed the contribution of all three networks, with ResNet50V2 features showing the highest aggregate importance in the final LightGBM selection step. Secondly, the use of a stacking ensemble classifier, integrating predictions from LightGBM, XGBoost, CatBoost, and random forest via a logistic regression meta-learner, leveraged the strengths of multiple algorithms, while compensating for individual model weaknesses and enhancing predictive robustness and generalization. Ensemble methods, like the one used by Ragab et al.[14] (achieving 97.52% accuracy with VGG-16/19/SqueezeNet + multilayer perceptron with 780 images), often enhance predictive robustness and generalization. Thirdly, employing SMOTE during training helped mitigate the inherent class imbalance in the dataset, likely contributing to the strong performance observed, particularly for the benign and normal classes, which achieved excellent precision, recall, and AUC scores.

However, while the model demonstrated high precision (0.87) for the malignant class, indicating that positive predictions for malignancy are likely correct, the recall (sensitivity) of 0.71 presents a significant concern from a clinical perspective. This recall value implies that approximately 28-29% of actual malignant cases in the test set were misclassified as benign or normal (false negatives). This contrasts sharply with some studies reporting very high recall/sensitivity, such as Yadav et al.[15] who achieved 98.55% overall recall and 90.32% malignant recall using a modified ResNet-101 with 780 images, and Ragab et al.[14] reporting 96.01% overall sensitivity with their ensemble. Even Kalafi et al.,[6] who used an Attention-VGG16 for binary classification with 439 images, reported 96% sensitivity. In clinical practice, failing to detect malignancy has far more severe consequences than misclassifying a benign lesion as malignant, (false positive). The 29% false-negative rate for malignant cases could lead to delayed diagnoses, allowing disease progression that may result in advanced-stage cancer, increased mortality risk, and reduced treatment efficacy. Our lower sensitivity might stem from the inherent subtlety of some malignant lesions, potential limitations of SMOTE for this complex class, or the aggressive feature selection potentially removing crucial subtle features. Future improvements could focus on addressing class imbalance or incorporating additional features to better distinguish malignant characteristics.

Our approach aligns with trends in the literature that utilize deep learning for breast ultrasound analysis.[3-6] We extended the deep feature extraction concept used by Yu et al.,[5] by fusing features from three distinct architectures rather than focusing on specific regions within one architecture (Inception-V3). Our model achieved strong performance, particularly for the benign (AUC: 0.984) and normal (AUC: 0.969) classes. While direct comparison is challenging due to variations in datasets, preprocessing, metrics, and architectures across studies, these results are competitive with or exceed those reported previously (e.g., Zhang et al.[1] 90% AUC using Breast Imaging Reporting and Data System features; Vigil et al.[4] 78.5% accuracy using autoencoder/radiomics; Cao et al.[3] 87.5% accuracy using DenseNet for ternary classification; Gu et al.[12] 0.91 AUC using VGG19 for binary prediction). However, achieving the near-perfect scores reported by Jabeen et al.[16] (99.1% accuracy with augmented BUSI dataset using pre-trained DarkNet-53) or Kiran et al.[17] (100% accuracy with EfficientKNN on a small 780-image dataset) remains challenging, and may depend heavily on their reliance on extensive data augmentation and smaller, potentially less diverse datasets, which may not generalize as effectively to our ternary classification task across a larger, multi-center dataset of 2233 images.

Study Limitations

Several limitations should be acknowledged in this study. The multi-step feature selection, while necessary, might have discarded valuable information; the clinical relevance of the final 50 features needs further investigation. Although the dataset was compiled from multiple sources to enhance diversity, the model's performance was evaluated only on an internal test split. External validation on completely independent datasets from different institutions and ultrasound machines is crucial to assess its generalizability.

Future research should prioritize enhancing the sensitivity (recall) for malignant lesion detection. This may involve exploring cost-sensitive learning algorithms that better address the misclassification of malignant cases, experimenting with alternative data augmentation techniques techniques, advanced oversampling methods (e.g., ADASYN, class weighting within models), or optimizing feature selection. Investigating attention mechanisms (as in Kalafi et al.[6] or Lyu et al.[18] for segmentation) within the feature extractors, exploring architectures known for high performance, such as advanced ResNet and variants, could yield improvements. Investigating alternative approaches, such as end-to-end deep learning models that learn features and classify directly without explicit feature extraction/selection steps, or autoencoders for dimensionality reduction, could also be beneficial.

CONCLUSION

In conclusion, this study demonstrates the considerable potential of combining deep feature extraction from multiple CNNs with advanced feature selection and ensemble learning techniques for classifying breast ultrasound images. This approach represents a promising

frontier in breast cancer diagnostics. The developed model achieved high overall accuracy and discriminatory power, particularly excelling in the classification of benign and normal cases on a diverse, multi-center dataset. These results underscore the model's efficacy and potential as a robust tool for clinical decision support. By mitigating challenges associated with operator dependency and subjective interpretation, such automated methods can offer a reproducible approach to enhance early detection and potentially reduce unnecessary invasive procedures. However, the critical challenge of lower sensitivity (recall) for malignant lesions must be addressed through further research and refinement. While promising as a component of a computer-aided diagnosis system to support radiologists and enhance consistency, this model requires significant improvements in malignant detection and rigorous external validation are essential before it can be reliably integrated into clinical workflows to ultimately support improved patient outcomes.

Ethics

Ethics Committee Approval and Informed Consent: This study used anonymized, publicly available datasets with no personally identifiable information. As it involved no human interaction or new data collection, neither ethics committee approval nor patient consent was required.

Footnotes

Authorship Contributions

Concept: K.P., Design: K.P., Data Collection or Processing: K.P., S.S., Analysis or Interpretation: K.P., S.S., Literature Search: K.P., S.S., Writing: K.P., S.S.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Zhang E, Seiler S, Chen M, Lu W, Gu X. BIRADS features-oriented semi-supervised deep learning for breast ultrasound computer-aided diagnosis. Phys Med Biol. 2020;65:125005.
2. Yap MH, Pons G, Martí J, et al. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J Biomed Health Inform. 2018;22:1218-26.
3. Cao Z, Duan L, Yang G, Yue T, Chen Q. An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. BMC Med Imaging. 2019;19:51.
4. Vigil N, Barry M, Amini A, et al. Dual-intended deep learning model for breast cancer diagnosis in ultrasound imaging. Cancers (Basel). 2022;14:2663.
5. Yu H, Sun H, Li J, et al. Effective diagnostic model construction based on discriminative breast ultrasound image regions using deep feature extraction. Med Phys. 2021;48:2920-8.
6. Kalafi EY, Jodeiri A, Setarehdan SK, et al. Classification of breast cancer lesions in ultrasound images by using attention layer and loss ensemble in deep convolutional neural networks. Diagnostics. 2021;11:1859.
7. Vallez N, Bueno G, Deniz O, Rienda MA, Pastor C. BUS-UCLM: breast ultrasound lesion segmentation dataset. Mendeley Data, V1. 2024. Available from: https://data.mendeley.com/datasets/7fvgj4jsp7/1
8. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data Brief. 2020;28:104863.
9. Abbasian Ardakani A, Mohammadi A, Mirza-Aghazadeh-Attari M, Acharya UR. An open-access breast lesion ultrasound image database: applicable in artificial intelligence studies. Comput Biol Med. 2023;152:106438.
10. Iqbal A. BUSC Dataset. Mendeley Data, V1. 2023. Available from: https://data.mendeley.com/datasets/vckdnhtw26/1
11. Pawłowska A, Ćwierz-Pieńkowska A, Domalik A, et al. Curated benchmark dataset for ultrasound based breast lesion analysis. Sci Data. 2024;11:148.
12. Gu Y, Xu W, Lin B, et al. Deep learning based on ultrasound images assists breast lesion diagnosis in China: a multicenter diagnostic study. Insights Imaging. 2022;13:124.
13. Ellis J, Appiah K, Amankwaa-Frempong E, Kwok SC. Classification of 2D ultrasound breast cancer images with deep learning. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2024;5167-73.
14. Ragab M, Albukhari A, Alyami J, Mansour RF. Ensemble deep-learning-enabled clinical decision support system for breast cancer diagnosis and classification on ultrasound images. Biology (Basel). 2022;11:439.
15. Yadav A, Kolekar M, Zope M. ResNet-101 Empowered Deep Learning for Breast Cancer Ultrasound Image Classification. 2024;763-9.
16. Jabeen K, Khan MA, Alhaisoni M, et al. Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion. Sensors (Basel). 2022;22:807.
17. Kiran A, Ramesh JVN, Rahat IS, Khan MAU, Hossain A, Uddin R. Advancing breast ultrasound diagnostics through hybrid deep learning models. Comput Biol Med. 2024;180:108962.
18. Lyu Y, Xu Y, Jiang X, Liu J, Zhao X, Zhu X. AMS-PAN: breast ultrasound image segmentation model combining attention mechanism and multi-scale features. Biomed Signal Process Control 2023;81:104425.

DOI: 10.4274/forbes.galenos.2025.25338
Forbes J Med 2025;6(2):156-64

Orijinal Araştırma / Original Article

Böbrek Nakli Sonrası Çocuk Hastalarda Glukoz Metabolizması ve Değişiklikleri

Glucose Metabolism Alterations After Kidney Transplantation in Pediatric Patients

Hayrullah MANYAS[1], Caner ALPARSLAN[2], Fatma MUTLUBAŞ[3], Demet ALAYGUT[3], Belde KASAP-DEMİR[4], Önder YAVAŞCAN[5]

[1]Şanlıurfa Eğitim ve Araştırma Hastanesi, Çocuk Endokrinolojisi Kliniği, Şanlıurfa, Türkiye
[2]İzmir Demokrasi Üniversitesi Tıp Fakültesi, Çocuk Nefrolojisi Anabilim Dalı, İzmir, Türkiye
[3]Sağlık Bilimleri Üniversitesi, İzmir Tepecik Eğitim ve Araştırma Hastanesi, Çocuk Nefrolojisi Kliniği, İzmir, Türkiye
[4]İzmir Katip Çelebi Üniversitesi Tıp Fakültesi, Çocuk Nefrolojisi Anabilim Dalı, İzmir, Türkiye
[5]İstanbul Medipol Üniversitesi, Çocuk Nefrolojisi Anabilim Dalı, İstanbul, Türkiye

Atıf: Manyas H, Alparslan C, Mutlubaş F, Alaygut D, Kasap-Demir B, Yavaşcan Ö. Glucose metabolism alterations after kidney transplantation in pediatric patients. Forbes J Med. 2025;6(2):156-64

ÖZ

Amaç: Nakil sonrası dönemde, çoğunlukla immünosüpresif ajanlar ve hızlı kilo alımı nedeniyle glukoz metabolizması bozuklukları gelişebilir. Bu bozukluklar, insülin direnci, bozulmuş glukoz toleransı ve nakil sonrası ortaya çıkan diabetes mellitus olarak karşımıza çıkar. Değişen glukoz metabolizması kardiyovasküler komplikasyonlara, greft sorunlarının kötüleşmesine ve farklı morbiditelere neden olabilir. Bu durumların erken tespiti ve uygun tedavi planı, risk ve komplikasyon oranlarını azaltmak için son derece önemlidir. Bu çalışmada, çocuklarda nakil öncesi ve sonrası dönem arasındaki glukoz metabolizması değişikliklerini değerlendirmeyi amaçladık.

Yöntem: Bu çalışma, kliniğimizde böbrek nakli olan 36 çocuk hastanın klinik ve laboratuvar bulgularının retrospektif değerlendirilmesiyle yapıldı. Nakilden önceki son laboratuvar değerlendirmesi ve ayrıca nakil sonrası dönemdeki klinik ve laboratuvar parametrelerinin yıllık değerlendirmesi protokolümüze göre oluşturulan değerlendirme kayıtlarından alındı.

Bulgular: Beş yıllık izlem sonucunda, 9 (%25) hastada insülin-glukoz metabolizma bozukluğu (2 hasta nakil sonrası diabetes mellitus, 3 hasta insülin direnci, 4 hasta glukoz tolerans bozukluğu) tanısı aldı. Nakil sonrası diabetes mellitus, insülin direnci ve bozulmuş glukoz toleransı gelişen hastalarda; yaşın, geçmişinde periton diyalizi tedavisi uygulamış olmanın, perioperatif hiperglisemi nedeniyle insülin ihtiyacı duymuş olmanın, yüksek kreatinin düzeylerine sahip olmanın, yüksek vücut kitle indeksine sahip olmanın ve nakil sonrası puberteye girmenin insülin-glukoz metabolizması bozukluğu gelişme riski ile pozitif korelasyon gösterdiği saptandı ($p<0,05$).

Sonuç: Sonuç olarak, pediatrik böbrek nakli hastalarında glukoz metabolizmasında değişiklik (nakil sonrası diabetes mellitus, insülin direnci, glukoz intoleransı) önemli bir morbidite nedenidir. Glukoz metabolizması bozukluğu için risk değerlendirmesinin nakilden önce başlatılması ve nakilden sonra yıllık olarak devam ettirilmesi gerekmektedir.

Anahtar Kelimeler: Diabetes mellitus, glukoz intoleransı, insülin direnci, böbrek nakli, çocukluk çağı

ABSTRACT

Objective: In post-transplant period, glucose metabolism is altered mostly due to immunsupressive agents and fast weight gain. Insulin resistance, impaired glucose tolerance test results and new onset diabetes after transplantation are detected pathologies. Moreover, this altered glucose metabolism could be a cause of cardiovascular complications, deterioration of graft problems and morbidities.

Geliş/Received: 06.05.2025
Kabul/Accepted: 19.06.2025
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Sorumlu Yazar/ Corresponding Author:
Dr. Hayrullah MANYAS,
Şanlıurfa Eğitim ve Araştırma Hastanesi, Çocuk Endokrinolojisi Kliniği, Şanlıurfa, Türkiye
hayrullahmanyas@hotmail.com
ORCID: 0000-0002-4775-2950

Early detection and appropriate treatment plan have importance to reduce risk and complication rate. In this study, we aimed to assess glucose metabolism changes in children between pre-and post-transplantation period.

Methods: We conducted a retrospective study with the clinical and laboratory findings of 36 pediatric kidney transplant patients in our clinic. Last laboratory assesment before transplantation and also, clinical and laboratory parameters in post-transplant period were taken from evaluation records according to our annually assessment protocol.

Results: In 9 (25%) patients have altered glucose metabolism (new onset diabetes mellitus after transplantation in 2 patients, insulin resistance in 3 patients, glucose intolerance in 4 patients). There was a positive corelation between age, being on chronic periton dialysis programme, the need for insulin treatment in perioperative and first 24 hours in post-transplant period, low glomerular filtration rate, high body mass index, begining of puberty and new onset diabetes mellitus after transplantation, insulin resistance and glucose intolerance ($p<0.05$).

Conclusion: In conclusion, altered glucose metabolism in pediatric kidney transplant patients is an important co-morbidity. We believe that risk assessment for glucose metabolism disorders should be start before transplantation and continued annually after transplantation.

Keywords: Diabetes mellitus, glucose intolerance, insulin resistance, kidney transplantation, pediatric

GİRİŞ

Son dönem böbrek yetmezliği (SDBY) olan çocuklarda glukoz metabolizmasında görülen bozukluklar sık rastlanan bir problem haline gelmiştir ve kronik böbrek hastalığının her döneminde görülmektedir. İnsülin sekresyon bozukluğu, insülin direnci ve beta hücre hasarı bu durumda etkin rol oynayan mekanizmalardır.[1] Nakil sonrası dönemde de, çoğunlukla immünosüpresif ajanların kullanımı (örneğin; kortikosteroidler, kalsinörin inhibitörleri) ve hızlı kilo alımı nedeniyle glukoz metabolizması bozuklukları gelişebilir. Özellikle böbrek nakli yapılmış hastalarda diabetes mellitus (DM) gelişimi riskinin arttığı görülmüş, bu durumun majör kardiyovasküler olaylar, enfeksiyon ve artmış mortalite ile olan ilişkisi de kanıtlanmıştır.[2,3] Bu sebeplerden dolayı, glukoz metabolizmasındaki değişiklikleri yakından takip etmek ve artmış kan şekeri düzeylerini erken saptamak, bu hastaların yönetiminde oldukça önemli hale gelmiştir. Periton diyalizi (PD), hemodiyaliz (HD) ve böbrek naklinden oluşan renal replasman tedavi yöntemlerinin glukoz metabolizması üzerine etkileri çeşitli araştırmacılar tarafından incelenmiştir. Bu çalışmaların önemli bir bölümünde DM gelişimi önceden öngörülmeye çalışılmıştır. Bu çalışmalarda glukoz metabolizmasındaki değişiklikleri saptamak amacıyla açlık kan şekeri, HbA1C, tokluk kan şekeri, oral glukoz tolerans testi (OGTT), insülin düzeyi, homeostatic model assesment insülin rezistansı (HOMA-IR) gibi insülin ve glukoz dengesindeki bozuklukları gösteren ve DM tanısını koymada kullanılan laboratuvar tetkiklerinden yararlanılmıştır.[1,2,4-6] Bu çalışmada, böbrek nakli olan çocuklarda böbrek nakli öncesinde ve sonrasında glukoz metabolizmasında oluşmuş veya izlemde gelişmiş bozuklukların klinik ve laboratuvar veriler eşliğinde değerlendirilmesi amaçlanmıştır.

YÖNTEM

Kliniğimizde böbrek nakli olan 18 yaş altı 36 olgunun demografik, klinik ve biyokimyasal verileri böbrek nakli öncesi diyaliz programında ve nakil sonrası yıllık değerlendirmelerle geriye yönelik incelendi ve kaydedildi. Çalışma için Sağlık Bilimleri Üniversitesi, İzmir Tepecik Eğitim ve Araştırma Hastanesi Etik Kurulu'ndan onay alınmıştır (karar numarası: 8, tarih: 29.12.2014). Hastaların antropometrik ölçümleri (ağırlık, boy, bel çevresi) ve kan basıncı kaydedildi. Neyzi verilerine göre ağırlık, boy, vücut kitle indeksi (VKİ) persentilleri değerlendirildi. VKİ ağırlığın boyun metrekaresine bölünmesi ile [ağırlık/boy^2 (kg/m^2)] hesaplandı. VKİ <85. persentil olanlar normal kilolu, VKİ>85-<95. persentil olanlar aşırı kilolu, VKİ >95. persentil olanlar obez kabul edildi.

Hastaların nakil öncesi diyaliz tedavisi alıp almadığı, hangi diyaliz tedavisinin (HD, PD) uygulandığı, nakil yaşı, nakil sonrasında geçen süre, greft tipi (kadavra ve canlı), donör yaşı kaydedildi. Hastaların pubertal dönemde olup olmadıkları yıllık değerlendirmelerle Tanner evrelemesine göre kaydedildi. Hastanın SDBY nedeni ürolojik (vur, obstrüktif üropatiler, taş hastalığı, nörojenik mesane vb.), non-ürolojik nedenler (glomerülonefritler, herediter nefritler vb.) olarak iki grupta değerlendirildi.

Glukoz metabolizmasındaki değişiklikleri öngörmede açlık kan şekeri, açlık insülin düzeyi, OGTT, HOMA-IR parametreleri yine nakil öncesi, sonrası 12. ay ve bunu izleyen 5 yıllık zaman diliminde yıllık olarak değerlendirildi. DM tanısında Dünya Sağlık Örgütü ölçütleri kullanıldı. Açlık kan şekeri değerinin ≥126 mg/dL olması, OGTT'nin 2. saatinde kan glukoz değerinin ≥200 mg/dL olması, semptomu olan hastada herhangi bir zamanda bakılan kan şekerinin ≥200 mg/dL olarak saptanması veya HbA1C değerinin ≥%6,5 saptanması DM olarak tanımlandı.

OGTT 1,75 gr/kg glukozun oral alımı sonrası 0.-120. dakikalarda kan şekeri düzeylerinin ölçümü yapılarak değerlendirildi. Bozulmuş glukoz toleransı açlık plazma glukozunun 100-125 mg/dL arasında olması, OGTT'de 2. saat glukoz düzeyinin 140-200 mg/dL arası olması olarak tanımlandı.

İnsülin direncini değerlendirmek amacıyla HOMA-IR (açlık kan şekeri x açlık kan insülin düzeyi/405) değeri kullanıldı. İnsülin direnci HOMA-IR düzeyinin yüksek saptanması

olarak tanımlandı. Prepubertal hastalarda >2,5 ve pubertal hastalarda >3,5 değerler yüksek olarak kabul edildi.

Steroid tedavisine ek olarak aldıkları immünosüpresif tedaviler kayıt altına alındı. Steroidin hiperglisemik etkisi düşünülürek hastalarımızın aylık kümülatif steroid dozu mg/kg/ay olarak not edildi. Siklosporin ve takrolimus kullanan hastaların ortalama ilaç kan düzeyi yıllık olarak hesaplandı.

Hastalar sitomegalovirüs (CMV) enfeksiyonu açısından nakil öncesi, nakil sonrası 12. ay ve sonrasında yıllık olarak polimeraz zincir reaksiyonu ile CMV-deoksiribonükleik asit tetkikleri kullanılarak araştırıldı ve 100/mL kopyanın üstündeki değerler pozitif olarak değerlendirildi. Tanı biyopsi ile desteklendi.

Büyük çocuklarda ve adölesanlarda serum kreatinin seviyesinin bazal düzeyinden 0,3 mg/dL, küçük çocuklarda ise %20 artış göstermesi akut rejeksiyon kabul edildi. Akut rejeksiyon kabul edilip yoğunlaştırılmış immünosüpresif tedavilere maruz kalan hastaların akut rejeksiyon sayıları kayıt altına alındı. Akut rejeksiyon tanısı alan hastalara tanı biyopsi ile konuldu.

Testler sonucunda glukoz tolerans bozukluğu, insülin direnci, DM tanısı alan hastaların tamamı insülin-glukoz metabolizma bozukluğu olarak tanımlandı.

İstatistiksel Analiz

İstatistiksel analizler IBM SPSS 20.0 (SPSS, Chicago, Illinois, USA) programı kullanılarak yapıldı. Normal dağılım gösteren değişkenler için gruplar arası karşılaştırmalarda, bağımsız gruplarda Student's t-test kullanıldı ve tanımlayıcı istatistiksel ortalama ± standart sapma biçiminde verildi. Kategorik olmayan verilerin yıllar arasındaki anlamlılık araştırmaları için Friedman's two-way analizi kullanıldı. Kategorik verilerin karşılaştırılmasında Binomiyal test ve Pearson ki-kare testi kullanıldı. Kategorik verilerin periyotlar arasındaki farklılığını araştırmak için Kruskal-Wallis testi kullanıldı. Glukoz metabolizma bozukluğu olarak tanımlanan hastaların öngörülmesi amacıyla tüm risk faktörleri üzerinden korelasyon analizi uygulandı. P<0,05 değeri istatistiksel olarak anlamlı kabul edildi.

BULGULAR

Çalışmaya 17'si erkek (%47,2), 19'u kız (%52,8) olmak üzere toplam 36 hasta dahil edildi. Hastaların nakil öncesi 24'ünün (%67) PD, 8'inin (%22) HD tedavisi ile izlendiği, 4 (%11) olgunun ise diyaliz tedavisi almadan böbrek nakli olduğu (preemptif nakil) belirlendi. Diyaliz tedavisi almış olan hastaların ortalama diyalizde kalma süreleri ortanca değer 12 ay, tüm hastaların ortalama nakil olma yaşları ise ortanca değer 12 yaş olarak saptandı. Hastaların 15'inin (%42) kadavradan, 21'inin ise (%58) canlıdan nakil olduğu saptandı. Hastaların nakil sonrası ortalama izlem süresi ortanca değer 41 ay olarak bulundu. Nakil sonrası hastalarda kullanılan immünsüpresyon rejimleri değerlendirildiğinde 3 (%8,3) hastanın içeriğinde siklosporin bazlı, 33 (%91,7) hastanın ise içeriğinde takrolimus bazlı rejimleri kullandığı saptandı. Ailesinde DM tanısı olan hasta belirtilmedi. Nakil sonrası ilk dönemde eve gidene kadar olan dönemde hiperglisemi nedeniyle insülin tedavisi alan 11 (%30,5) hasta saptandı (Tablo 1). Nakil sonrası insülin-glukoz metabolizması bozukluğu toplamda 9 (%25) hastada saptandı. Bu hastalardan 2'sinde (%5,6) böbrek nakli sonrası gelişen diabetes mellitus (NSDM), 3'ünde (%3,8) insülin direnci, 4'ünde (%1,11) glukoz tolerans bozukluğu saptandı (Şekil 1).

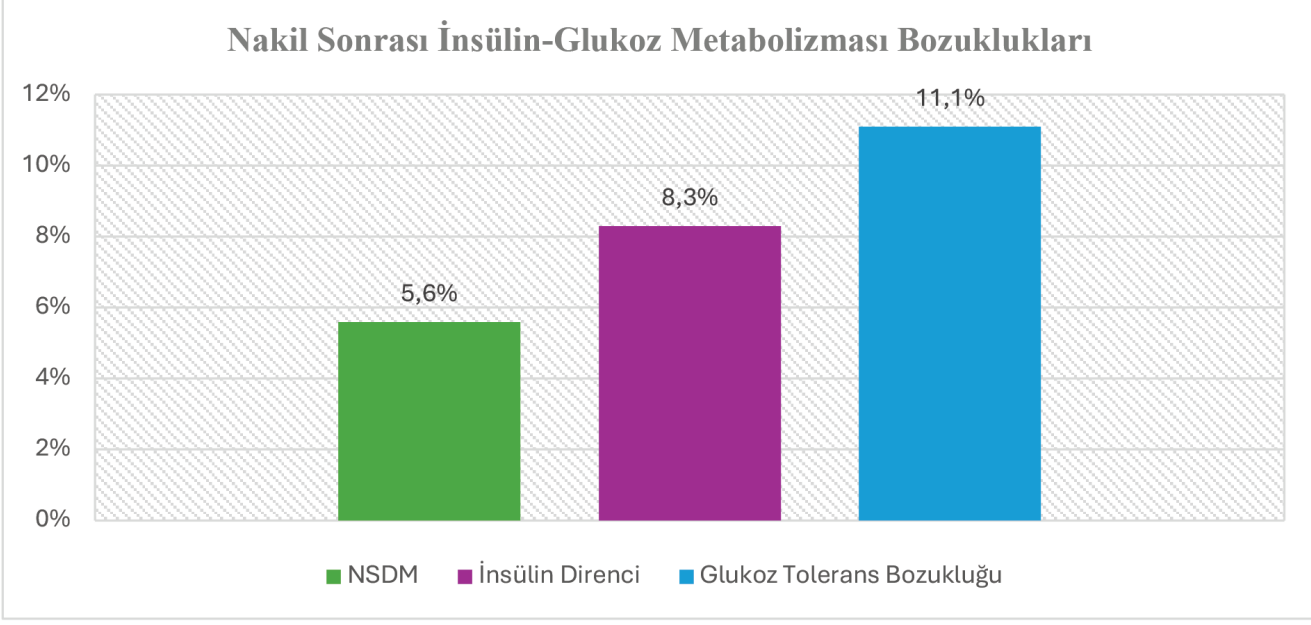
Nakil öncesi hastaların hiçbiri obez veya aşırı kilolu olarak tanımlanmadı. Hastalar insülin-glukoz metabolizması bozuklukları açısından değerlendirildiklerinde nakil öncesi yalnız 1 (%2,7) hastada glukoz tolerans bozukluğu, bu hastadan farklı 1 (%2,7) hastada da insülin direnci saptandı. Hiçbir hastada DM saptanmadı.

Hastalar 1. yılında insülin-glukoz metabolizma bozuklukları açısından değerlendirildiğinde 2 hasta (%6,6) NSDM tanısı aldı. Bu hastalara tedavi başlanıp, hastalar izleme alındı. Nakil öncesi insülin direnci saptanan bir hastaya ikincisi de eklenerek nakil sonrası 1. yılda insülin direnci olan hasta sayısı 2'ye çıktı. Nakil öncesinde glukoz tolerans bozukluğu saptanan 1 hasta bulunurken, nakilden sonraki 1. yılda glukoz tolerans bozukluğu gelişen yeni bir hasta daha saptandı ve böylece glukoz tolerans bozukluğu olan hasta sayısı 2'ye yükseldi. İnsülin-glukoz metabolizma bozukluğu saptanan bu 6 hastadan 2'si aşırı kilolu olarak değerlendirildi.

Hastaların 26'sının (%72,2) böbrek nakli sonrası 2. yılını doldurduğu belirlendi. İkinci yılda 2 (%7,7) hasta obez olarak kabul edildi. Fazla tartılı hasta sayısı 2 hasta ile değişmedi. İkinci yıl içinde yeni NSDM hastası saptanmazken 1. yıla göre 1 yeni insülin direnci ve 2 yeni glukoz tolerans bozukuluğu hastasına tanı konularak 2. yıl sonunda insülin-glukoz metabolizma bozukluğu saptanan toplam hasta sayısı 9 olarak belirlendi. Böylelikle insülin-glukoz metabolizma bozukluğu saptanan tüm hastalara tanıları ilk 2 yılda konmuş oldu (Tablo 2).

Tablo 2'de de hastaların 5 yıllık izlemleri sonucu elde edilen klinik ve laboratuvar verilerin ortalamalarının karşılaştırılmış hali verildi. VKİ değerlerinin tüm izlem sürecinde istatistiksel olarak anlamlı bir artış göstermediği saptandı. Ancak nakil sonrası dönem içerisinde yıllık izlemler ayrı ayrı olarak nakil öncesi dönem ile karşılaştırıldığında tüm periyotlarda VKİ değerleri nakil öncesi döneme göre istatistiksel olarak anlamlı bir biçimde yüksek bulundu (p=0,02).

Tablo 1. Hastaların genel özellikleri	
Hasta sayısı, (n)	36
Erkek/Kız (n), (%)	17/19, 47.2/52.8
Nakil öncesi diyaliz tipi (n), (%)	
PD	24 (67)
HD Preemptif	8 (22) 4 (11)
Diyalizde kalma süresi, (ay)	16,16±1,47/12
Hastaların nakil olma yaşları	11,82±4,7/12
Nakil tipi (n), (%)	
Canlı	21 (58)
Kadavra	15 (42)
Donör yaşı	25,62±16,41/31
Nakil sonrası izlem süresi, (ay)	45,54±26,5/41
İmmünosüpresif ilaç rejimi (n), (%)	
Siklosporin içeren rejim	**3 (8,3)**
Takrolimus içeren rejim	**33 (91,7)**
Nakil sonrası eve çıkana kadar geçici insülin kullanan hasta sayısı, (n), (%)	11 (30,5)
Nakil öncesi ve sonrası insülin-glukoz metabolizma bozukluğu saptanan toplam hasta sayısı (n), (%)	9 (25)
NSDM	2 (5,6)
İnsülin direnci	3 (8,3)
Glukoz tolerans bozukluğu	4 (11,1)
Değerler ortalama ± standart sapma/ortanca değer olarak verilmiştir. PD: Periton diyalizi, HD: Hemodiyaliz, DM: Diabetes mellitus, NSDM: Nakil sonrası gelişen diabetes mellitus	



Şekil 1. Böbrek nakli sonrası ortaya çıkan insülin-glukoz metabolizması bozuklukları görülme sıklığı

NSDM: Nakil sonrası gelişen diabetes mellitus

Tablo 2. Hastaların böbrek nakli sonrası izleminde yıllık kan şekeri, insülin dengesi değişim parametreleri ve ilişkili faktörler							
	0. Yıl	**1. Yıl**	**2. Yıl**	**3. Yıl**	**4. Yıl**	**5. Yıl**	**p**
Hasta sayısı (n)	36	31	26	19	12	10	
VKİ (kg/m^2)[a,b]	16,86±1,59	19,49±3,53	19,02±2,86	18,97±2,54	18,93±1,84	19,63±2,31	0,30[a]/**0,02**[b]
Açlık kan şekeri (mg/dL)[a,c]	88,88±10,62	87,30±16,55	84,69±8,30	87,50±12,85	88,25±13,59	83,70±10,32	0,35
İnsülin (IU/mL)[a]	12,43±15,13	6,08±4,92	7,66±5,43	7,01±4,39	5,76±3,16	8,09±3,81	0,06[a]/**0,01**[c]
HOMA-IR (mg/dL)[a]	2,60±2,81	1,28±1,05	1,56±1,14	1,47±1,01	1,52±1,11	1,72±0,77	0,06[a]
OGTT (mg/dL)[a]							
0. Dakika	94±21,21	85,44±8,53	83,37±7,74	81,38±10,60	81,30±13,27	79,43±7,48	0,2[a]
120. Dakika	95±4,24	123,59±74,17	99,89±17,58	117,81±35,07	111,60±35,38	94,14±19,45	0,07[a]
İnsülin-glukoz metabolizma bozukluğu saptanan hasta sayısı (n), (%)[e]	2 (5,4)	6 (19,2)	9 (34,6)	9 (47,3)	0 (0)	0 (0)	**0,03**[e]
NSDM'li hasta sayısı (n), (%)[e]	0 (0)	2 (6,4)	2 (7,7)	2 (10,4)	0 (0)	0 (0)	**0,04**[e]
İnsülin direnci*[,e]	1 (2,7)	2 (6,4)	3 (11,5)	3 (15,8)	0 (0)	0 (0)	**0,04**[e]
Glukoz tolerans bozukluğu[e]	1 (2,7)	2 (6,4)	4 (15,4)	4 (21,1)	0 (0)	0 (0)	**0,02**[e]

*Hastanın puberte durumuna göre değerlendirildi.

[a]Freidman's two-way analizi.

[b]Student's t testi; VKİ 0. yıl değerleri post transplant diğer yıllara göre anlamlı düşük (p=0,02).

[c]Student's t testi; VKİ ve insülin düzeyleri 0. yıl değerleri post transplant diğer yıllara göre anlamlı düşük (sırasıyla p=0,02, p=0,01).

[d]Kruskal-Wallis testi: Nakil öncesi döneme göre nakil sonrası dönemde puberteye giren hasta oranı istatistiksel olarak anlamlı bir biçimde süreklilik göstererek artmış (p=0,03).

[e]Kruskal-Wallis testi: nakil öncesi ve nakil sonrası ilk 3 içinde insülin-glukoz metabolizma bozukluğu saptanan hasta oranı istatistiksel olarak anlamlı bir biçimde süreklilik göstererek artmış (p=0,03).

VKİ: Vücut kitle indeksi, IU: İnternasyonel ünite, HOMA-IR: Homeostatic model assesment insülin rezistansı, OGTT: Oral glukoz tolerans testi, NSDM: Nakil sonrası gelişen diabetes mellitus

Benzer şekilde ortalama insülin değerlerinin tüm izlem sürecinde istatistiksel olarak anlamlı bir değişim göstermediği saptandı. Ancak nakil sonrası dönem içerisinde yıllık izlemler ayrı ayrı olarak nakil öncesi dönem ile karşılaştırıldığında tüm periyotlarda ortalama insülin değerleri nakil öncesi döneme göre istatistiksel olarak anlamlı bir biçimde düşük bulundu (p=0,01). Nakil sonrası dönemde istatistiksel olarak anlamlı bir biçimde puberteye girme oranlarının yükselerek gittiği görüldü (p=0,04), Yıllar içerisinde görülen değişim içerisinde sadece insülin-glukoz metabolizma sorunu saptanan toplam hasta oranının (p=0,03), NSDM saptanan hasta oranının (p=0,04), insülin direnci saptanan hasta oranın (p=0,04) ve glukoz tolerans bozukluğu saptanan hasta oranının (p=0,02) nakil öncesi dönemden nakil sonrası 3 yıl boyunca istatistiksel olarak anlamlı biçimde artış gösterdiği saptandı. İnsülin-glukoz metabolizma bozukluğu saptanan ve saptanmayan hastaların klinik, laboratuvar özellikleri Tablo 3'te karşılaştırmalı olarak verildi. Glukoz metabolizma bozukluğu saptanan grupta sayıca obez hasta ve nakil sonrası eve çıkana kadar geçici insülin kullanan hasta sayısı daha fazla saptandı. Yine bu grupta kreatin klirensi istatistiksel olarak daha düşük bulundu (p=0,03).

İnsülin-glukoz metabolizma bozukluğu (insülin direnci, bozulmuş glukoz toleransı ve NSDM) saptanan hastaların tamamının korelasyon sonuçları Tablo 4'te verildi. Sadece anlamlı bulunan sonuçlar tabloya alındı. Nakil yaşları 10 yaşından büyük olan hastalarda NSDM, insülin direnci ve bozulmuş glukoz toleransı gelişimi arasında pozitif korelasyon saptandı (p=0,01, r=0,45). Nakil öncesi diyaliz yöntemi olarak kronik PD tedavisi altında izlenmiş olmak ile insülin-glukoz metabolizma bozulukları arasında pozitif korelasyon saptandı (p=0,005, r=0,45). Nakil öncesi kreatinin sonuçları ile insülin-glukoz metabolizması bozuklukları arasında pozitif korelasyon bulundu (p=0,02, r=0,52). Hem 1. yılda (p=0,009, r=0,66) hem de 2. yılda (p=0,03, r=0,42) fazla tartılı ve/veya obez olmak ile insülin-glukoz metabolizması bozuklukları arasında pozitif korelasyon saptandı. Son olarak insülin-glukoz metabolizması bozuklukları saptandığı sırada puberteye girmiş olmak ile glukoz metabolizma bozuklukları arasında da pozitif korelasyon saptandı (p=0,01, r=0,40).

Tablo 3. İnsülin-glukoz metabolizma bozukluğu saptanan ve saptanmayan hastaların klinik, laboratuvar özellikleri			
	İnsülin-glukoz metabolizma bozukluğu saptanmayan hasta	**İnsülin-glukoz metabolizma bozukluğu saptanan hasta**	**p**
Hasta sayısı, (n)/(%)	27 (75)	9 (25)	
VKİ (kg/m^2)			
Pretransplant	16,6±1,45	17,76±1,89	0,59
1. Yıl	18,1±2,58	21,73±4,55	0,68
2. Yıl	18,27±2,34	21,27±3,31	0,57
3. Yıl	18,95±2,79	19,05±1,67	0,39
4. Yıl	19,11±1,77	18,05±2,72	0,44
5. Yıl	19,41±2,4	20,51±2,42	0,96
Fazla tartılı ve/veya obez hasta sayısı n, (%)	1 (3,7)	6 (96,3)	**0,01**
Pretransplant CrCl (mL/dak/1,73 m^2)	19,05±5,65	10,4±2,36	**0,03**
Nakil sonrası eve çıkana kadar geçici insülin kullanımı n, (%)	4 (14,8)	7 (77,7)	**0,002**
İmmünosüpresif tedavi n, (%)			
Siklosporin tedavi rejimi	3 (11)	0 (0)	0,29
Takrolimus tedavi rejimi	24 (89)	9 (100)	
Nakil tipi			
Canlı	17	4	0,32
Kadavra	10	5	
Pretransplant RRT tipi (n), (%)			
PD	19 (70,3)	4 (44,4)	0,15
HD	3 (11,1)	5 (55,6)	
Preemptif	5 (18,6)	0 (0)	
VKİ: Vücut kitle indeksi, CrCl: Kreatinin klirensi, RRT: Renal replasman tedavisi, PD: Periton diyalizi, HD: Hemodiyaliz			

TARTIŞMA

Çeşitli çalışmalarda NSDM gelişim oranları konusunda çok farklı sonuçlara rastlanmaktadır. Bu verilerin büyük çoğunluğu yetişkin hastalara ait olup, bu oranlar organ nakli olmayan hasta grubu ile karşılaştırıldığında organ nakli olmuş hastalarda anlamlı artışlar gözlenmektedir.[7-12] Ancak çocukluk çağında NSDM, bozulmuş glukoz toleransı ve insülin direnci konusunda bilgi ve deneyim son derece kısıtlıdır. Bu çalışma 5 yıllık izlemi içeren ve böbrek nakli yapılmış 36 hastayı kapsayan bir çalışma olup, konuyla ilgili veriler içermesi açısından oldukça önemlidir.

Son zamanlarda yapılan randomize kontrollü çalışmalarda nakil sonrası ilk 1 yılda NSDM gelişme oranı %2-50 gibi çok geniş bir aralıkta bildirilmektedir.[7-13] Ülkemiz verileri incelendiğinde, nakil sonrası hastalarda DM gelişimi ve insülin direnci gelişimi ile ilgili yapılan araştırmaların kısıtlılığı dikkati çekmektedir. Karakan ve ark.[14] erişkin yaş grubunda %19,4 oranında NSDM saptadıklarını bildirmişlerdir. Ülkemizde yapılan çalışmaların kısıtlı olmasına rağmen organ nakli olmayan hasta grubu ile karşılaştırıldığında NSDM gelişiminin nakil sonrası hastalarda daha yüksek oranda saptandığı rahatlıkla söylenebilir. Çalışmamızda, NSDM oranı %6,5, insülin direnci tanısı oranı %8,3 ve glukoz tolerans bozukluğu oranı %11,1 olarak saptandı (Tablo 2). Yurt içi ve yurt dışında yapılan çalışmalarda bildirilen oranlar bizim çalışmamızda saptanan oranlara benzer bulundu.

Bozulmuş glukoz toleransı, NSDM gelişimi için önemli bir risk faktörüdür. Prediyabeti olan hastalar, NSDM'ye ilerleme riski yüksek olduğundan, nakil sonrası dönemde yakından izlenmelidir. Bu hastalar, düzenli kan glukozu değerlendirmelerine ek olarak, kilo yönetimi, sağlıklı beslenme ve fiziksel aktivite gibi yaşam tarzı değişiklikleri konusunda danışmanlık almalıdır. Bu önlemler, özellikle nakil sonrası immünosüpresif tedavi bağlamında, diyabet gelişimi için ek bir risk faktörü olan aşırı kilo alımını önlemeye yardımcı olabilir.[13] Bu çalışma, az sayıdaki pediatrik çalışmalardan farklı olarak hastaların nakil öncesi insülin glukoz metabolizması ile ilgili bulguları da içermektedir. Bu bulgulara göre insülin düzeyleri nakil öncesi dönemde diğer post transplant dönemlere göre istatistiksel olarak anlamlı bir biçimde yüksek bulundu

Tablo 4. NSDM, insülin direnci ve bozulmuş glukoz toleransı gelişen hastalarda korelasyon analizi sonuçları (sadece anlamlı bulunan sonuçlar verildi)

	p	r
Yaş (≥10 yaş)	0,02	0,52
PD yapmış olmak	0,005	0,45
Nakil sonrası eve çıkana kadar geçici insülin kullanımı	0,004	0,57
Pretransplant diyaliz yetersizliği (CrCl ≤10 mL/dak/1,73 m^2)	0,01	0,51
1. Yılda fazla tartılı ve/veya obez olmak (VKİ ≥85 p)	0,009	0,66
2. Yılda fazla tartılı ve/veya obez olmak (VKİ ≥85 p)	0,03	0,42
Puberteya girmiş olmak	0,01	0,40

NSDM: Nakil sonrası gelişen diabetes mellitus, PD: Periton diyalizi, CrCl: Kreatinin klirensi, VKİ: Vücut kitle indeksi, HOMA-IR: Homeostatic model assesment insülin rezistansı

(p=0,01). HOMA-IR ortalama düzeyleri pretransplant dönemde yüksek bulunsa da posttransplant diğer dönemler ile karşılaştırıldığında istatistiksel olarak anlamlı bir sonuç bulunmadı. Bizim olgularımızda pretransplant dönemde DM hastası saptanmadı. Ancak birbirinden farklı olan 1 (%2,7) hastada bozulmuş glukoz toleransı, 1 (%2,7) hastada ise insülin direnci saptandı.

Post-transplant dönemde postoperatif ilk günlerde cerrahi stres, yüksek doz steroid kullanımı ile ilişkili olarak hiperglisemi saptanma olasılığı yüksektir. Ancak sonraki birkaç hafta da dahil olmak üzere bu stres hiperglisemisi döneminde insüline ihtiyaç duymuş olmak, posttransplant uzun dönemde NSDM açısından artmış risk faktörü olarak bildirilmektedir.[15] Bu çalışmada posttransplant hiperglisemiye bağlı olarak 11 hastada (%30,5) geçici olarak insülin kullanılmıştır. NSDM, insülin direnci ve bozulmuş glukoz toleransı saptanan hastalar, saptanmayan hastalar ile karşılaştırıldığında hasta grubunda bu tedaviyi alma oranı istatistiksel olarak anlamlı bulundu (p=0,002), (Tablo 3). Ayrıca literatür ile uyumlu olarak NSDM, insülin direnci ve bozulmuş glukoz toleransı bir havuzda değerlendirildiğinde bu havuza giren hastalar ile posttransplant erken dönem insülin kullanma ihtiyacı arasından pozitif kolerasyon saptandı (p=0,004, r=0,57) (Tablo 4).

Nakil sonrası ilk 1 yılda NSDM gelişme riskinin en yüksek oranda bulunduğu ifade edilmektedir.[16] İlk altı ayda NSDM insidansı %20,5'tir. Bu sürenin ardından insidans düşer ve yıllık yaklaşık %6 seviyesinde sabitlenir; bu oran, transplantasyon bekleme listelerindeki hastalardaki diyabet oranıyla benzerdir.[17] Eğer nakil sonrası dönemin ilerleyen zamanlarında diyabet teşhis edilirse, bu durum NSDM yerine tip 2 DM olarak da değerlendirilebilir.[18] Bu çalışmada posttransplant 5 yıllık izlem sürecinde anormal glukoz metabolizmasına sahip 9 hasta saptandı. Literatür ile uyumlu olarak bu hastaların 6'sına (% 66,6) ilk 1 yılda tanı konulmuş olup tamamı ilk 2 yıl içinde tanı aldı. Bu çalışmadan çıkarılacak önemli sonuçlardan birisi de nakil olan hastaların nakil sonrası özellikle ilk 1 yılda glukoz metabolizma bozukluğu açısından yakından izlenmesinin önemidir.

Nakil sonrası DM gelişimi, bozulmuş glukoz tolerans ve insülin direncine neden olabilecek değiştirilebilir risk faktörleri içerisinde ise özellikle aşırı kilo, obezite, sedanter yaşam, sigara içme gibi kişinin yaşam alışkanlıklarına bağlı müdahale edilebilir nedenler bulunmaktadır.[19,20] Nakil olmamış hastalar gibi nakil olan hastalarda da obezite varlığı, özellikle VKİ >29,9 kg/m^2 saptanan hastalar, DM gelişimi açısından kuvvetli risk taşımaktadırlar. Çalışmamızda nakil sonrası hem 1. yılda (p=0,009, r=0,66) hem de 2. yılda (p=0,03, r=0,42) fazla tartılı ve/veya obez olmak (VKİ ≥85 p) ile NSDM, glukoz tolerans bozukluğu ve insülin direnci arasında pozitif korelasyon saptandı (Tablo 4).

Kuzey Amerika Pediatrik Nefroloji Çalışma Grubu verilerine göre 1987-2002 yılları arasında obezite prevelansı nakil olan çocuk böbrek hastalarında da %9,7'den %12,4'e çıktığı tespit edilmiştir. Obezite oranlarındaki bu artışın nakil sonrası gelişen bozulmuş glukoz toleransı, insülin direnci ve NSDM insidansının artmasında en önemli katkı veren problem olduğu düşünülmektedir.[21] Bu çalışmada nakil sonrası süratle artan obezite göstergesi olarak hem 1. yıl VKİ (p=0,008, r=0,42), hem de 2. yıl VKİ (p=0,02, r=0,42) sonuçları posttransplant insülin-glukoz metabolizma bozuklukları arasında istatistiksel olarak anlamlı bir biçimde pozitif korelasyon saptandı. Üremi, diyaliz tedavisi, renal tübüler disfonksiyon, asidoz, anemi gibi bir çok faktörün neden olduğu iştahsızlığın nakil uygulanması ile birlikte daha ilk haftalardan itibaren ortadan kalkması ve bu dönemlerde uygulanan daha yüksek doz steroid tedavisinin de katkısıyla iştahın daha da artması, birçok diyet yasağının sonlanmasının, hızlı kilo alımının en önemli nedenleri olduğunu düşünmekteyiz. Bunu destekler nitelikte pretransplant VKİ değerleri posttransplant süreçte yıllar içerisinde istatistiksel olarak anlamlı bir

biçimde artmamış olsa da (Tablo 2, Friedman's two-way analizi, p=0,30) pretransplant VKİ değerleri sonraki yılların her biri ile ayrı ayrı değerlendirildiğinde, istatistiksel olarak anlamlı bir biçimde düşük bulundu (Tablo 2, Student's t testi, p=0,02).

Takrolimus ve siklosporin gibi kalsinörin inhibitörleri, insülin salgısını azaltarak, insülin direncini artırarak ve pankreatik beta hücreleri üzerinde toksik etki yaparak NSDM gelişimine katkıda bulunur.[22-24] Diabetojenik etkileri, yüksek doz kortikosteroidlerle birlikte kullanıldığında daha da artar.[25] Özellikle takrolimus, akut greft reddini önlemede ve greft sağkalımını iyileştirmede siklosporinden 100 kat daha etkilidir. Ancak, takrolimus, NSDM gelişimi ile siklosporine kıyasla daha güçlü bir şekilde ilişkilidir.[26] Siklosporin ile tedavi edilen gruplarda NSDM saptanma oranı %23,7 olarak bulunmuştur.[27] Takrolimus ile tedavi edilen bir başka nakil grubunda ise 6 ay sonraki NSDM oranı %57,1 gibi daha yüksek bir oranda saptanmıştır.[28] Dört bin dört yüz elli iki kalp nakli hastası üzerinde yapılan büyük bir çalışmada, takrolimusun NSDM gelişimi ile en yüksek ilişkiye sahip olduğu (HR=1,459), ardından rifampisin ve siklosporinin geldiği bulunmuştur.[29] Son yıllarda yapılan başka bir çalışmada ise takrolimus ve siklosporin içeren rejimler NSDM gelişimi açısından karşılaştırılmış fakat istatistiksel olarak herhangi bir fark saptanmamıştır.[30] Takrolimus genellikle immünosüpresif protokollerde birinci basamak tedavi olarak kullanılır ve mTOR inhibitörleri, mikofenolat mofetil veya azatioprin gibi diğer ilaçlarla kombin edilerek dozlar azaltılır ve toksisite en aza indirilmeye çalışılır.[31-33] Çalışmamızda ise insülin-glikoz metabolizması bozukluğu olan ve olmayan hastalar almakta oldukları takrolimus ve siklosporin içeren rejimler açısından karşılaştırıldığında istatistiksel olarak anlamlı farklılık saptanmadı (p=0,29) (Tablo 3).

Hepatit C enfeksiyonu ve CMV enfeksiyonu geçirilmesi de bazı çalışmalarda NSDM gelişimi için risk faktörleri olarak tanımlanmıştır.[10] Otozomal dominant polikistik böbrek hastalığı varlığı, PD tedavisi almış olmak, metabolik sendromun diğer komponentlerinin varlığının da NSDM gelişimi ile ilişkili olduğu bildirilmektedir.[7] Obezite dışında önlebilir risk faktörleri içerisinde sadece PD tedavisini uygulamış olmanın hasta grubumuzda nakil sonrası gelişen insülin-glukoz metabolizma bozuklukları açısından istatistiksel olarak anlamlı bir şekilde pozitif korelasyona sahip olduğu görüldü (p=0,005, r=0,45), (Tablo 3).

Bu çalışma önemli veriler içerse de bazı handikaplara sahiptir. Geriye dönük olarak planlanması en önemli handikapı olarak düşünülebilir. Veriler 5 yılı ilgilendirse de çalışmaya alınan 36 hastadan henüz 5 yılını dolduran hasta sayısı sadece 10'dur. Daha objektif sonuçlara ulaşmak için çok merkezli, daha çok hasta sayısına ulaşmış ileriye dönük çalışmalara ihtiyaç olduğunu düşünmekteyiz.

SONUÇ

Sonuç olarak çalışmamızda böbrek nakli sonrası 2 hasta NSDM, 3 hasta insülin direnci, 4 hasta glukoz tolerans bozukluğu olmak üzere toplamda 9 (%25) hasta İnsülin-glukoz metabolizma bozukluğu tanısı aldı. İnsülin direnci ve glukoz tolerans bozukluğu saptanan hastaların tamamına ilk 2 yılda, NSDM hastalarının tamamına ilk bir yılda DM tanısı konuldu. Böbrek nakli olan hastalarda normal popülasyona ve SDBY hastalarına göre DM riskinin daha yüksek olduğu desteklenmiş oldu. Bu nedenle nakil sonrası hastalarda DM gelişimini önlemek amacıyla olgular nakil öncesi ve sonrası klinik ve laboratuvar bulgularıyla yakından izlenmelidir. NSDM gelişimi, insülin direnci ve glukoz tolerans bozukluğu saptandığı takdirde gerekli tedavi girişimleri uygulanmalıdır.

Etik

Etik Kurul Onayı: Çalışma için Sağlık Bilimleri Üniversitesi, İzmir Tepecik Eğitim ve Araştırma Hastanesi Etik Kurulu'ndan onay alınmıştır (karar numarası: 8, tarih: 29.12.2014).

Hasta Onayı: Retrospektif çalışma.

Dipnotlar

Yazarlık Katkıları

Cerrahi ve Medikal Uygulama: C.A., B.K.D., Ö.Y., Konsept: H.M., C.A., Ö.Y., Dizayn: H.M., Ö.Y., Veri Toplama veya İşleme: H.M., C.A., Analiz veya Yorumlama: H.M., Ö.Y., Literatür Arama: H.M., C.A., Ö.Y., Yazan: H.M., Ö.Y.

Çıkar Çatışması: Yazarlar bu makale ile ilgili olarak herhangi bir çıkar çatışması bildirmemiştir.

Finansal Destek: Çalışmamız için hiçbir kurum ya da kişiden finansal destek alınmamıştır.

KAYNAKLAR

1. Dong M, Parsaik AK, Eberhardt NL, Basu A, Casio FG, Kudva YC. Cellular and physiological mechanisms of new onset diabetes mellitus after solid organ transplantation. Diabet Med. 2012;29:1-12.
2. Franz Schaefer. Endocrine and growth disorders in chronic kidney disease. In Ellis D. Avner, William E. Harmon, Ptrick Niaudet, Norishige Yoshikawa, eds. Pediatric Nephrology 6th Edition. 2009; p.1713-54.
3. Solhjoo, M, Kumar SC. New Onset Diabetes After Transplant. 2023. In StatPearls; StatPearls Publishing: Treasure Island, FL, USA, 2024.
4. Nagaraja P, Ravindran V, Stiff GM, Baboolal K. Role of insulin resistance indices in predicting new-onset diabetes after kidney transplantation. Tranplant Int. 2012;26:273-80.
5. Venna IC. Nelson textbook of pediatrics. Indian J Pediatr. 2003;70,892.

6. UpToDate. Son erişim tarihi: 17.07.2025. Erişim linki: www.uptodate.com/New-onset-diabetes-after-transplant-(NODAT)-in-renal-transplant-recipients
7. Juan Khong M, Ping Chong Ch. Prevention and management of new-onset diabetes mellitus in kidney transplantation NJM. 2014;72:127-34.
8. Prakash J, Rathore SS, Brojen Singh T, et al. New onset diabetes after transplantation (NODAT): analysis of pre-transplant risk factors in renal allograft recipients. Indian J Transplant. 2012;6:77-82.
9. Pham PT, Pham PM, Pham SV, Pham PA, Pham PC. New onset diabetes after transplantation (NODAT): an overview. Diabetes Metab Syndr Obes. 2011;4:175-86.
10. Sarno G, Muscogiuri G, De Rosa P. New-onset diabetes after kidney transplantation: prevalence, risk factors, and management. Transplantation. 2012;93:1189-95.
11. Koshy SM, Gutmann A, Hebert D, Parkes RK, Logan AG. Incidence and risk factors for cardiovascular events and death in pediatric renal transplant patients: a single center longterm outcome study. Pediatr Transplant. 2009;13:1027-33.
12. Gomes V, Ferreira F, Guerra J, Bugalho MJ. New-onset diabetes after kidney transplantation: incidence and associated factors. World J Diabetes. 2018;9:132-7.
13. Jenssen T, Hartmann A. Post-transplant diabetes mellitus in patients with solid organ transplants. Nat Rev Endocrinol. 2019;15:172-88.
14. Karakan Ş, Sezer S, Acar Ö, Nurhan F. Böbrek nakli sonrası yeni diyabetes mellitus gelişmesine etki eden faktörler; obezite ve inflamasyonun rolü. Fırat Tıp Derg. 2013:18;122-5.
15. Chakkera HA, Knavoler WC, Devarapilli Y, et al. Relationship between inpatient hyperglicemia and insulin treatment after kidney transplantation and future new onset diabetes mellitus. Clind Am Soc Nephrol. 2010;5:1669-75.
16. Kidney Disease: Improving Global Outcomes (KDIGO) clinical practice guideline for the care of kidney transplant recipients. Am J Transplant. 2009;9:1-155.
17. Ghisdal, L.; Van Laecke, S.; Abramowicz, M.J.; Vanholder, R.; Abramowicz, D. New-onset diabetes after renal transplantation: risk assessment and management. Diabetes Care. 2012;35:181-8.
18. Porrini EL, Díaz JM, Moreso F, et al. Clinical evolution of post-transplant diabetes mellitus. Nephrol Dial. Transplant. 2016;31:495-505.
19. Ouni A, Sahtout W, Hadj BM, et al. New-onset diabetes as a complication after kidney transplant: incidence and outcomes. Exp Clin Transplant. 2022;20:129-31.
20. Chang S, Jiang J. Association of body mass index and the risk of new-onset diabetes after kidney transplantation: a meta-analysis. Transpl P. 2018;50:1316-25.
21. Hanevold CD, Ho PL, Talley L, Mitsnefes MM. Obesity and renal transplant outcome: a report of the North American Pediatric Renal Transplant Cooperative Study. Pediatrics 2005;115:352-6.
22. Katwal PC, Jirjees S, Htun ZM, Aldawudi I, Khan S. The Effect of anemia and the goal of optimal HbA1c control in diabetes and non-diabetes. Cureus. 2020;12:8431.
23. Hjelmesaeth J, Sagedal S, Hartmann A, et al. Asymptomatic cytomegalovirus infection is associated with increased risk of new-onset diabetes mellitus and impaired insulin release after renaltransplantation. Diabetologia. 2004;47:1550-6.
24. Rodríguez-Rodríguez AE, Porrini E, Hornum M, et al. Post-transplant diabetes mellitus and prediabetes in renal transplant recipients: an update. Nephron. 2021;1405:317-29.
25. Ponticelli C, Favi E, Ferraresso M. New-onset diabetes after kidney transplantation. Medicina. 2021;57:250.
26. Di Maira T, Little EC, Berenguer M. Immunosuppression in liver transplant. Best Pract Res Clin Gastroenterol. 2020;46-7.
27. Nam JH, Mun JI, Kim SI, et al. Beta cell disfunction rather than insulin resistance is the main contributing factor for the devolopment of postrenal transplantation diabetes mellitus. Transplantation. 2001;71:1417-23.
28. Cho YM, Park KS, Jung HS, et al. High incidance of tacrolimus associated posttransplantation diabetes in the Korean renal allograft recipients according to American Diabetes Association criteria. Diabetes Care. 2003; 26: 1123-8.
29. Wang R, Zhang Y, Fan J, Wang Z, Liu Y. Risk factors for new-onset diabetes mellitus after heart transplantation: a nomogram approach. Transplant Proc. 2022;54:762-8.
30. Yousif E, Abdelwahab A. Post-transplant diabetes mellitus in kidney transplant recipients in Sudan: a comparison between tacrolimus and cyclosporine-based immunosuppression. Cureus. 2022;14:22285.
31. Noble J, Terrec F, Malvezzi P, Rostaing L. Adverse effects of immunosuppression after liver transplantation. Best Pract Res. Clin Gastroenterol. 2021;54-5.
32. Nankivell BJ, P'Ng CH, O'Connell PJ, Chapman JR. Calcineurin inhibitor nephrotoxicity through the lens of longitudinal histology: comparison of cyclosporine and tacrolimus eras. Transplantation. 2016;100:1723-31.
33. Braithwaite HE, Darley DR, Brett J, Day RO, Carland JE. Identifying the association between tacrolimus exposure and toxicity in heart and lung transplant recipients: a systematic review. Transplant Rev. 2021;35:100610.

DOI: 10.4274/forbes.galenos.2025.27880
Forbes J Med 2025;6(2):165-71

Original Article / Orijinal Araştırma

Assessing the Association Between Troponin I Changes Below the Diagnostic Cut-off with Mortality and Major Cardiac Events

Tanı Eşiğinin Altındaki Troponin I Değişimlerinin Mortalite ve Majör Kardiyak Olaylarla İlişkisinin Değerlendirilmesi

Alper TORUN[1], Hayriye GÖNÜLLÜ[2], Caner SAĞLAM[3], İsmet PARLAK[4†]

[1]Dokuz Eylül University Faculty of Medicine, Department of Emergency Medicine, İzmir, Türkiye
[2]Bakırçay University Faculty of Medicine, Department of Emergency Medicine, İzmir, Türkiye
[3]University of Health Sciences Türkiye, İzmir City Hospital, Clinic of Emergency Medicine, İzmir, Türkiye
[4]Aksaray University Faculty of Medicine, Department of Emergency Medicine, Aksaray, Türkiye

[†]The author is deceased.

Cite as: Torun A, Gönüllü H, Sağlam C, Parlak İ. Assessing the association between troponin I changes below the diagnostic cut-off with mortality and major cardiac events. Forbes Med J. 2025;6(2):165-71

ABSTRACT

Objective: In this study, we aimed to investigate the effect of variations in cardiac troponin I (cTnI) (delta troponin) levels-within the normal reference range-on 30-day mortality and non-fatal major adverse cardiac events (MACE) in patients who were monitored in the emergency department with a preliminary diagnosis of acute coronary syndrome.

Methods: After obtaining ethics committee approval, this retrospective study included patients who presented to the emergency department with chest pain and/or related symptoms, underwent serial cTnI measurements, and had measured cTnI levels ≤0.05 ng/mL. Patients were divided into two groups based on the presence or absence of delta troponin. The association between delta troponin and 30-day mortality and non-fatal MACE was examined.

Results: The group with positive delta cTnI had a significantly higher risk of non-fatal MACE compared to the group without delta cTnI (p<0.001). The presence of cardiac-type chest pain, ischemic ST-segment changes on electrocardiographic, and a history of hypertension, diabetes, hyperlipidemia, and coronary artery disease were all significantly associated with an increased risk of 30-day mortality and non-fatal MACE. Similarly, advancing patient age was significantly associated with these outcomes.

Conclusion: Our findings demonstrate that changes in delta troponin levels below the 99th percentile have a significant impact on the risk of 30-day mortality and non-fatal MACE.

Keywords: 99th percentile, delta troponin, chest pain, mortality, major adverse cardiac event

ÖZ

Amaç: Bu çalışmada, acil serviste akut koroner sendrom ön tanısıyla izlenen hastalarda, normal referans aralığı içindeki kardiyak troponin I (cTnI) düzeylerindeki değişimlerin (delta troponin) 30 günlük mortalite ve ölümcül olmayan majör kardiyak olaylar (MACE) üzerindeki etkisini araştırmayı amaçladık.

Yöntem: Bu çalışmaya, etik kurul onayı alındıktan sonra göğüs ağrısı ve/veya ilişkili semptomlar ile acil servise başvuran, seri cTnI takibi yapılan ve ölçülen cTnI ≤0,05 ng/mL olan hasta grubu geriye dönük olarak alındı. Hastalar delta troponin olan ve olmayan şeklinde iki gruba ayrıldı ve 30 günlük mortalite ve mortalite dışı MACE ile ilişkili durumları incelendi.

Bulgular: Delta cTnI pozitif olan grupta olmayan gruba göre mortalite dışı MACE riski daha yüksek bulundu (p<0,001). Kardiyak tipte göğüs ağrısı ve elektrokardiyogramda iskemik ST segment değişikliğinin olması, hasta özgeçmişinde hipertansiyon, diabet, hiperlipidemi ve koroner arter hastalığı öyküsünün olması, 30 günlük mortalite ve mortalite dışı MACE riskini anlamlı olarak artırdı. Benzer anlamlılık hasta yaşı arttıkça da görüldü.

Received/Geliş: 15.05.2025
Accepted/Kabul: 02.07.2025
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Corresponding Author/ Sorumlu Yazar:
Alper TORUN, MD,
Dokuz Eylül University Faculty of Medicine, Department of Emergency Medicine, İzmir, Türkiye
✉ alpertorun87@gmail.com
ORCID: 0000-0002-9045-7922

Sonuç: Çalışmamız bulguları, 99 persantil altındaki delta troponin değişiminin 30 günlük mortalite ve mortalite dışı MACE riski üzerinde etkisinin anlamlı olduğunu gösterdi.

Anahtar Kelimeler: 99 persantil, delta troponin, göğüs ağrısı, mortalite, majör kardiyak olay

INTRODUCTION

Chest pain is one of the most common complaints presented to the emergency department. In the majority of cases, the underlying cause is benign, non-cardiac in nature, and does not require hospitalization. However, a smaller yet potentially life-threatening subset of these patients is diagnosed with acute coronary syndromes (ACS).[1,2] According to the literature, approximately 2% of ACS patients are mistakenly discharged from the emergency department, and this group demonstrates increased 30-day morbidity and mortality rates.[3] This poses a significant concern for both emergency physicians and cardiologists, complicating the overall management of patients presenting with chest pain.[1]

One of the key challenges lies in the subjective nature of chest discomfort, which may not be clearly articulated by certain patient populations, including women, the elderly, individuals with diabetes, and those with mental health conditions. In such groups, ACS may present with atypical or equivalent symptoms such as dyspnea, fatigue, epigastric pain, dizziness or nausea and vomiting, rather than classic chest pain.[1,2]

In ST-elevation myocardial infarction (STEMI), which accounts for a small proportion of ACS, the diagnosis can be made rapidly through electrocardiographic (ECG) evaluation. The identification of non-STEMI (non-STEMI) cases-characterized by ischemic ECG changes and elevated cardiac enzymes-is relatively straightforward. However, the primary challenge lies in detecting and managing low-risk ACS cases in which cardiac enzyme levels are either within normal limits or show only minimal elevation.[2,4]

The diagnostic approach to ACS relies on a combination of clinical history, ECG, and cardiac biomarkers. While the interpretation of history and ECG findings has remained largely unchanged over the years, cardiac biomarkers have undergone frequent updates in recent years.[2,4] Among cardiac troponin T (cTnT) and troponin I (cTnI) are considered reliable markers for both the diagnosis and exclusion of ACS.[2,4-6] cTnI, a myocardial-specific marker of infarction, begins to rise within 2 to 4 hours following acute myocardial injury and typically peaks at around 24 hours.[7]

In current ACS guidelines when neither STEMI nor non-STEMI is suspected, the evaluation of chest pain is recommended to incorporate validated risk-stratification tools.[8,9] This approach is particularly important in emergency departments that rely on conventional cTn assays rather than high-sensitivity cardiac troponin (hs-cTn), as risk scores improve the identification of low-risk chest pain patients in these settings.[9] Notably, the History, ECG, Age, Risk Factors, and Troponin (HEART) score-encompassing HEART, its successor, the HEARTS[3] score, facilitates the early discharge of chest pain patients without necessitating exercise testing, advanced cardiac investigations, or imaging.[5] Moreover, these tools permit reliable estimation of 30-day major adverse cardiac event (MACE) risk-including non-fatal myocardial infarction, revascularization, and out-of-hospital cardiac arrest-thereby optimizing patient management in the emergency department.[4,9]

Delta troponin refers to the difference between two troponin values evaluated at different times during a patient's follow-up. In our study, we aimed to evaluate whether changes in troponin values, even when below the normal limits, could be clinically predictive. In this study, we aimed to evaluate the impact of an increase in delta troponin levels on 30-day mortality and non-fatal MACE among patients monitored in the emergency department with a provisional diagnosis of ACS and normal-range cTnI values.

METHODS

This retrospective cross-sectional study was approved by the Clinical Research Ethics Committee of University of Health Sciences Türkiye, İzmir Bozyaka Training and Research Hospital (decision number: 7, date: 04.07.2018). We included patients aged ≥18 years who presented to the emergency department of this tertiary center between January 1, 2017, and December 31, 2017, with chest pain or chest pain-related symptoms (dyspnea, fatigue, epigastric pain, dizziness, nausea/vomiting), and whose serial cTnI measurements remained within the normal range (≤0.05 ng/mL). Exclusion criteria comprised pregnancy; diagnoses of STEMI, non-STEMI, sepsis, acute or chronic renal failure, pulmonary thromboembolism, or severe anemia; and presentation due to trauma.

Demographic characteristics, presenting complaints, duration and character of chest pain, and ECG findings were extracted retrospectively from manually maintained patient files and the hospital's electronic health record system. Chest pain type and ECG abnormalities were classified according to the HEART risk score.

The first cTnI level measured upon arrival in the emergency department was defined as the baseline troponin.

A second blood sample for cTnI was drawn at the fourth hour. Patients whose baseline and fourth-hour cTnI values both remained below the assay cut-off yet differed from each other were designated the "delta+positive" group. Those whose cTnI levels remained unchanged and below the assay cut-off at four hours were classified as the "delta-negative" group.

All cTnI assays were performed on the Beckman Coulter Access 2 analyzer (Beckman Coulter, Canada, USA) using the manufacturer's original Access AccuTnI kit. To assess the association between changes in troponin values below the assay cut-off and 30-day outcomes-mortality and non-fatal MACE-patients were contacted by telephone and their electronic medical records were reviewed. Collected data were recorded on a standardized research form, and 30-day mortality and MACE information were obtained via telephone follow-up or hospital database review.

There is no previously established delta threshold reported in the literature for this context. Therefore, the 0.005 ng/mL cutoff for delta cTnI was determined empirically through receiver operating characteristic (ROC) curve analysis. Using Youden's index, we identified the optimal threshold that maximizes the sum of sensitivity and specificity for predicting both 30-day mortality and MACE. This value represents the point on the ROC curve where the trade-off between sensitivity and specificity is most favorable.

Statistical Analysis

Descriptive statistics were used to summarize the data. Continuous variables were presented as mean ± standard deviation (SD) when normally distributed, or median, with interquartile range when non-normally distributed. Categorical variables were expressed as counts and percentages.

Comparisons of delta troponin status, age groups, chest pain type, ECG findings, and sex were made with respect to 30-day mortality and non-fatal MACE using the chi-square test for categorical data.

All statistical analyses were conducted using IBM SPSS Statistics version 25 (SPSS Inc., Chicago, IL, USA). A two-tailed p value of <0.05 was considered statistically significant. ROC curve analysis was employed to evaluate the association between delta troponin and 30-day mortality or non-fatal MACE. ROC curves were generated to assess the discriminatory performance of troponin change for survival and non-fatal MACE. Optimal cut-off values, along with their 95% confidence intervals (CI) and areas under the curve (AUC), were determined using Youden's index and the DeLong method in MedCalc Statistical Software (trial version 15.8; MedCalc Software bvba, Ostend, Belgium; http://www.medcalc.org).

RESULTS

During the study period, troponin measurements were performed in 4,650 patients. Of these, 1,259 were excluded because at least one troponin value (baseline or 4-hour) exceeded 0.05 ng/mL, and 410 met predefined exclusion criteria. Consequently, 2,981 patients were included in the analysis: 43.7% (n=1,304) in the delta+positive group and 56.3% (n=1,677) in the delta-negative group.

There were statistically significant differences in both sex distribution and age groups between the delta+positive and delta-negative cohorts (Table 1). The prevalence of risk factors, symptom characteristics, ECG findings, and comorbidities is also detailed in Table 1. Overall, hypertension was the most common risk factor (50.6%). When comparing risk factors by sex, only coronary artery disease was significantly more frequent in men than in women (p=0.036); there were no significant sex differences for diabetes mellitus, hypertension, or hyperlipidemia (all p>0.05).

Baseline cTnI levels were 0.009±0.012 ng/mL (mean ± SD; range 0.000–0.050), and 4-hour cTnI levels were 0.014±0.015 ng/mL (range: 0.000-0.060). The mean delta troponin was 0.006±0.008 ng/mL (range: 0.000-0.050). Compared to the delta-negative group, the delta+positive group had significantly higher rates of non-fatal MACE (p<0.001) and mortality (p=0.013) (Table 2). A total of 112 patients (8.6%) in the delta+positive group and 100 patients (6%) in the delta-negative group required admission to the coronary intensive care unit.

Among patients presenting with cardiac-type chest pain, there was no statistically significant difference in 30-day mortality between the two groups (p=0.173), whereas non-fatal MACE rates differed significantly (p<0.001). A history of hypertension, diabetes mellitus, hyperlipidemia, or coronary artery disease was significantly associated with non-fatal MACE within 30 days (p<0.001), but none of these risk factors showed a significant relationship with mortality (all p>0.05).

ECG abnormalities were significantly associated with both non-fatal MACE (p<0.001) and mortality (p=0.003). Multivariable logistic regression analysis identified independent predictors of 30-day mortality and non-fatal MACE; the results are presented in Table 2.

In ROC curve analysis, a delta troponin threshold ≥0.005 ng/mL yielded a sensitivity of 63.4% and specificity of 37.7% for predicting non-fatal MACE (AUC: 0.635; 95% CI, 0.607-0.664). For predicting 30-day mortality, the same threshold demonstrated a sensitivity of 61.2% and specificity of 41.1% (AUC: 0.609; 95% CI, 0.528-0.690). The ROC curves for delta troponin versus non-fatal MACE and mortality are shown in Figure 1.

Table 1. Demographic and general clinical characteristics of the patients

		Delta+group (%)		Delta-group (%)		p
Sex	Female	616 (47.2)		861 (51.3)		0.026
	Male	688 (52.8)		816 (48.7)		
Age groups	<45 y	186 (14.3)		365 (18.5)		<0.001
	45-65 y	481 (36.9)		711 (42.4)		
	>65 y	637 (48.8)		601 (35.8)		
Symptoms	CPRS	528 (40.5)		677 (40.4)		0.016
	NCCP	706 (54.1)		945 (56.4)		
	CCP	70 (5.4)		50 (3.3)		
ECG	Normal	927 (71.1)		1365 (81.4)		<0.001
	NSC	352 (27)		292 (17.4)		
	ST-segment depression	25 (1.9)		20 (1.2)		
Comorbidities	DM	Present	301 (23.1)	Present	337 (20.1)	0.051
		Absent	1003 (76.9)	Absent	1340 (79.9)	
	HT	Present	718 (55.1)	Present	790 (47.1)	<0.001
		Absent	586 (44.9)	Absent	887 (52.9)	
	HL	Present	465 (35.7)	Present	489 (29.2)	<0.001
		Absent	839 (64.3)	Absent	1188 (70.8)	
	CAD	Present	325 (25.9)	Present	293 (17.5)	<0.001
		Absent	929 (75.1)	Absent	1384 (82.5)	

CPRS: Chest pain–related symptoms, NCCP: Non-cardiac chest pain, CCP: Cardiac chest pain, NSC: Non-specific ECG changes, ECG: Electrocardiographic, DM: Diabetes mellitus, HT: Hypertension, HL: Hyperlipidemia, CAD: Coronary artery disease

Table 2. Factors affecting non-fatal major adverse cardiac events and mortality

Variable	Non-fatal MACE odds ratio (95% CI)	p	Mortality odds ratio (95% CI)	p
Cardiac chest pain	7.27 (4.43-11.95)	<0.05		
ECG-non-specific changes	4.36 (3.38-5.63)	<0.05	4.36 (3.38-5.63)	<0.05
ST-segment depression	11.75 (2.96-46.64)	<0.05	32.21 (13.29-78.07)	<0.05
Age >65			12.34 (2.87-53.08)	<0.05
Hyperlipidemia	4.16 (2.92-5.91)	<0.05		
Absence of hypertension			2.72 (1.29-5.77)	<0.05
Absence of CAD	1.53 (1.10-2.14)	<0.05		
Presence of delta troponin	1.61 (0.89-2.91)	0.12	2.67 (2.09-3.43)	<0.05

CAD: Coronary artery disease, MACE: Major adverse cardiac events, CI: Confidence interval, ECG: Electrocardiographic

DISCUSSION

Studies on delta troponin have largely focused on hs-cTn assays, while data regarding delta changes in conventional troponin I (cTnI) levels remain limited in the literature. In our study, we found that even minor changes in cTnI values within the normal reference range were statistically associated with 30-day MACE. Among patients presenting to the emergency department with chest pain and monitored for cardiac biomarkers, temporal variations in cTnI-despite remaining below the diagnostic threshold-appeared to be associated with subsequent cardiac events. These findings suggest that even modest delta increases in cTnI may warrant closer clinical monitoring due to their potential prognostic relevance.

cTn became popular for ACS diagnosis in the early 1990s and are now considered fundamental laboratory parameters alongside clinical assessment and ECG.[10] The delta cTn value is defined as the difference between an

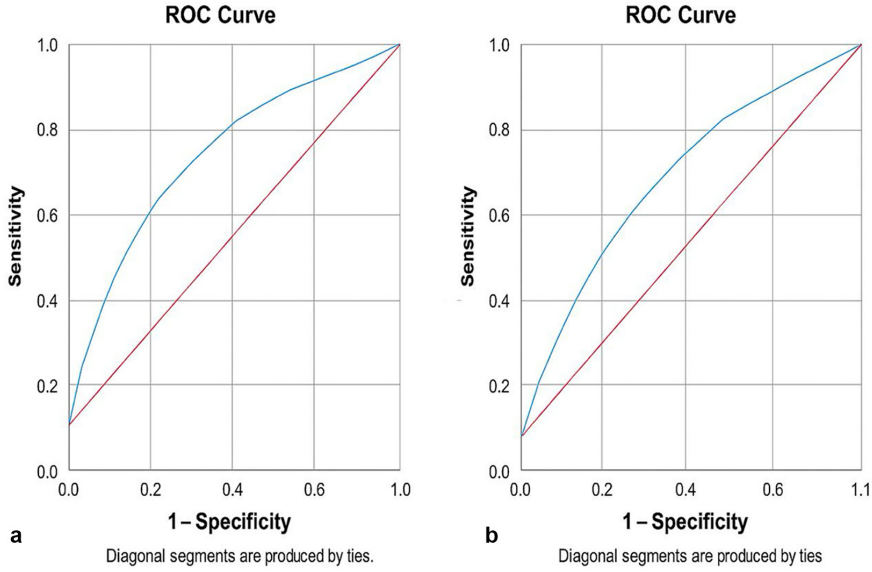


Figure 1. (a) Delta troponin-MACE ROC curve (b) delta troponin-mortality ROC curve

MACE: Major adverse cardiac events, ROC: Receiver operating characteristic

initial cTn concentration and a subsequent measurement; expressing this change as a percentage has been recommended.[11] High-sensitivity troponin I (hs-TnI) and its delta rate (delta hs-TnI) have recently been adopted for early ACS diagnosis with excellent performance. Studies have shown that a negative hs-TnI result safely facilitates discharge of chest pain, patients, whereas elevated hs-TnI values-irrespective of ACS-signal poorer prognosis and necessitate further evaluation.[12,13] Weir et al.[13] reported that delta hs-TnI changes of less than 20% can classify patients as low risk for ACS.

Because hs-TnI assays are not universally available, many emergency departments continue to rely on conventional cTnI measurements, which require longer monitoring intervals.[11,12] In our center, cTnI is expected to rise between 2 and 4 hours following myocardial injury. Accordingly, we compared baseline (0-hour) and 4-hour cTnI values and examined their relationship with 30-day mortality and non-fatal MACE.

In a prospective observational study, Suh et al.[14] evaluated the 0-to 1-hour delta hs-TnI alongside the HEART score to predict 30-day MACE. They designated patients as low risk if they had a modified HEART score of 0-3, symptom onset >3 hours before presentation, and hs-TnI <6 ng/L, and discharged them after a single hs-TnI measurement. Women with hs-TnI ≤14 ng/L and men with hs-TnI ≤22 ng/L had a follow-up hs-TnI obtained, and those with an hourly delta hs-TnI increase of <2.5 ng/L were deemed low risk and discharged. Conversely, patients exhibiting hourly increases >2.5 ng/L or elevated follow-up hs-TnI (>22 ng/L in men, >14 ng/L in women) were considered high risk.

Age-related increases in chest pain presentations and ACS incidence have also been reported.[15,16] Advanced age serves as a risk factor partly because other cardiovascular risk factors become more prevalent over time, increasing overall disease burden. In our cohort, delta+positive and delta-negative groups differed significantly by age and sex ($p=0.020$ and $p<0.001$, respectively); older age correlated with greater delta cTnI changes. Nevertheless, the rising incidence of ACS among individuals under 45-likely due to sedentary lifestyles and urbanization-should not be overlooked. In our study, male sex was independently associated with delta troponin elevation, irrespective of age.

James et al.[17] found that patients with mildly elevated cTnT levels (0.01-0.1 µg/L) faced a higher risk compared to those with completely normal values. Cullen et al.[11] examined cTnI changes at 2 and 6 hours post-presentation, reporting that delta changes exceeding the 99th percentile were highly indicative of ACS. Consistent with these findings, our delta+positive group experienced significantly higher non-fatal MACE rates ($p<0.05$).

Although hs-TnI measurement allows rapid risk stratification-particularly useful in busy emergency departments-our center's reliance on conventional cTnI necessitated longer observation periods. Given increasing patient volumes and operational challenges in Turkish

emergency departments, broader adoption of hs-TnI assays could enhance efficient and accurate patient management.

Several scoring systems support safe discharge of low-risk ACS patients by demonstrating <1% 30-day MACE risk; these incorporate hypertension, hyperlipidemia, diabetes mellitus, and coronary artery disease history among their variables.[18,19] The Thrombolysis in Myocardial Infarction score, developed by Antman et al.[18] in 2000, similarly employs these risk factors. We evaluated these risk factors, chest pain characteristics, and ECG findings across delta+positive and delta-negative groups. Our observations regarding chest pain character and ECG ischemic changes (ST-segment depression and other ischemic signs) aligned with existing literature. However, in contrast to previous studies, our findings revealed that the absence of a history of coronary artery disease was associated with a higher incidence of non-fatal MACE, while the absence of hypertension was linked to increased mortality. This discrepancy may be attributed to the exclusion of STEMI and non-STEMI patients from our study, as well as the likelihood that individuals with known coronary artery disease or hypertension are under regular cardiology follow-up and receive optimal medical therapy, potentially mitigating the impact of these risk factors on mortality. Future prospective studies are warranted to further investigate the influence of hypertension on mortality in these populations.

Fesmire et al.[19] reported that a cTnI increase ≥0.2 ng/mL predicted MACE with 87.7% sensitivity and 61.4% specificity. That study included all emergency department presentations and focused on changes above the cut-off. Mueller et al.[20] examined patients with at least one troponin value above the assay cut-off, finding 88% specificity and 60% sensitivity for delta troponin in predicting MACE. Our study differs by investigating delta changes below the assay cut-off and their association with 30-day MACE.

Although delta troponin changes in our study were found to be clinically significant, the obtained AUC values indicate a limited discriminatory power. Therefore, combining delta troponin changes with established risk scoring systems in the literature (e.g., the HEART score) may enhance predictive accuracy. Prospective, multicenter studies are needed to test this hypothesis.

Study Limitations

The retrospective nature of the study limited the comprehensive evaluation of dynamic parameters and clinical decision-making in patient management. Key limitations include the inability to determine the precise onset time of symptoms, the lack of standardized follow-up procedures, and the absence of data on certain established risk factors such as family history, central obesity, and smoking status. Prospective follow-up studies are needed to validate our findings. Furthermore, the single-center design of the study restricts the generalizability of the results.

CONCLUSION

In conclusion, although delta changes in cTnI within the normal range did not predict 30-day mortality, they were significantly associated with non-fatal MACE. Monitoring delta cTnI in chest pain patients may thus aid in early identification of those at higher risk for adverse cardiac events.

Ethics

Ethics Committee Approval: This study was approved by the Clinical Research Ethics Committee of University of Health Sciences Türkiye, İzmir Bozyaka Training and Research Hospital (decision number: 7, date: 04.07.2018).

Informed Consent: Retrospective study.

Note: This study is derived from a medical specialty thesis.

Footnotes

Authorship Contributions

Concept: A.T., H.G., C.S., İ.P., Design: A.T., H.G., C.S., İ.P., Data Collection or Processing: A.T., Analysis or Interpretation: A.T., Literature Search: A.T., Writing: A.T., H.G., C.S.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Stepińska J, Lettino M, Ahrens I, et al. Diagnosis and risk stratification of chest pain patients in the emergency department: focus on acute coronary syndromes. Eur Heart J Acute Cardiovasc Care. 2020;9:76-89.
2. Gulati M, Levy PD, Mukherjee D, et al. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR Guideline for the Evaluation and Diagnosis of Chest Pain: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2021;144:368-454.
3. Pope JH, Aufderheide TP, Ruthazer R, et al. Missed diagnoses of acute cardiac ischemia in the emergency department. N Engl J Med. 2000;342:1163-70.
4. Amsterdam EA, Kirk JD, Bluemke DA, et al. Testing of low-risk patients presenting to the emergency department with chest pain: a scientific statement from the American Heart Association. Circulation. 2010;122:1756-76.
5. Peacock WF, Baumann BM, Bruton D, et al. Efficacy of high-sensitivity troponin T in identifying very-low-risk patients with possible acute coronary syndrome. JAMA Cardiol. 2018;3:104-11.
6. Chapman AR, Anand A, Boeddinghaus J, et al. Comparison of the efficacy and safety of early rule-out pathways for acute myocardial infarction. Circulation. 2017;135:1586-96.

7. Wu AH, Apple FS, Gibler WB, Jesse RL, Warshaw MM, Valdes RJ Jr. National academy of clinical biochemistry standards of laboratory practice: recommendations for the use of cardiac markers in coronary artery diseases. Clin Chem. 1999;45:1104-21.
8. Salimi A, Zolghadrasli A, Jahangiri S, et al. The potential of HEART score to detect the severity of coronary artery disease according to SYNTAX score. Sci Rep. 2023;13:7228.
9. Long B, Oliver J, Streitz M, Koyfman A. An end-user's guide to the HEART score and pathway. Am J Emerg Med. 2017;35:1350-5.
10. Vaz HA, Guimaraes RB, Dutra O. Challenges in high-sensitive troponin assay interpretation for intensive therapy. Rev Bras Ter Intensiva. 2019;31:93-105.
11. Cullen L, Parsonage WA, Greenslade J, et al. Delta troponin for the early diagnosis of AMI in emergency patients with chest pain. Int J Cardiol. 2013;168:2602-8.
12. Collinson PO. Sensitive troponin assays. J Clin Pathol. 2011;64:845-49.
13. Weir RAP, Osmanska J, Docherty KF, Petrie CJ. Chest pain with less than 20% change in high-sensitivity troponin T-a low risk cohort? Acta Cardiol. 2020;75:149-55.
14. Suh EH, Probst MA, Tichter AM, et al. Flexible-interval high-sensitivity troponin velocity for the detection of acute coronary syndromes. Am J Cardiol. 2023;203:240-7.
15. Ruigómez A, Rodríguez LAG, Wallander M-A, Johansson S, Jones R. Chest pain in general practice: incidence, comorbidity and mortality. Fam Pract. 2006;23:167-74.
16. Mokhtari A, Dryver E, Söderholm M, Ekelund U. Diagnostic values of chest pain history, ECG, troponin and clinical gestalt in patients with chest pain and potential acute coronary syndrome assessed in the emergency department. Springerplus. 2015;4:219.
17. James S, Armstrong P, Califf R, et al. Troponin T levels and risk of 30-day outcomes in patients with the acute coronary syndrome: prospective verification in the GUSTO-IV trial. Am J Med. 2003;115:178-84.
18. Antman EM, Cohen M, Bernink PJ, et al. The TIMI risk score for unstable angina/non-ST elevation MI. JAMA. 2000;284:835-42.
19. Fesmire FM, Hughes AD, Fody EP, et al. The Erlanger chest pain evaluation protocol: a one-year experience with serial 12-lead ECG monitoring, two-hour delta serum marker measurements, and selective nuclear stress testing to identify and exclude acute coronary syndromes. Ann Emerg Med. 2002;40:584-94.
20. Mueller M, Biener M, Vafaie M, et al. Absolute and relative kinetic changes of high-sensitivity cardiac troponin T in acute coronary syndrome and in patients with increased troponin in the absence of acute coronary syndrome. Clin Chem. 2012;58:209-18.

DOI: 10.4274/forbes.galenos.2025.53244
Forbes J Med 2025;6(2):172-9

Original Article / Orijinal Araştırma

AI-Driven Clinical Guidance in Necrotizing Enterocolitis: Concordance with European Neonatal Care Standards

Nekrotizan Enterokolitte Yapay Zeka Destekli Klinik Rehberlik: Avrupa Yenidoğan Bakım Standartları ile Uyumun Değerlendirilmesi

Saime SÜNDÜS UYGUN[1], Mehmet SARIKAYA[2]

[1]Selçuk University Faculty of Medicine, Department of Pediatrics, Division of Neonatology, Konya, Türkiye
[2]Selçuk University Faculty of Medicine, Department of Pediatric Surgery, Konya, Türkiye

Cite as: Sündüs Uygun S, Sarıkaya M. AI-driven clinical guidance in necrotizing enterocolitis: concordance with European Neonatal care standards. Forbes Med J. 2025;6(2):172-9

ABSTRACT

Objective: Necrotizing enterocolitis (NEC) is a life-threatening emergency in neonatal medicine, especially among premature infants, with high morbidity and mortality. Despite the availability of evidence-based resources such as the European Standards of Care for Newborn Health, variations in clinical practice persist. Artificial intelligence (AI) systems based on large language models have recently gained attention as tools to support clinical decision-making. This study aimed to evaluate the alignment of two widely used AI applications-Chat Generative Pre-trained Transformer (ChatGPT) version 4.0 and Gemini-with the European neonatal care standards in providing clinical recommendations for the management of NEC.

Methods: Forty clinical questions were prepared based on the European guidelines, covering diagnosis, treatment, nutrition, follow-up, and ethical issues. Both AI models were queried under identical conditions. Their responses were independently evaluated by a pediatric surgeon and a neonatologist using a five-point Likert scale. Inter-rater agreement and statistical comparisons were analyzed using Spearman's correlation, Cohen's kappa coefficient, and the Wilcoxon signed-rank test.

Results: ChatGPT received mean scores of 4.53 from both reviewers, while Gemini received scores of 4.33 and 4.40. Median scores for both models ranged from four to five. Spearman's correlation indicated moderate agreement between reviewers, while Cohen's kappa showed weak agreement. No statistically significant differences were found between reviewers for either model.

Conclusion: Both AI models showed high compliance with European neonatal care standards in NEC management. These findings support their potential role as supportive tools in neonatal clinical decision-making.

Keywords: Necrotizing enterocolitis, large language models, artificial intelligence, clinical guidelines, ESCNH, neonatal care, ChatGPT, Gemini

ÖZ

Amaç: Nekrotizan enterokolit (NEC), özellikle prematüre yenidoğanlarda yüksek morbidite ve mortaliteye yol açan ciddi bir klinik tablodur. European Standards of Care for Newborn Health adlı kanıta dayalı rehbere rağmen klinik uygulamalarda anlamlı düzeyde farklılıklar devam etmektedir. Son yıllarda, büyük dil modeli temelli yapay zeka (AI) uygulamaları klinik karar destek sistemleri olarak dikkat çekmektedir. Bu çalışmada, *Chat Generative Pre-trained Transformer (*ChatGPT)-4.0 ve Gemini isimli iki AI uygulamasının, NEC yönetimine yönelik olarak European Standards of Care for Newborn Health rehberi temelinde oluşturulan klinik önerilerle uyum düzeylerinin değerlendirilmesi amaçlanmıştır.

Yöntem: Söz konusu rehbere dayanarak tanı, tedavi, beslenme, izlem ve etik konularını içeren 40 açık uçlu soru hazırlanarak her iki AI uygulamasına aynı koşullarda yöneltilmiştir. Yanıtlar, bir çocuk cerrahı ve bir yenidoğan uzmanı tarafından beş puanlık Likert ölçeğiyle bağımsız şekilde değerlendirilmiştir. Değerlendiriciler arası uyum ve iki modelin karşılaştırması Spearman korelasyonu, Cohen kappa katsayısı ve Wilcoxon işaretli sıra testi ile analiz edilmiştir.

Received/Geliş: 12.05.2025
Accepted/Kabul: 04.07.2025
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Corresponding Author/ Sorumlu Yazar:
Saime SÜNDÜS UYGUN, MD,
Selçuk University Faculty of Medicine, Department of Pediatrics, Division of Neonatology, Konya, Türkiye
✉ uygunsaime@hotmail.com
ORCID: 0000-0002-6694-8115

Bulgular: ChatGPT-4.0 her iki uzman tarafından 4,53 puan ortalamasıyla değerlendirilirken, Gemini için puanlar sırasıyla 4,33 ve 4,40 olarak bulunmuştur. Medyan puanlar her iki model için dört ile beş arasında değişmiştir. Değerlendiriciler arasında istatistiksel fark saptanmamıştır.

Sonuç: AI uygulamaları, NEC yönetiminde rehber uyumunu yüksek düzeyde sağlamış olup klinik karar süreçlerinde destekleyici araçlar olarak potansiyel taşımaktadır.

Anahtar Kelimeler: Nekrotizan enterokolit, büyük dil modelleri, yapay zeka, klinik kılavuzlar, ESCNH, yenidoğan bakımı, ChatGPT, Gemini

INTRODUCTION

Necrotizing enterocolitis (NEC) is a serious gastrointestinal emergency that primarily affects premature and low birth weight infants. The disease is characterized by intestinal inflammation, bacterial invasion, and, in advanced cases, necrosis and perforation of the intestinal wall. Despite advances in neonatal intensive care, NEC remains associated with elevated morbidity and mortality rates. In cases of NEC requiring surgical intervention, the mortality rate has been documented to exceed 30% in several series.[1] It has been shown that surviving infants are at a heightened risk of experiencing long-term complications, including but not limited to short bowel syndrome, growth retardation, and neurodevelopmental disorders.[2,3] The unpredictable clinical course and multifactorial etiology of NEC make timely diagnosis and effective management challenging; therefore, a standards-based, evidence-based care approach is necessary.

The clinical management of NEC poses significant challenges due to the highly variable course of the disease, the potential for rapid progression, and the absence of definitive diagnostic markers. The decisions regarding medical or surgical treatment, the timing of treatment steps, and nutritional strategies can vary significantly between institutions and clinicians. Despite the existence of international guidelines, such as the European Standards for Neonatal Health (ESCNH), which provide structured frameworks for the management of NEC, there is a paucity of research on the real-time implementation of these guidelines at the bedside.[4] This is due to limitations imposed by clinical complexities, differences in experience, and access issues. These discrepancies in implementation underscore the necessity for support systems that can seamlessly integrate clinical guideline recommendations into decision-making processes.

In recent years, there has been a marked increase in the integration of artificial intelligence (AI) into clinical applications, with its use extending to many areas, including diagnostic imaging, risk stratification, treatment planning, and clinical decision support systems.[5] In the field of neonatal care, AI has been employed to facilitate early prediction of diseases such as sepsis, respiratory distress syndrome, and intraventricular haemorrhage in studies numbered.[5-8] Recent developments have demonstrated the integration of AI applications into the domain of nutrition planning, as evidenced by the implementation of a system known as TPN 2.0, which has been shown to enhance adherence to parenteral nutrition guidelines and reduce the incidence of complications in premature infants.[9,10] Concurrently, machine learning models that amalgamate imaging and clinical data for the diagnosis of NEC have been developed and demonstrated an enhancement in diagnostic accuracy.[11-13] These developments demonstrate the expanding role of AI in neonatal intensive care and its potential contribution to the delivery of timely and standardised care to vulnerable patient groups.

Despite the encouraging outcomes observed in diverse domains of neonatal care, ranging from early diagnosis to nutrition planning, there is a lack of studies that systematically assess the clinical guideline alignment of LLMs in the context of NEC. Existing literature has focused more on predictive algorithms or imaging-based diagnostic systems, neglecting the assessment of textual clinical reasoning capabilities.[11-13] To address this need, this study aims to evaluate the degree to which large language model (LLM)-based AI applications align with established clinical guidelines in NEC scenarios. Demonstrating that such systems can produce consistent, guideline-compliant responses is critical not only for validating their safety and reliability but also for facilitating their integration into real-world neonatal care. AI systems capable of aligning with structured standards such as the ESCNH could assist clinicians by reinforcing best practices, reducing variability in care, and supporting decision-making processes in time-sensitive and resource-limited settings.

METHODS

Study Design

Both AI applications were accessed in their publicly available web-based versions: Chat Generative Pre-trained Transformer (ChatGPT)-4 (OpenAI, May 2025) and Gemini Pro 1.5 (Google Bard, accessed May 2025). No prompt engineering, temperature adjustment, or context priming was used. Each of the 40 clinical questions was presented as a plain-text prompt during a single uninterrupted session per model. The ESCNH guideline document was not uploaded or attached to the prompt; the AI models were expected to generate responses based on their internal training data and knowledge up to the date of access. To ensure standardization, each question was

asked only once per model, without regeneration or multiple attempts. All responses were recorded in their original form, reflecting a single-use, real-time interaction to simulate a realistic clinical consultation scenario. No post-processing or modification of AI responses was performed. To minimize evaluator bias, the responses from ChatGPT and Gemini were anonymized and randomly ordered before being presented to the reviewers. Both evaluators were blinded to the source of the AI responses during the scoring process. To enhance transparency and clinical relevance, one example case-including the prompt and responses generated by both AI systems-has been provided in the supplementary material (Appendix A).

A total of 40 open-ended clinical questions were created based on the official recommendations of the European Standards of Care for ESCNH, covering core areas such as diagnosis, treatment, follow-up, communication, nutrition, and ethical considerations in NEC care. These questions were presented separately to both AI applications under identical conditions, and each response was recorded in its original form without modification.

The responses generated by each AI application were then independently assessed by two specialists: a paediatric surgeon and a neonatologist, both with more than 10 years of clinical experience in neonatal care. The reviewers scored each AI response using a structured 5-point Likert scale to determine its level of agreement with the ESCNH guideline recommendations. The 5-point Likert scale used by the reviewers was structured as follows: 1 = not compliant with the guideline, 2 = low compliance, 3 = moderate compliance, 4 = high compliance, and 5 = fully compliant with the guideline. This methodological design was chosen to simulate real-world use scenarios and to assess the comparative performance of AI models in providing evidence-based clinical guidance.

Statistical Analysis

All statistical analyses were conducted using IBM SPSS Statistics version 29.0 (IBM Corp., Armonk, NY). Descriptive statistics were presented as medians with interquartile ranges (IQR), in accordance with the non-parametric nature of the data. Inter-rater agreement was assessed using Spearman's rank correlation coefficient (ρ) for ordinal data and Cohen's kappa (κ) for categorical agreement. The Wilcoxon signed-rank test was used to compare the Likert scores assigned by the pediatric surgeon and the neonatologist for each AI system, as the data were not normally distributed. A p value <0.05 was considered statistically significant.

Ethical Consideration

Since this study did not involve human or animal subjects and was limited to AI-based textual response evaluation, it was exempt from ethics committee approval. As the evaluated responses were AI-generated and did not involve patient data or human participation, the study posed no ethical risk.

RESULTS

The mean Likert score for ChatGPT responses was 4.53±0.72 based on pediatric surgeon evaluations and 4.53±0.64 based on neonatologist ratings. Similarly, Gemini responses received an average score of 4.33±0.69 from the pediatric surgeon and 4.40±0.55 from the neonatologist. Median values for both AI systems were 5, with an IQR of 4-5 for both reviewers across all models.

A moderate and statistically significant correlation was observed between the two reviewers' Likert scores for both AI systems (Figure 1 for Gemini, Figure 2 for ChatGPT). For ChatGPT, Spearman's rho was 0.41 (p=0.008), while for Gemini, the correlation coefficient was 0.35 (p=0.027). These findings suggest reasonable consistency in expert assessments, despite minor variations.

Cohen's kappa statistic revealed weak categorical agreement between the reviewers, with κ=0.10 for ChatGPT and κ=0.21 for Gemini. This low kappa is likely due to the ordinal nature of the Likert scale rather than true disagreement.

Although the vast majority of responses received scores of 4 or 5, a few instances of moderate compliance (score=3) were observed. These cases often involved complex, context-sensitive issues such as ethics or multidisciplinary coordination, where even human experts may interpret guideline recommendations differently. Such variability underscores the importance of expert oversight when integrating AI tools into clinical workflows.

Wilcoxon signed-rank tests demonstrated no statistically significant differences in Likert scores between the pediatric surgeon and neonatologist for either AI system (Table 1). None of the AI responses received a score of 1 (not compliant) or 2 (low compliance) from either expert. Only 3 of the total 80 ratings (3.75%) were scored as 3 (moderate compliance), reflecting some ambiguity in guideline interpretation in specific ethical or interdisciplinary domains. The p value was 1.000 for ChatGPT and 0.513 for Gemini; however, this does not necessarily indicate strong consistency in reviewer scoring within each AI system. Although ChatGPT consistently received slightly higher scores than Gemini, the differences were not statistically significant for either reviewer (p=0.133 for

pediatric surgeon; p=0.219 for neonatologist), indicating comparable levels of guideline adherence between the two AI systems (Table 2).

DISCUSSION

In this prospective comparative study, two widely used LLMs, ChatGPT and Gemini, were evaluated for their adherence to the ESCNH in the context of NEC. Both systems demonstrated a high level of guideline compliance, with ChatGPT achieving slightly higher mean Likert scores than the two expert raters. Notably, 96.25% of the expert scores fell within the 4-5 range on the Likert scale, reflecting a consistently high or full compliance with ESCNH recommendations across both AI platforms. These findings highlight the potential of LLMs as decision support tools in neonatal care, particularly in the management of

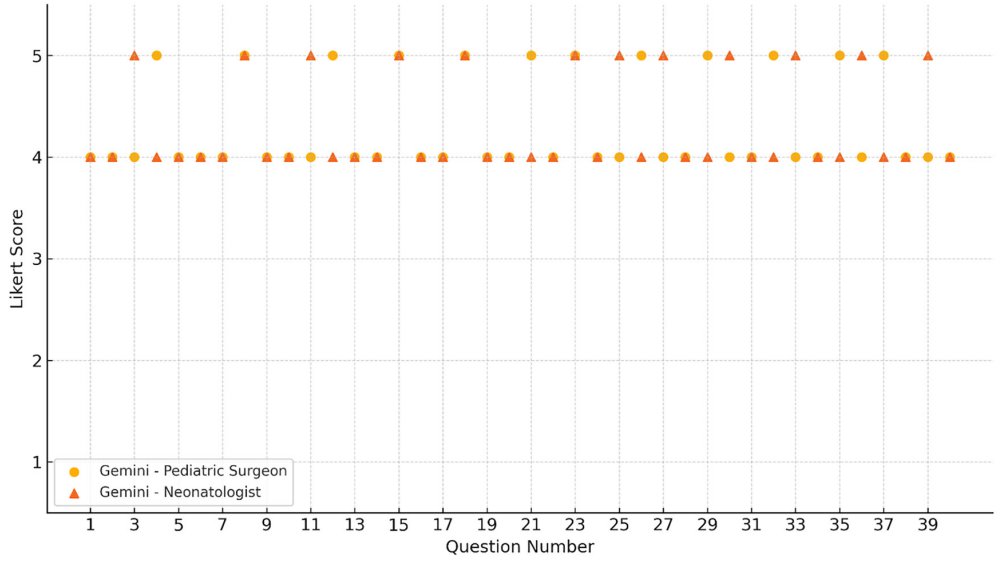


Figure 1. Reviewer scores for Gemini across 40 clinical questions. Scatterplot shows Likert scores assigned by the pediatric surgeon and neonatologist for each Gemini-generated response. Scores range from 1 (not compliant) to 5 (fully compliant) based on ESCNH guideline adherence

ESCNH: European Standards for Neonatal Health

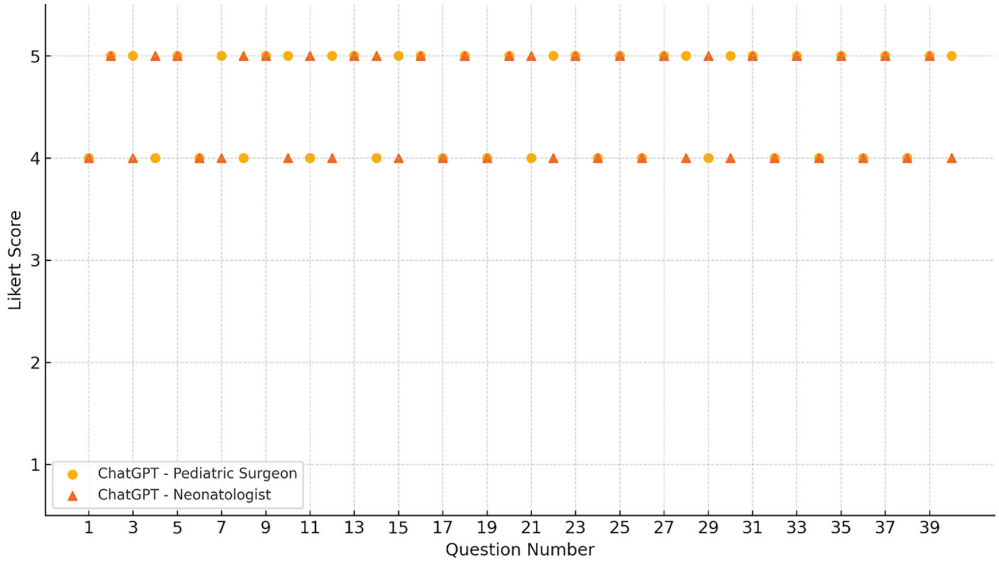


Figure 2. Reviewer scores for ChatGPT across 40 clinical questions. Scatterplot illustrates the distribution of Likert scores assigned by the pediatric surgeon and neonatologist for ChatGPT-generated responses. Responses were evaluated for alignment with ESCNH recommendations

ChatGPT: Chat Generative Pre-trained Transformer, ESCNH: European Standards for Neonatal Health

Table 1. Reviewer scores and inter-rater agreement metrics for ChatGPT and Gemini					
AI system	**Reviewer**	**Median (IQR)**	**Spearman ρ (p)**	**Cohen's κ**	**Wilcoxon p**
ChatGPT	Pediatric surgeon	5 (4-5)	0.41 (p=0.008)	0.10	1.000
ChatGPT	Neonatologist	5 (4-5)			
Gemini	Pediatric surgeon	4 (4-5)	0.35 (p=0.027)	0.21	0.513
Gemini	Neonatologist	4 (4-5)			
(Scoring based on a 5-point Likert scale: 1 = not compliant, 2 = low compliance, 3 = moderate, 4 = high, 5 = fully compliant with ESCNH guidelines) ChatGPT: Chat Generative Pre-trained Transformer, IQR: Interquartile range					

Table 2. Comparison of mean scores for ChatGPT and Gemini by Reviewer				
Reviewer value	**ChatGPT median (IQR)**	**Gemini median (IQR)**	**Mann-Whitney U**	**p value**
Pediatric surgeon	5 (4-5)	4 (4-5)	940.00	0.133
Neonatologist	5 (4-5)	4 (4-5)	913.50	0.219
(Higher scores reflect stronger alignment with ESCNH guidelines) ChatGPT: Chat Generative Pre-trained Transformer, IQR: Interquartile range, ESCNH: European Standards for Neonatal Health				

complex and time-sensitive situations where adherence to standard protocols, such as NEC, is critical.

Furthermore, the moderate inter-rater agreement observed using Spearman's correlation suggests a reassuring level of consistency in expert judgement. Although Cohen's kappa values were relatively low, this is likely to be due to methodological limitations associated with the application of categorical agreement statistics to ordinal data, rather than a genuine lack of consensus. Overall, the findings position LLMs as promising adjunct tools in neonatal clinical reasoning when linked to structured clinical frameworks such as ESCNH guidelines.

While numerous studies have explored applications of AI in general medicine and adult subspecialties, relatively few studies have systematically evaluated LLMs in neonatal care using structured, guideline-based assessments. For example, Singer et al.[9] and Phongpreecha et al.[10] reported encouraging results with AI systems in neonatal nutrition planning and precision parenteral nutrition strategies, but these studies focused on algorithm-driven interventions rather than textual clinical reasoning. Similarly, Sullivan et al.[7] highlighted the role of AI in predicting neonatal sepsis and respiratory failure through data modeling, but these approaches did not include direct guideline adherence.[6]

In contrast, our study offers several novel directions: it addresses a critical NEC, it uses a rigorously developed question set fully derived from the ESCNH guidelines, and it uses a dual expert assessment model involving both a paediatric surgeon and a neonatologist. To our knowledge, this is the first study to evaluate guideline concordance of LLM-generated clinical responses specifically in the context of NEC.

Previous research on AI-assisted NEC diagnosis has primarily focused on image-based tools, such as deep learning models for interpreting abdominal radiographs or integrating clinical data with imaging for early diagnosis.[11-13] While these approaches are valuable, they do not capture the narrative reasoning and context-sensitive decision making that is central to LLM performance. Therefore, our findings help to fill a critical gap by assessing how well LLMs reflect structured neonatal guideline recommendations through textual outputs. In addition, Sarikaya et al.[14] evaluated the guideline compatibility of multiple AI platforms in the management of vesicoureteral reflux and found that LLMs showed high levels of agreement with pediatric urology guidelines, strengthening their applicability in structured pediatric decision-making frameworks.

The structured nature of this study enhances both the content validity and the direct applicability of its findings to neonatal clinical decision-making. By formulating all questions strictly based on ESCNH recommendations, the study reflects practical scenarios that clinicians frequently encounter in neonatal intensive care settings. In addition, presenting the same set of questions to both AI models under identical conditions eliminates variability, and the blinded assessment of anonymized responses by two independent experts helps minimize contextual and cognitive biases. This methodology increases the internal validity of the results and supports the comparability of AI performance across platforms-an essential factor when considering their future integration into standardized care protocols or decision-support systems in NICUs. Second, the inclusion of two independent raters (a pediatric surgeon and a neonatologist) from different but complementary specialties provides a multidisciplinary perspective and increases the generalizability of the results across neonatal

care settings. Third, the prospective and standardized assessment design, including same-day referral and the same set of questions for both LLMs, minimizes potential contextual bias and ensures comparability.

However, certain limitations should also be recognized. Although widely used in health informatics research, the 5-point Likert scale used to assess guideline adherence is inherently subject to subjective interpretation, which may affect inter-rater reliability measures such as Cohen's kappa. Although Spearman's correlation indicates a moderate level of agreement between raters, low kappa values may reflect the ordinal nature of the scale rather than significant disagreement. In addition, only two LLM-based systems were evaluated, limiting generalizability to other existing or new models.

The results of this study support the potential integration of LLM-based systems into neonatal clinical settings, particularly in roles involving decision support, guidelines training and clinical education. Given the high level of alignment with ESCNH recommendations demonstrated by both ChatGPT and Gemini, such tools may be particularly useful in settings where access to neonatal subspecialists is limited or where rapid dissemination of standardized information is required. LLMs can be incorporated into electronic health records as real-time, just-in-time decision support tools that reinforce evidence-based practice and reduce variability in care.

However, caution should be exercised when interpreting these findings for real-world application. Although the LLMs evaluated provided largely guideline-compliant responses, occasional inconsistencies or omissions were noted, highlighting the ongoing need for expert review and validation. Future research should aim to develop domain-specific LLMs trained on neonatal data, validate them in different clinical settings, and explore their impact on diagnostic accuracy, clinician efficiency, and parental satisfaction. In addition, ethical considerations, including transparency, privacy, and liability issues in AI-assisted clinical decision-making, should be systematically addressed before widespread implementation.

This study provides the first evidence that LLMs can comply with neonatal clinical guidelines in complex scenarios such as necrotising enterocolitis, providing a promising basis for future AI-assisted care in neonatology.

The integration of LLM-based systems into NICU workflows may offer several practical advantages, such as providing real-time clinical decision support, reinforcing adherence to standardized guidelines, and serving as educational tools for junior clinicians or training simulations. Particularly in resource-limited or high-pressure settings, AI systems may act as supplementary aids to improve consistency in care. However, their implementation must be approached cautiously. Medicolegal concerns-such as liability in case of AI-related error, the interpretability of outputs, and the transparency of data sources-must be addressed through clear institutional policies. Furthermore, ethical considerations, including patient safety, data privacy, and clinician responsibility, remain critical to responsible deployment. Continuous human oversight should be maintained to validate AI outputs and ensure alignment with patient-specific contexts.

The low Cohen's kappa values observed in our study may suggest weak inter-rater agreement. However, this likely stems from the statistical limitations of using kappa with ordinal data, especially when most ratings are concentrated at the high end of the scale (i.e., scores of 4 and 5). In such skewed distributions, even minor differences can disproportionately affect the kappa coefficient. To provide a more appropriate measure, we also reported Spearman's rank correlation, which demonstrated moderate and statistically significant agreement between reviewers. Although other methods such as Gwet's AC1 or intraclass correlation coefficient could be employed, the 5-point Likert scale remain widely used in clinical guideline adherence studies and was selected for its clarity, ease of interpretation, and suitability for the scoring task assigned to expert raters.

Study Limitations

This study's findings should be interpreted with caution due to several limitations. Primarily, the evaluation was confined to two LLM-based AI systems, ChatGPT-4 and Gemini, which may restrict the generalizability of the results to other existing or forthcoming models. Secondly, each clinical prompt was presented to the AI systems only once, without employing multiple iterations or prompt refinement strategies. This approach may have influenced the observed response variability. Lastly, the ESCNH guideline was not directly integrated or linked to the AI systems. Consequently, the assessment reflects the models' inherent knowledge bases as of May 2025, which are subject to continuous evolution. Future research endeavors should aim to incorporate a broader spectrum of LLMs, simulate real-time clinical scenarios, and include a more comprehensive evaluation of patient outcomes to further validate these findings.

CONCLUSION

This study provides preliminary evidence that LLMs, such as ChatGPT-4 and Gemini, can generate clinical responses that align closely with the ESCNH in the context of NEC. The findings demonstrate high levels of guideline compliance, particularly in structured clinical scenarios,

and suggest that LLMs hold potential as supplementary decision-support tools in neonatal intensive care units.

While limitations exist-including the use of a single response per prompt, the inherent variability of AI outputs, and the reliance on Likert-based expert scoring-these do not compromise the study's overall validity. Instead, they highlight the importance of continued validation, transparent implementation, and responsible oversight.

As AI models continue to evolve, future studies should explore their integration into clinical workflows, training environments, and decision-making frameworks, ensuring that such technologies enhance, rather than replace, expert judgment in neonatal care.

Ethics

Ethics Committee Approval and Informed Consent: This study was exempt from ethics committee approval because it did not involve human or animal subjects and was limited to AI-based textual response evaluation. No ethics committee approval or patient consent was required for the study, as the evaluated responses were generated by AI and did not involve patient data or human participation.

Footnotes

Authorship Contributions

Concept: S.S.U., Design: M.S., Data Collection or Processing: M.S., Analysis or Interpretation: S.S.U., Literature Search: M.S., Writing: S.S.U.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Fitzgibbons SC, Ching Y, Yu D, et al. Mortality of necrotizing enterocolitis expressed by birth weight categories. J Pediatr Surg. 2009;44:1072-5.
2. Patel RM, Kandefer S, Walsh MC, et al. Causes and timing of death in extremely premature infants from 2000 through 2011. N Engl J Med. 2015;372:331-40.
3. Neu J, Walker WA. Necrotizing enterocolitis. N Engl J Med. 2011;364:255-64.
4. Lindacher V, Altebaeumer P, Marlow N, et al. European Standards of Care for Newborn Health-a project protocol. Acta Paediatr. 2021;110:1433-8.
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25:44-56.
6. Sullivan BA, Kausch SL, Fairchild KD. Artificial and human intelligence for early identification of neonatal sepsis. Pediatric Res. 2023;93:350-6.
7. Sullivan BA, Beam K, Vesoulis ZA, et al. Transforming neonatal care with artificial intelligence: challenges, ethical consideration, and opportunities. J Perinatol. 2024;44:1-11.
8. De Rose DU, Ronchetti MP, Martini L, et al. Diagnosis and management of neonatal bacterial sepsis: current challenges and future perspectives. Trop Med Infect Dis. 2024;9:199.
9. Singer P, Robinson E, Raphaeli O. The future of artificial intelligence in clinical nutrition. Current opinion in clinical nutrition and metabolic care. Curr Opin Clin Nutr Metab Care. 2024;27(2):200-6.
10. Phongpreecha T, Ghanem M, Reiss JD, et al. AI-guided precision parenteral nutrition for neonatal intensive care units. Nat Med. 2025;31:1882-94.
11. Sitek A, Seliga-Siwecka J, Płotka S, et al. Artificial intelligence in the diagnosis of necrotising enterocolitis in newborns. Pediatric Res. 2023;93:376-81.
12. Duci M, Verlato G, Moschino L, Uccheddu F, Fascetti-Leon F. Advances in artificial intelligence and machine learning for precision medicine in necrotizing enterocolitis and neonatal sepsis: a state-of-the-art review. Children (Basel). 2025;12:498.
13. Weller JH, Scheese D, Tragesser C, Yi PH, Alaish SM, Hackam DJ. Artificial intelligence vs. doctors: diagnosing necrotizing enterocolitis on abdominal radiographs. J Pediatr Surg. 2024;59:161592.
14. Sarikaya M, Ozcan Siki F, Ciftci I. Use of artificial intelligence in vesicoureteral reflux disease: a comparative study of guideline compliance. J Clin Med. 2025;14:2378

Appendix A.

Sample Case Scenario and AI Responses

Prompt

"A 28-week premature infant with abdominal distension, bile-stained gastric residuals, and signs of systemic instability is suspected of having necrotizing enterocolitis (NEC). According to European Standards for Neonatal Health (ESCNH) guidelines, how should this patient be managed in terms of diagnostics, nutrition, and surgical consultation?"

ChatGPT-4b Response

Based on the ESCNH guidelines, the management of a 28-week premature infant suspected of having NEC should include:

Diagnostics

- Immediate cessation of enteral feeds.
- Serial abdominal examinations and measurements.
- Abdominal X-ray to identify pneumatosis intestinalis, portal venous gas, or pneumoperitoneum.
- Laboratory testing including complete blood count, C-reactive protein, blood culture, and serum electrolytes.

Nutrition

- Initiation of total parenteral nutrition (TPN).
- Monitoring of fluid balance and metabolic parameters.

Surgical Consultation

- Early involvement of the pediatric surgery team.
- Indications for surgical evaluation include abdominal wall discoloration, fixed bowel loop on imaging, and clinical deterioration despite conservative measures.

Close multidisciplinary collaboration is essential, and transfer to a tertiary neonatal surgical center should be considered if surgical capacity is limited locally.

Gemini Response

For a 28-week gestational age infant showing signs suggestive of NEC, ESCNH guidelines recommend the following approach:

Diagnostic Approach

- Discontinue oral or tube feeding.
- Perform serial abdominal radiographs to detect early signs such as pneumatosis intestinalis.
- Laboratory evaluation with blood counts, inflammatory markers, and metabolic panels.

Nutritional Support

- Initiate TPN and maintain adequate hydration and electrolyte balance.
- Withhold enteral feeds until clinical and radiologic improvement is observed.

Surgical Considerations

- Consult pediatric surgery early.
- Signs like abdominal rigidity, erythema, metabolic acidosis, and free air are key indicators for potential surgical intervention.

A team-based approach involving neonatologists, radiologists, and pediatric surgeons is emphasized to ensure timely decision-making.

DOI: 10.4274/forbes.galenos.2025.78309
Forbes J Med 2025;6(2):180-7

Original Article / Orijinal Araştırma

Comparison of Prognostic Risk Models (IMDC, MSKCC, CFF) in Patients Diagnosed with Metastatic Renal Cell Cancer

Metastatik Renal Hücreli Kanser Tanılı Hastalarda Prognostik Risk Modellerinin (IMDC, MSKCC, CCF) Karşılaştırılması

Merve KESKİNKILIÇ, Kübra CANASLAN, Hüseyin Salih SEMİZ, Tuğba YAVUZŞEN

Dokuz Eylül University Faculty of Medicine, Department of Internal Medicine, Division of Medical Oncology, İzmir, Türkiye

Cite as: Keskinkılıç M, Canaslan K, Semiz HS, Yavuzşen T. Comparison of prognostic risk models (IMDC, MSKCC, CFF) in patients diagnosed with metastatic renal cell cancer. Forbes Med J.2025;6(2):180-7

ABSTRACT

Objective: The aim of this study is to reveal the relationship and correlation between the International metastatic renal cell carcinoma (mRCC) Database Consortium (IMDC) and Memorial Sloan Kettering Cancer Center (MSKCC) risk models used to determine first-line treatment in metastatic mRCC and, less commonly, the Cleveland Clinic Foundation (CCF) prognostic risk model.

Methods: The IMDC, MSKCC and CCF scores of mRCC patients who received immunotherapy (IO) and molecular targeted therapy were calculated retrospectively at the time of diagnosis.According to the score results, the patients were grouped as favorable,intermediate and poor risk.According to these risk groups,the median progression-free survival (mPFS) and median overall survival (mOS) of the patients were calculated and the correlation with each other was considered significant using appropriate statistical analyses, and $p<0.05$ was considered significant.

Results: The median follow-up time of 189 patients in the study was 45.5 months, mPFS 23.6 months [95% confidence interval (CI): 18.6-28.5 months] and mOS 34.6 months (95% CI: 23.3-45.9 months).The distribution of patients according to risk groups was similar in all three prognostic risk models. In the poor-risk group, both mPFS and mOS were statistically significantly shorter according to all three risk models (mPFS, IMDC: 14.2 months, MSKCC: 15.6 months, CCF: 17.1 months; mOS, IMDC: 17.6 months, MSKCC: 17.7 months, CCF: 22.4 months, $p<0.001$). A statistically significant positive correlation was observed between CCF, MSKCC and IMDC ($r=0.656$ vs. $r=0.690$, $p<0.001$). A stronger and statistically significant positive correlation was observed between MSKCC and IMDC ($r=793$, $p<0.001$).

Conclusion: Our study is the first study in the literature that we know of comparing the IMDC, MSKCC and CCF risk models in mRCC receiving IO and targeted therapy and as a result of our study, it was shown that all three risk models were correlated with each other.

Keywords: Cleveland Clinic Foundation (CCF), International mRCC Database Consortium (IMDC), Memorial Sloan Kettering Cancer Center risk (MSKCC), prognostic model, metastatic renal cell carcinoma (RCC)

ÖZ

Amaç: Bu çalışmanın amacı, metastatik renal hücreli karsinomda (mRHK) birinci basamak tedaviyi belirlemek için kullanılan Uluslararası mRCC Veritabanı Konsorsiyumu (IMDC) ve Memorial Sloan Kettering Kanser Merkezi (MSKCC) risk modelleri ile daha az yaygın olarak Cleveland Clinic Foundation (CCF) prognostik risk modeli arasındaki ilişkiyi ve korelasyonu ortaya koymaktır.

Yöntem: İmmünoterapi ve moleküler hedefli tedavi gören mRCC hastalarının IMDC, MSKCC ve CCF skorları tanı anında retrospektif olarak hesaplandı. Skor sonuçlarına göre hastalar düşük, orta ve kötü riskli olarak gruplandırıldı. Bu risk gruplarına göre hastaların medyan progresyonsuz sağkalım (mPSK) ve medyan genel sağkalım (mGSK) değerleri hesaplandı ve uygun istatistiksel analizler kullanılarak birbirleriyle korelasyon anlamlı kabul edildi ve $p<0,05$ anlamlı kabul edildi.

Received/Geliş: 16.06.2025
Accepted/Kabul: 09.07.2025
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Corresponding Author/ Sorumlu Yazar:
Merve KESKİNKILIÇ, MD,
Dokuz Eylül University Faculty of Medicine, Department of Internal Medicine, Division of Medical Oncology, İzmir, Türkiye
mervekeskinkilic90@gmail.com
ORCID: 0000-0002-3342-3144

Bulgular: Çalışmaya katılan 189 hastanın medyan takip süresi 45,5 ay, mPSK 23,6 ay [95% güven aralığı (CI): 18,6-28,5 ay] ve mGSK 34,6 ay (95% CI: 23,3-45,9 ay) idi. Hastaların risk gruplarına göre dağılımı her üç prognostik risk modeline göre benzerdi. Düşük risk grubunda hem mPSK hem de mGSK her üç risk modeline göre istatistiksel olarak anlamlı şekilde daha kısaydı (mPSK, IMDC: 14,2 ay, MSKCC: 15,6 ay, CCF: 17,1 ay; mGSK, IMDC: 17,6 ay, MSKCC: 17,7 ay, CCF: 22,4 ay, p<0,001). CCF, MSKCC ve IMDC arasında istatistiksel olarak anlamlı pozitif bir korelasyon gözlendi (r=0,656 vs. r=0,690, p<0,001). MSKCC ve IMDC arasında daha güçlü ve istatistiksel olarak anlamlı pozitif bir korelasyon gözlemlendi (r=793, p<0,001).

Sonuç: Çalışmamız, immünoterapi ve hedefli tedavi alan mRHK'de IMDC, MSKCC ve CCF risk modellerini karşılaştıran literatürdeki bildiğimiz ilk çalışmadır ve çalışmamızın sonucunda, üç risk modelinin de birbirleriyle ilişkili olduğu gösterilmiştir.

Anahtar Kelimeler: Cleveland Clinic Foundation (CCF), International mRCC Database Consortium (IMDC), Memorial Sloan Kettering Cancer Center (MSKCC), prognostik model, metastatik renal hücreli kanser (RHK)

INTRODUCTION

Metastatic renal cell carcinoma (mRCC) presents significant challenges in clinical management due to its heterogeneous nature and variable patient outcomes. Accurate prognostic models are essential for guiding treatment decisions and optimizing patient care. Among the most widely recognized models are the Memorial Sloan Kettering Cancer Center (MSKCC) criteria and the International mRCC Database Consortium (IMDC) criteria. These models stratify patients based on clinical and laboratory parameters, facilitating personalized therapeutic approaches.

The MSKCC model, established in the cytokine therapy era, identifies five adverse prognostic factors: low Karnofsky performance status (PS) (<80%), elevated lactate dehydrogenase (>1.5 times the upper limit of normal), anemia, hypercalcemia, and a diagnosis-to-treatment interval of less than one year. Patients are categorized into favorable (0 factors), intermediate (1-2 factors), and poor (≥3 factors) risk groups, with corresponding median overall survival (mOS) times of 20, 10, and 4 months, respectively.[1]

Recognizing the advancements in targeted therapies, the IMDC model was developed to provide prognostic insights in the context of modern treatments. This model incorporates six factors: anemia, hypercalcemia, neutrophilia, thrombocytosis, poor PS, and a diagnosis-to-treatment interval of less than one year. Similar to the MSKCC criteria, the IMDC stratifies patients into favorable (0 factors), intermediate (1-2 factors), and poor (≥3 factors) risk categories. Studies have demonstrated that the IMDC model effectively predicts outcomes in patients receiving targeted therapies, with mOS times of 35.3, 16.6, and 5.4 months for favorable, intermediate, and poor risk groups, respectively.[2]

In addition to the most commonly used MSKCC and IMDC criteria, the Cleveland Clinic Foundation (CCF) criteria are also used as a prognostic risk model, and this model uses elevated lactate dehydrogenase (>1.5 times the upper limit of normal), anemia, hypercalcemia, a diagnosis-to-treatment interval of less than one year, prior radiotherapy, and presence of hepatic, lung, or retroperitoneal lymph node metastases as risk factors.[3] The CCF prognostic risk model stratifies patients into favorable (0-1 factors), intermediate (2 factors), and poor (>3 factors) risk categories.[4] Using these criteria, 353 patients were retrospectively evaluated. Thirty-seven percent were at favorable risk, 35% at intermediate risk, and 28% at poor risk. mOS for these groups was 26.0, 14.4, and 7.3 months, respectively.[4] This study has also shown that the CCF prognostic risk model criteria contribute to prognostic modeling in mRCC.[3]

In studies comparing the correlation between the widely used MSKCC and IMDC prognostic risk model criteria in the literature, it was shown that these two prognostic models were statistically significantly correlated. In one study, 19% of the MSKCC intermediate group were reclassified as belonging to the IMDC poor risk group.[5-7] In the external validation study of the IMDC prognostic risk model, the CCF model, the International Kidney Cancer Working Group (IKCWG) model, the french model, and the MSKCC prognostic risk model were compared, and the MSKCC and IMDC prognostic risk models showed the highest concordance.[8] Although there are studies in the literature comparing the IMDC and MSKCC prognostic risk models one-on-one, and the CCF prognostic risk model indirectly for validation purposes, there is no study comparing the CCF prognostic risk model, which includes the metastasis site and the treatment received, with the IMDC and MSKCC prognostic risk models. In this study, we aimed to evaluate the prognostic value and correlation of IMDC, MSKCC, and CCF prognostic risk models in terms of selecting appropriate treatment strategies and counseling patients about prognosis in patients with mRCC receiving treatment with an anti-vascular endothelial growth factor (VEGF) inhibitor and immunotherapy (IO) treatment.

METHODS

Study Design and Population

In this retrospective study, patients diagnosed with mRCC between January 2010 and September 2023 at the Dokuz Eylül University Faculty of Medicine, Department of Medical Oncology were evaluated. The inclusion criteria of the study were determined as follows: i) being diagnosed with RCC, ii) receiving treatment for at least 3 months, iii) having complete blood count and serum biochemical

values at the beginning of treatment, iv) having complete data, v) being male and female aged 18 years and over. The exclusion criteria of the study were determined as follows: (i) patients with non-RCC histology; (ii) patients with less than 3 months follow-up; (iii) clinical trial patients; (iv) patients with a second malignancy were excluded. At the time of diagnosis, the demographic characteristics, clinicopathological characteristics, complete blood count, and biochemical laboratory values of the patients were recorded through the hospital database retrospectively.

Ethical Approval

This retrospective study was designed in accordance with the principles of the Declaration of Helsinki. Use of participant data was permitted without obtaining informed consent with the permission of the hospital administration. Based on this, the study was approved by the Dokuz Eylül University Faculty of Medicine Non-Interventional Research Ethics Committee (decision number: 2022/42-12, date: 28.12.2022).

Cleveland Clinic Foundation, International Metastatic Renal Cell Carcinoma Database Consortium and Memorial Sloan Kettering Cancer Center Risk Models

CCF, IMDC, and MSKCC prognostic risk models were used to determine the risk group of the patients. The parameters specified in Table 1 were used in these risk models, and the study population was grouped as favorable (no risk factors), intermediate (1-2 risk factors) and poor risk (>2 risk factors) according to the presence of the parameters in these models.

Response and Toxicity Assessment

Positron emission tomography/computed tomography (CT) and CT were performed every three to four months during the treatment. T Tumor staging was performed according to "Eighth Edition of American Joint Committee on Cancer and the Union for International Cancer Control TNM stage classification." Response assessments. It was conducted in accordance with the "Response Evaluation Criteria in Solid Tumors v1.1 guidelines". Toxicity assessments were made according to the National Cancer Institute Common Toxicity Criteria.

Statistical Analysis

Demographic and clinicopathological features of the patients were obtained from the hospital database. The Statistical Package for the Social Sciences version 24.0 (SPSS 24.0, IBM Corporation, Armonk, New York, USA) was used in the analysis of variables. The Kolmogorov-Smirnov Test was used to evaluate the conformity of the data to normal distribution. Independent Samples t-test and the Mann-Whitney U test, were used to compare two independent groups according to quantitative data. The Pearson chi-square test and Fisher's exact test were used to compare categorical variables. Multiple linear regression analyses were used to evaluate the relationship between the CCF, IMDC, and MSKCC prognostic risk models. Median progression-free survival (PFS) in the entire population was defined as the time from the start date of systemic therapy to progression, death, or last follow-up, whichever occurred first. OS was defined as the time from the date of diagnosis to death or last follow-up. Median follow-up time in the study was calculated using the inverse Kaplan-Meier. The Kaplan-Meier (product limit method) and Log Rank (Mantel-Cox) analyses were used to examine the effects of variables on survival times, according to the determined cut-off value. Quantitative variables are expressed in the tables as mean ± standard deviation and median (minimum/maximum), while categorical variables are shown as n (%). Variables were evaluated at a 95%

Table 1. CCF, IMDC and MSKCC prognostic risk models

Risk factor	CCF risk factors	IMDC risk factors	MSKCC risk factors
Time from diagnosis to systemic treatment <1 year	✓	✓	✓
Hemoglobin < lower limit of normal	✓	✓	✓
Calcium level > upper limit of normal	✓	✓	✓
Performance status <80% (Karnofsky)		✓	✓
LDH > 1.5x upper limit of normal	✓	✓	
Neutrophils > upper limit of normal			✓
Platelets > upper limit of normal			✓
Prior radiotherapy	✓		
Presence of hepatic, lung or retroperitoneal node metastases	✓		

CCF: Cleveland Clinic Foundation, IMDC: International Metastatic Renal Cell Carcinoma Database Consortium, MSKCC: Memorial Sloan Kettering Cancer Center, LDH: Laktat dehidrogenaz

confidence interval (CI) and statistical significance was determined as $p<0.05$.

RESULTS

Patients and Clinicopathological Characteristics

The median age of the 189 patients included in the study population was 62.4 years (range: 27.4-89.9). The male population in the study was approximately 3 times as many as the female population [male/female: 75.1% (n=142)/24.9% (n=47)]. The majority of the patients [62.4% (n=118)] had Eastern Cooperative Oncology Group PS 0, and the majority of the population (68.8%) had comorbidities. The distributions of tumor localization were quite similar, although the localization in the right kidney was slightly higher [right kidney/left kidney: 51.9% (n=98)/48.1% (n=91)]. It was observed that 69.8% (n=132) of the patients were de novo metastatic at the time of diagnosis, and the most common metastatic site was the lung [64.6% (n=122)]. The most common histological subtype in the population was clear cell RCC [66.1% (n=125)]. Clinicopathological features of the population are shown in Table 2.

Patients Treatment Characteristics

In the study population, the most frequently preferred treatment for metastatic cancer first-line was sunitinib (34.4%), followed by nivolumab (28%), axitinib (19.6%), and cabozantinib (2.6%). 79.9% of patients (n=151) who received treatment had progressed after first-line treatment. In addition, 70.9% of patients were able to reach second-line treatment, 37.6% were able to reach third-line treatment, and 6.9% were able to reach fourth-or more-line treatment.

Distribution of Patients According to Prognostic Models

When we evaluated the distribution of the patient population according to the IMDC prognostic risk model, 9.5% (n=18) of the patients were in the favorable-risk group, 62.4% (n=118) were in the intermediate-risk group, and 28% (n=53) were in the poor-risk group. According to the MSKCC prognostic risk model, 7.7% (n=14) were in the favorable-risk group, 60.8% (n=115) were in the intermediate-risk group, and 31.7% (n=60) were in the poor-risk group. In contrast, in the CCF prognostic model, a distribution almost similar to the MSKCC prognostic model was observed, with 7.9% (n=15) in the favorable-risk group, 61.9% (n=117) in the intermediate-risk group, and 30.2% (n=57) in the poor-risk group. The distribution of the population is shown in Table 3.

Survival Analysis

The mean follow-up period in the study was 45.5 months, and 29.6% (n=56) of patients were still alive. Without stratification according to prognostic risk group and stage, mPFS was 23.6 months (95% CI: 18.6-28.5 months) and mOS was 34.6 months (95% CI: 23.3-45.9 months) in the entire group.

Survival Analyses According to Prognostic Risk Models

The IMDC prognostic risk model reliably discriminated three risk groups to predict survival: the mPFS mOS for the favorable, intermediate, and poor risk groups were 40.9,

Table 2. Clinicopathological features of the patients

Characteristic	(n) %
Gender	
Male	75.1% (n=142)
Female	24.9% (n=47)
ECOG performance score	
0	62.4% (n=118)
1	28.6% (n=54)
2	17 % (n=9)
Comorbidity	69.8 % (n=132)
Hypertansion	46% (n=87)
Tip 2 diabetes mellitus	23.8% (n=45)
Chronic renal disease	20% (n=38)
Coronary arter disease	18.5% (n=35)
Other	16.9% (n=32)
Smoking status	
Present	22.8% (n=43)
Absent	42.9% (n=81)
Unknown	34.4% (n=65)
Tumor location	
Right kidney	51.9% (n=98)
Left kidney	48.1% (n=91)
Stage at the time of diagnosis	
Stage 4	69.8% (n=132)
Stage 3	15.3% (n=29)
Stage 2	7.4% (n=14)
Stage 1	6.3% (n=12)
Unknown	1.1% (n=2)
Histologic subtype	
Clear cell	66.1% (n=125)
Unclassifiable	7.9% (n=15)
Sarcomatoid	7.9% (n =15)
Papillary type type 1	4.2% (n=8)
Chromophobe	4.2% (n=8)
Papillary type 2	1.1% (n=2)
Others	8.6% (n=10)
Site of metastasis	
Lung	64.6% (n=122)
Lymph node	44.4% (n=84)
Non-regional	13.2% (n=25)
Regional	31.7% (n=60)
Bone	24.3% (n=46)
Liver	10.1% (n=19)
Brain	

ECOG: Eastern Cooperative Oncology Group.

Table 3. Distribution of prognostic risk models			
Prognostic risk models	**Favorable risk (%/n)**	**Intermediate risk (%/n)**	**Poor risk (%/n)**
IMDC	9.5% (n=18)	62.4% (n=118)	28 % (n=53)
MSKCC	7.7% (n=14)	60.8% (n=115)	31.7% (n=60)
CCF	7.9% (n=15)	61.9% (n=117)	30.2% (n=57)

CCF: Cleveland Clinic Foundation, IMDC: International Metastatic Renal Cell Carcinoma Database Consortium, MSKCC: Memorial Sloan Kettering Cancer Center

29.7, and 14.2 months (p=0.001); 75, 47.7, and 17.6 months (p<0.001), respectively. The MSKCC prognostic risk model also reliably distinguished three risk groups, quite similar to the IMDC, with survival times of 40.6, 33.5, and 15.6 months (p=0.001), and 76.3, 47.7, and 17.7 months (p=0.001). Again, as in the IMDC and MSKCC prognostic risk models, the CCF prognostic risk model statistically and significantly separated survival in the three risk groups, although patients had shorter PFS and OS times than the other two prognostic models: 35.7, 25.8, and 17.1 months (p=0.011); 67.9, 40.8, and 22.4 months (p=0.016), respectively. Survival analyses are shown in Figure 1.

Comparison of Cleveland Clinic Foundation, International Metastatic Renal Cell Carcinoma Database Consortium and Memorial Sloan Kettering Cancer Center Risk Models

When correlation indices were calculated for each model, it was observed that the CCF prognostic risk model showed a statistically significant positive correlation with both the MSKCC prognostic risk model (r=0.656) and the IMDC prognostic risk model (r=0.690), with p<0.001 for both). A statistically significant stronger positive correlation is observed between the MSKCC prognostic risk model and the IMDC prognostic risk model, which are more frequently used in current practice, respectively (r=793, p<0.001).

DISCUSSION

To our knowledge, this study is the first to evaluate the comparison of the most commonly used IMDC and MSKCC models in determining the treatment strategy for mRCC, and the CCF prognostic risk model, which includes the metastasis site and the given treatment (radiotherapy) as prognostic risk criteria. This is in contrast to two other less commonly used risk models. In our study population diagnosed with metastatic clear cell RCC treated with tyrosine kinase inhibitor (TKI) and IO, these three prognostic risk models were similarly effective in determining good, intermediate, and poor risk groups and were related to each other in terms of prognostic value. Although there is no direct comparison of these IMDC, MSKCC and CCF prognostic risk models in the literature, Heng et al.[8] reported that the IMDC prognostic risk model was comparable to the CCF, IKCWG, french and MSKCC risk models in their study. This supports our study focusing on only three prognostic risk models.

When the history of the development of prognostic risk models in mRCC was evaluated, the MSKCC prognostic risk model, developed by Motzer et al.[9], combined five poor prognostic factors: not having undergone nephrectomy, a Karnofsky-PS score below at the beginning of treatment, anemia, elevated serum laktat-dehidrogenaz, and hypercalcemia (corrected Ca >10), during the period of systemic chemotherapy and cytokine therapy in mRCC. This model was validated and reported as the CCF prognostic risk model. After previous radiotherapy, the presence of liver, lung, and retroperitoneal nodal metastases were found to be independent prognostic factors.[4,8] With the introduction of VEGF molecular targeted therapies over time, the treatment paradigm in mRCC has changed, and new prognostic profiles are needed. The IMDC prognostic risk model, derived from a multicenter cohort, was constructed consisting of six independent predictive factors [Karnofsky PS <80%, time from diagnosis to treatment <1 year, anemia (hemoglobin concentration < lower limit of normal), hypercalcemia (corrected calcium concentration > upper limit of normal), neutrophilia (neutrophil count > upper limit of normal), and thrombocytosis (platelet count > upper limit of normal)].[8,10]

Currently, in the determination of risk groups in Phase 3 studies, combination therapies (TKIs/IO and IO) recommended by international guidelines as first-line treatment in both favorable and intermediate/poor risk groups at category 1 level in the treatment of mRCC have received Food and Drug Administration approval. The IMDC and MSKCC risk models constitute the two most commonly used risk models, with the IMDC risk model validated for cytokine and targeted therapies being more prevalent.[11-14] It has been observed that other prognostic risk models, including CCF, are not used for determining treatment decisions in clinical phase studies other than these two risk models.[15] In our study, in addition to the IMDC and MSKCC prognostic risk models, the CCF prognostic risk model was also evaluated, and it was shown that in the study population including mRCC patients receiving targeted therapy and IO treatment, PFS and OS were similar according to all three risk models in all risk groups.

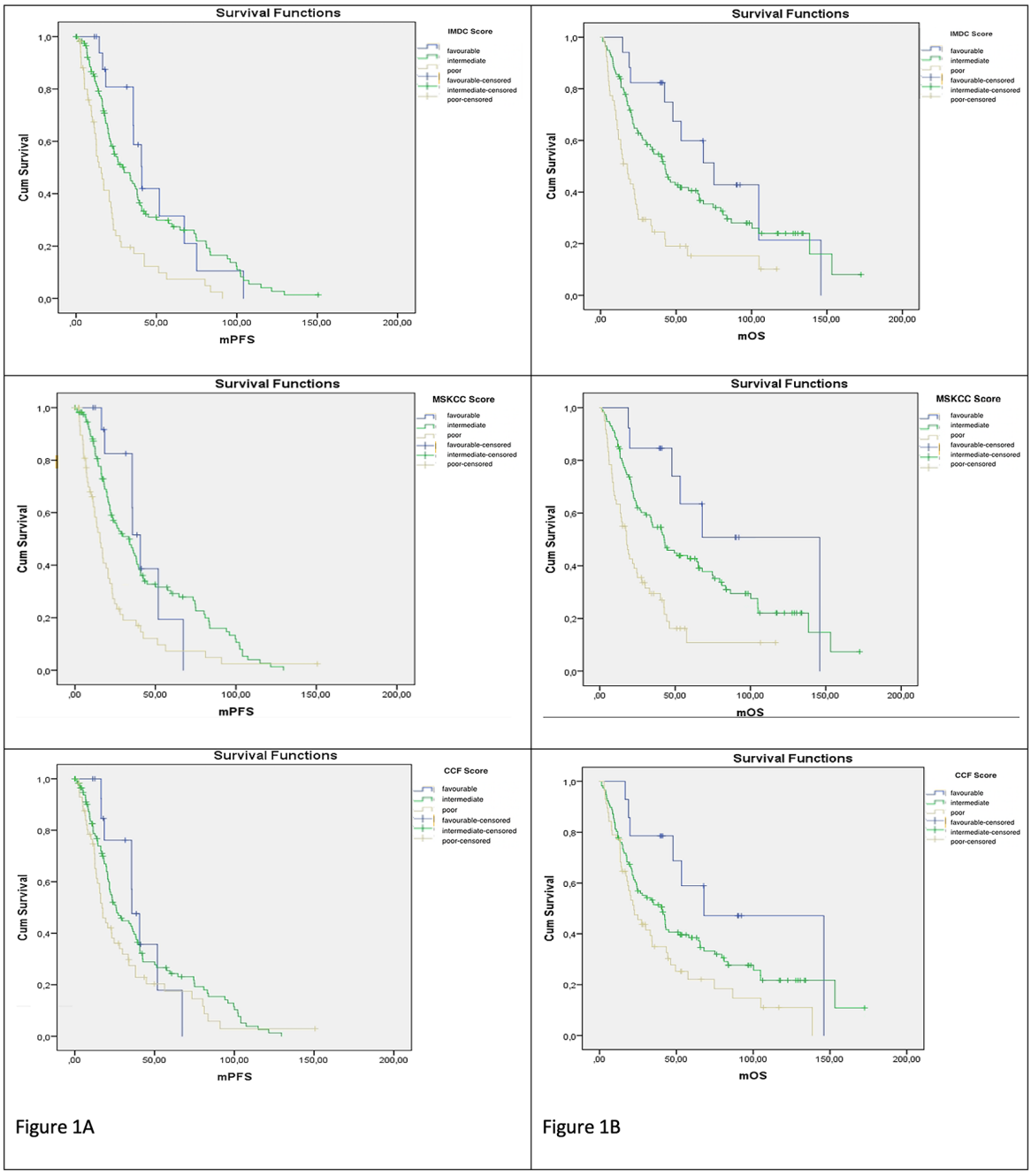


Figure 1. (A) PFS comparison of risk groups in IMDC, MSKCC and CCF risk models. (B) OS comparison of risk groups in IMDC, MSKCC and CCF risk models. PFS: Progression-Free Survival, IMDC: International Metastatic Renal Cell Carcinoma Database Consortium: MSKCC: Memorial Sloan Kettering Cancer Center, CCF: Cleveland Clinic Foundation, OS: Overall survival

The IMDC and MSKCC prognostic risk models were shown to be more highly correlated with each other, consistent with the literature.[16-19] Although there is no study in the literature comparing the CCF with the IMDC prognostic risk model and the MSKCC prognostic risk model, the results of the study by Heng et al.[8] comparing the IMDC with the CCF, IKCWG, the french, and the MSKCC prognostic risk models in mRCC patients reported that all risk models showed concordance with each other, as in our study.

In the studies in the literature where the distribution of risk groups was evaluated according to the IMDC and the MSKCC risk model, it was shown that patients in the intermediate risk group according to the MSKCC risk model were included in the poor risk group according

to the IMDC risk model at varying rates (19%, 31.2%).[7,16] In previous studies, although it was reported that the IMDC and MSKCC risk models showed concordance with each other in terms of PFS and OS, it was shown that, secondary to this difference in risk group distribution, the group that was intermediate according to the MSKCC risk model had shorter OS than the group that was intermediate according to the IMDC risk model. In our study, contrary to these studies, the distribution rate of the risk groups (favorable, intermediate, and poor group) was similar in all three prognostic risk models. As a result, no disagrrement was observed in terms of PFS and OS between both the IMDC and MSKCC risk models and between the risk groups according to all three prognostic risk models.

Study Limitations

Our study had some limitations. The first two of these limitations are that the study results belong to a single center and the study design is retrospective; therefore, we believe that the study should be planned and repeated as a prospective, large, multi-center study. Another limitation of the study is that since combination therapies (IO/TKI and IO) have become the standard treatment in mRCC today, this study involving patients receiving IO and VEGF-targeted therapy, does not reflect mRCC patients receiving combination therapy. It would be more appropriate to plan a study including this patient group. The last limitation of the study is that the CCF prognostic risk model was compared and correlated only with the most commonly used IMDC and MSKCC prognostic risk models. Since it was not compared with the french and IKCWG prognostic risk models, it is suggested that examining its correlation with these models may increase the power of the CCF prognostic risk model.

CONCLUSION

In our study, it was shown that the CCF prognostic risk model, which uses the given treatment and metastasis site as risk factors, is correlated with the most comprehensive and widely used MSKCC and IMDC prognostic risk models in clinical practice. It can be used as a prognostic risk model in mRCC in addition to other risk models, especially in the patient group where the metastasis site may be important in treatment decisions. These risk models are important for designing clinical trials, selecting appropriate treatment strategies, and counseling patients about their prognosis. Ongoing research continues to improve these models, combining new biomarkers and molecular insights to increase their predictive accuracy in the evolving environment of mRCC, where treatment strategies change over time.

Ethics

Ethics Committee Approval: This study was approved by the Dokuz Eylül University Faculty of Medicine Non-Interventional Research Ethics Committee (decision number: 2022/42-12, date: 28.12.2022).

Informed Consent: Retrospective study.

Presented in: The study was previously presented as an oral presentation at the 12th Symposium on Searches in Oncology held in Çeşme/İzmir on 13-15 December 2024.

Footnotes

Authorship Contributions

Surgical and Medical Practices: M.K., K.C., H.S.S., T.Y., Concept: M.K., K.C., H.S.S., T.Y., Design: M.K., Data Collection or Processing: M.K., K.C., Analysis or Interpretation: M.K., Literature Search: M.K., K.C., H.S.S., Writing: M.K.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Fitzgerald KN, Lee CH. Personalizing first-line management of metastatic renal cell carcinoma: leveraging current and novel therapeutic options. J Natl Compr Canc Netw. 2022;20.
2. Yang J, Wang K, Yang Z. Treatment strategies for clear cell renal cell carcinoma: past, present and future. Front Oncol. 2023;13:1133832.
3. Choueiri TK, Rini B, Garcia JA, et al. Prognostic factors associated with long-term survival in previously untreated metastatic renal cell carcinoma. Ann Oncol. 2007;18:249-55.
4. Mekhail TM, Abou-Jawde RM, Boumerhi G, et al. Validation and extension of the Memorial Sloan-Kettering prognostic factors model for survival in patients with previously untreated metastatic renal cell carcinoma. J Clin Oncol. 2005;23:832-41.
5. Ko JJ, Xie W, Kroeger N, et al. The International Metastatic Renal Cell Carcinoma Database Consortium model as a prognostic tool in patients with metastatic renal cell carcinoma previously treated with first-line targeted therapy: a population-based study. Lancet Oncol. 2015;16:293-300.
6. Kim JK, Kim SH, Song MK, et al. Application of the International Metastatic Renal Cell Carcinoma Database Consortium and Memorial Sloan Kettering Cancer Center risk models in patients with metastatic non-clear cell renal cell carcinoma: a multi-institutional retrospective study using the Korean Metastatic Renal Cell Carcinoma Registry. Cancer Res Treat. 2019;51:758-68.
7. Okita K, Hatakeyama S, Tanaka T, et al. Impact of disagreement between two risk group models on prognosis in patients with metastatic renal-cell carcinoma. Clin Genitourin Cancer. 2019;17:e440-6.
8. Heng DY, Xie W, Regan MM, et al. External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study. Lancet Oncol. 2013;14:141-8.

9. Motzer RJ, Mazumdar M, Bacik J, Berg W, Amsterdam A, Ferrara J. Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. J Clin Oncol. 1999;17:2530-40.
10. Heng DY, Xie W, Regan MM, et al. Prognostic factors for overall survival in patients with metastatic renal cell carcinoma treated with vascular endothelial growth factor-targeted agents: results from a large, multicenter study. J Clin Oncol. 2009;27:5794-9.
11. Tannir NM, Albigès L, McDermott DF, et al. Nivolumab plus ipilimumab versus sunitinib for first-line treatment of advanced renal cell carcinoma: extended 8-year follow-up results of efficacy and safety from the phase III CheckMate 214 trial. Ann Oncol. 2024;35:1026-38.
12. Plimack ER, Powles T, Stus V, et al. Pembrolizumab plus axitinib versus sunitinib as first-line treatment of advanced renal cell carcinoma: 43-month follow-up of the phase 3 KEYNOTE-426 study. Eur Urol. 2023;84:449-54. Epub 2023;25. Erratum in: Eur Urol. 2023;84:e123-4. Erratum in: Eur Urol. 2024;85:e58-9. Available from: 10.1016/j.eururo.2023.11.016.
13. Choueiri TK, Powles T, Burotto M, et al. Nivolumab plus cabozantinib versus sunitinib for advanced renal-cell carcinoma. N Engl J Med. 2021;384:829-41.
14. Motzer R, Alekseev B, Rha SY, et al. Lenvatinib plus pembrolizumab or everolimus for advanced renal cell carcinoma. N Engl J Med. 2021;384:1289-1300.
15. Shinohara N, Abe T. Prognostic factors and risk classifications for patients with metastatic renal cell carcinoma. Int J Urol. 2015;22:888-97.
16. Kubackova K, Melichar B, Bortlicek Z, et al. Comparison of two prognostic models in patients with metastatic renal cancer treated with sunitinib: a retrospective, registry-based study. Target Oncol. 2015;10:557-63.
17. Noe A, de Bruijn RE, Blank C, Horenblas S, Haanen J, Bex A. Comparison of pre-treatment MSKCC and IMDC prognostic risk models in patients with synchronous metastatic renal cell carcinoma treated in the era of targeted therapy. World J Urol. 2016;34:1067-72.
18. Sella A, Michaelson MD, Matczak E, Simantov R, Lin X, Figlin RA. Heterogeneity of patients with intermediate-prognosis metastatic renal cell carcinoma treated with sunitinib. Clin Genitourin Cancer. 2017;15:291-299.e1.
19. Tanaka N, Mizuno R, Ito K, et al. External validation of the MSKCC and IMDC risk models in patients treated with targeted therapy as a first-line and subsequent second-line treatment: a Japanese multi-institutional study. Eur Urol Focus. 2016;2:303-9.

DOI: 10.4274/forbes.galenos.2025.52385
Forbes J Med 2025;6(2):188-95

Original Article / Orijinal Araştırma

Retrospective Evaluation of Synovial Chondromatosis with Histopathological and Clinical Features

Sinovyal Kondromatozisin Histopatolojik ve Klinik Özellikleriyle Retrospektif Değerlendirilmesi

Özden ÖZ[1], Murat SEZAK[2], Başak DOĞANAVŞARGİL[2], İpek TAMSEL[3], Hüseyin KAYA[4], Elcil KAYA BİÇER[4], Emin BÜYÜKTALANCI[5], Gülçin BAŞDEMİR[6]

[1]İzmir Democracy University Buca Seyfi Demirsoy Training and Research Hospital, Department of Pathology, İzmir, Türkiye
[2]Ege University Faculty of Medicine, Department of Medical Pathology, İzmir, Türkiye
[3]Ege University Faculty of Medicine, Department of Radiodiagnostics, İzmir, Türkiye
[4]Ege University Faculty of Medicine, Department of Orthopedics and Traumatology, İzmir, Türkiye
[5]University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, Clinic of Medical Pathology, İstanbul, Türkiye
[6]İstanbul Galata University, Health Sciences Basic Field Medical Pathology, İstanbul, Türkiye

Cite as: Öz Ö, Sezak M, Doğanavşargil B, et al. Retrospective evaluation of synovial chondromatosis with histopathological and clinical features. Forbes J Med. 2025;6(2):188-95

ABSTRACT

Objective: Synovial chondromatosis (SC) is a rare, benign condition characterized by the formation of nodular cartilage within the synovium of joints, bursae, or tendon sheaths. The condition typically presents as loose body-like nodules, which often exhibit calcification and ossification. SC primarily affects adults and most commonly involves large joints. Histologically, it is defined by synovium-lined cartilaginous nodules displaying varying degrees of peripheral cellularity. Due to overlapping features, SC may be misdiagnosed as a chondroid malignancy, particularly in limited biopsy specimens.

Methods: We performed a retrospective review of cases diagnosed as SC in the Department of Pathology at Ege University Faculty of Medicine between 1986 and 2020. Archived materials were analyzed for epidemiological characteristics, clinical presentation, macroscopic and microscopic histopathological features, and imaging findings.

Results: A total of 114 cases were identified, with a nearly equal gender distribution (male-to-female ratio: 0.95). The median age was 54±16.7 years (range: 18-86 years). The knee was the most affected site (56.14%, n=64), followed by the ankle (14.03%), hip (12.28%), wrist (9.64%), elbow (5.26%), and shoulder (2.63%). Histopathological reevaluation was performed on 62 cases. No correlation was found between histological features and patient age, gender, lesion location, or recurrence.

Conclusion: Most cases involved the knee joint, although rarer sites such as the ankle were also observed. Notably, synovial lining over the nodules was not consistently present. Many nodules exhibited increased cellularity, nuclear pleomorphism, and hyperchromatism-features that can mimic low-grade chondrosarcoma. Histological and cytological features did not correlate with recurrence in our cohort.

Keywords: Joint diseases, synovium, chondromatosis, osteochondromatosis

ÖZ

Amaç: Sinovyal kondromatozis (SK), eklemlerin, bursaların veya tendon kılıflarının sinovyumunda nodüler kıkırdak oluşumuyla karakterize nadir, iyi huylu bir eklem hastalığıdır. Tipik olarak, sıklıkla kalsifikasyon ve ossifikasyon gösteren eklem içerisinde serbest nodüller olarak ortaya çıkar. SK öncelikle yetişkinleri etkiler ve çoğunlukla büyük eklemleri tutar. Histolojik olarak, çeşitli derecelerde özellikle periferik hücresellik gösteren sinovyumla kaplı kıkırdak nodülleri olarak tanımlanır. Çakışan özellikleri nedeniyle, SK özellikle sınırlı biyopsi örneklerinde kıkırdak malignitesi olarak yanlış teşhis edilebilir.

Received/Geliş: 02.07.2025
Accepted/Kabul: 01.08.2025
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Corresponding Author/ Sorumlu Yazar:
Özden ÖZ, MD,
İzmir Democracy University Buca Seyfi Demirsoy Training and Research Hospital, Department of Pathology, İzmir, Türkiye
ozdenozpat@gmail.com
ORCID: 0000-0001-5601-1567

Yöntem: 1986 ile 2020 yılları arasında Ege Üniversitesi Tıp Fakültesi, Patoloji Bölümü'nde SK tanısı konulan olguların retrospektif bir incelemesini gerçekleştirdik. Arşivlenen materyaller epidemiyolojik, klinik, makroskobik, mikroskobik, histopatolojik özellikleri ve görüntüleme bulguları açısından analiz edildi.

Bulgular: Toplam 114 olgu tespit edildi ve cinsiyet dağılımı neredeyse eşitti (erkek-kadın oranı: 0,95). Ortanca yaş 54±16,7 (aralığı: 18-86) idi. Diz en çok etkilenen bölgeydi (%56,14, n=64), ardından ayak bileği (%14,03), kalça (%12,28), bilek (%9,64), dirsek (%5,26) ve omuz (%2,63) geldi. Altmış iki olguda histopatolojik yeniden değerlendirme yapıldı. Histolojik özellikler ile hasta yaşı, cinsiyeti, lezyon yeri veya tekrarlama arasında bir korelasyon bulunamadı.

Sonuç: Olguların çoğu diz eklemini içeriyordu, ancak ayak bileği gibi daha nadir bölgeler de gözlendi. Özellikle nodüllerin üzerindeki sinovyal doku daima mevcut değildi. Birçok nodülde artmış hücresellik, nükleer pleomorfizm ve hiperkromazi görüldü; bunlar düşük dereceli kondrosarkomu taklit edebilen özelliklerdi. Histolojik ve sitolojik özellikler kohortumuzda hastalığın tekrarlaması ile ilişkili bulunmadı.

Anahtar Kelimeler: Eklem hastalıkları, sinovyum, kondromatozis, osteokondromatozis

INTRODUCTION

Synovial chondromatosis (SC) is a rare, benign yet locally aggressive neoplastic joint disorder characterized by the formation of multiple lobulated hyaline cartilaginous or osteocartilaginous nodules (chondromas) within the synovial membrane of large joints. Typically originating from synovial connective tissue in adults, SC most often presents intra-articularly, although extra-articular manifestations have been rarely reported. The condition is classified as primary SC when idiopathic, and as secondary SC when associated with pre-existing joint pathology.[1,2] Due to its clinical similarity to other joint disorders, SC is frequently challenging to diagnose accurately.[2-4] Although its etiology remains largely unknown, SC is widely believed to result from metaplasia of synovial connective tissue.[3,5] It most commonly affects individuals in the third to fifth decades of life,[2,3,5,6] with a higher prevalence in males, and predominantly involves large joints, particularly the knee.[2,7]

Clinically, SC poses a diagnostic challenge for orthopedic specialists, as its presentation often mimics other joint pathologies. Patients may experience a prolonged symptomatic period-sometimes extending up to five years-before a definitive diagnosis is established.[4] Common clinical features include joint effusion, tenderness, pain, swelling, crepitus, limited range of motion, and occasionally palpable nodules, resembling osteoarthritis.[2-4,7] Symptoms such as pain at rest, pain that worsens with movement, and joint deformity due to synovial hypertrophy are notable. Even when SC is a primary neoplastic or metaplastic lesion, it may present with secondary osteoarthritic changes resulting from mass effect. In both primary and secondary forms, long-term complications frequently include cartilage erosion and progressive joint degeneration.

Malignant transformation to synovial chondrosarcoma is exceedingly rare and is primarily reported in patients with long-standing disease and multiple recurrences.[5,6,8-11] The estimated incidence of malignant transformation in primary SC ranges from 1-5%,[4,6,9,10,12] although this remains uncertain. The highest reported rate was observed in a large orthopedic oncology database, where 5 out of 78 patients with primary SC developed chondrosarcoma.[4]

Recent molecular investigations have identified recurrent gene rearrangements involving fibronectin 1 (FN1) and activin receptor 2A (ACVR2A) as potential drivers of SC pathogenesis.[5,13] In one cohort, the FN1-ACVR2A gene fusion-detected via fluorescence *in situ* hybridization-was implicated in neoplastic transformation.[5] Mutations in IDH1/2, which are commonly observed in malignant chondroid tumors, have not been identified in SC to date.[14]

Definitive diagnosis of SC requires histopathological examination of biopsied or surgically excised tissue, demonstrating cartilaginous nodule formation within the synovium or as free-floating intra-articular bodies.

The objective of this study is to analyze a large surgical pathology database of SC cases to describe the clinical and histopathological characteristics of primary SC.

METHODS

Patient Selection of Synovial Chondromatosis Cases

Archival records of patients diagnosed with SC between 1986 and 2020 were obtained from the Department of Pathology at Ege University Faculty of Medicine. The study included only cases involving large joints-specifically the knee, hip, elbow, shoulder, and ankle. Clinical, demographic, histopathological, treatment, and imaging data available in the reporting system were re-evaluated. Of the 124 cases identified from the records, two cases diagnosed under the age of 18 and eight cases with insufficient clinical, demographic, and histopathological data were excluded from the study. The study was therefore conducted with 114 cases.

Histopathological Examination of Synovial Chondromatosis Cases

Since only slides or paraffin blocks of cases after 2000 could be obtained from the archive, hematoxylin and eosin (H&E) stained slides from 62 cases diagnosed between 2000 and 2019 were retrospectively reviewed and confirmed based on current histopathological criteria. Nuclear pleomorphism (focal vs. diffuse),

cellularity (slight, medium, and high), and the presence of calcification and ossification were assessed. Whether these histological features were statistically associated with clinicopathological findings was investigated using the chi-square test. Cases diagnosed prior to 2000 were evaluated solely through the review of archival reports. The study was approved by the Ege University Medical Research Ethics Committee (decision number: 22-3.1T/61, date: 24.03.2022).

Statistical Analysis

Statistical analyses were conducted using IBM SPSS Statistics v21.0. The relationship between demographic features and mucosal pathological characteristics was evaluated with the chi-square test. Fisher's exact test was used for comparisons of categorical variables. A p value of less than 0.05 was considered statistically significant in all analyses.

RESULTS

A total of 114 cases were identified, with a nearly equal gender distribution (male-to-female ratio: 0.95). The median age was 46.96±15.79 years (range: 18-86 years). The most commonly affected joint was the knee (56.14%, n=64), followed by the ankle (14.03%, n=16), hip (12.28%, n=14), wrist (9.64%, n=11), elbow (5.26%, n=6), and shoulder (2.63%, n=3) (Table 1, Supplementary Figure 1). The distribution of cases by decade is also included in the same table.

Radiologically, several cases showed densely radiopaque bodies within both the synovium and joint space (Figure 1a, c), with nodules generally measuring less than 30 mm. Degenerative joint changes, including osteophyte formation, were frequently observed. Plain radiographs often revealed multiple (Figure 1a) and occasionally single (Figure 1c) intra-articular bodies, when calcified. In cases lacking calcification, intra-articular bodies appeared radiolucent and could be missed. Magnetic resonance imaging (MRI) and computed tomography (CT) proved useful for detecting non-calcified nodules. Degenerative changes on the joint surface were also noted (Figure 1c, white arrow). Arthroscopic images of the case in Figure 1a are shown in Figure 1b, and of the case in Figure 1c are shown in Figure 1d.

Gross examination revealed round or oval loose bodies resembling a joint mouse, ranging from 0.1 to 3 cm in diameter, which can sometimes form mass formations (Figure 2a, b). In patients who underwent synovectomy, these nodules were occasionally embedded within the synovium (Figure 3a, black arrow; Figure 3c; Figure 4a). Histologically, the nodules consisted of cartilage covered by synovium with varying degrees of cellularity (Figure 3a, b; Figure 4b) or consisted of degenerated free cartilage nodules (Figure 2c). Most nodules were composed of chondrocyte clusters embedded within a hyaline cartilage matrix and frequently exhibited central calcification and/or ossification. Notably, increased peripheral cellularity was commonly observed, which could potentially be misinterpreted as a chondroid malignancy on biopsy specimens (Figure 4b).

Among the 62 cases with available follow-up data from 2000 to 2019 (excluding 52 patients for whom clinical follow-up could not be obtained), 21 cases (33.87%) were initially diagnosed as tumoral masses, 29 (46.77%) as SC, and 12 (19.35%) as synovitis. Of these, 56 patients (90.33%) underwent total synovectomy, while 6 (9.67%) were treated by removal of joint mice alone. Recurrence occurred in 4 cases (6.45%), whereas 58 cases (93.56%) remained recurrence-free. No statistically significant correlation was found between histological or nuclear features and lesion localization, or between cartilage cellularity and variables such as age, gender, anatomical site, or recurrence (p>0.05, chi-squared test).

Table 1. Clinicopathologic findings in the synovial chondromatosis cases

Age	46.96+-15.79 (18-86)	
Gender	Women	42.98%, n=49
	Men	57.02%, n=65
Localization	Right	57.01%, n=65
	Left	42.99%, n=49
Site	Knee	56.14%, n=64
	Ankle	14.03%, n=16
	Hip	12.28%, n=14
	Wrist	9.64%, n=11
	Elbow	5.26%, n=6
	Shoulder	2.63%, n=3
The case distribution of year	1969-1979	14.91%, n=17
	1980-1989	14.03%, n=16
	1990-1999	16.66%, n=19
	2000-2009	26.31%, n=30
	2010-2020	28.07%, n=32
Treatment	Remove of joint mass	5.26%, n=6
	Synovial excision	49.13%, %, n=56
	NA	45.61%, %, n=52
Recurrence	Present	3.51%, n=4
	Absent	50.87%, n=58
	NA	45.61%, n=52

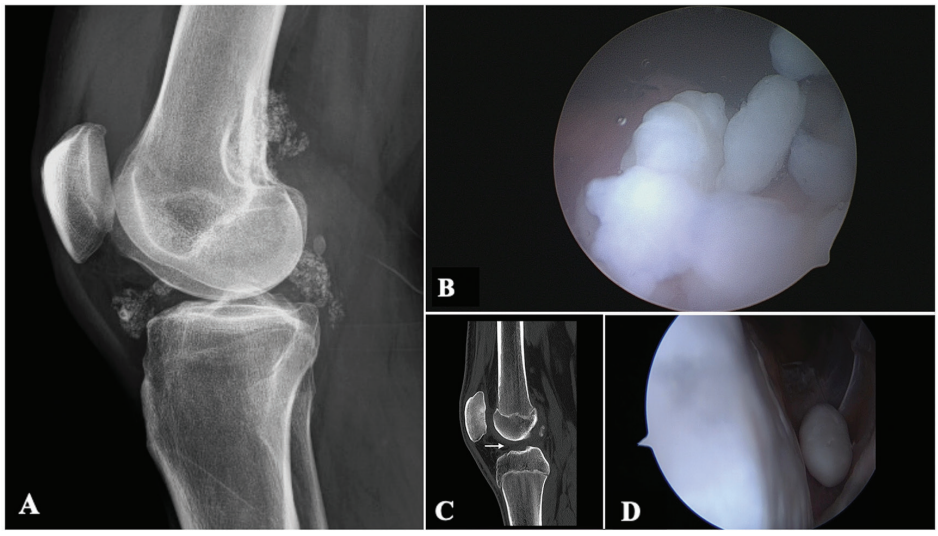


Figure 1. (a) Lateral knee radiograph of a 37-year-old male patient showing multiple well-defined, rounded opacities of similar size confined to the right knee joint space. No bony erosions were present. Concentric calcifications in the form of rings, arcs, and swirls were observed. (b) Arthroscopic view of a case with synovial chondromatosis. (c, d) Arthroscopic and radiological views of the same nodule. The degenerative change of the joint surface is shown in c (white arrow)

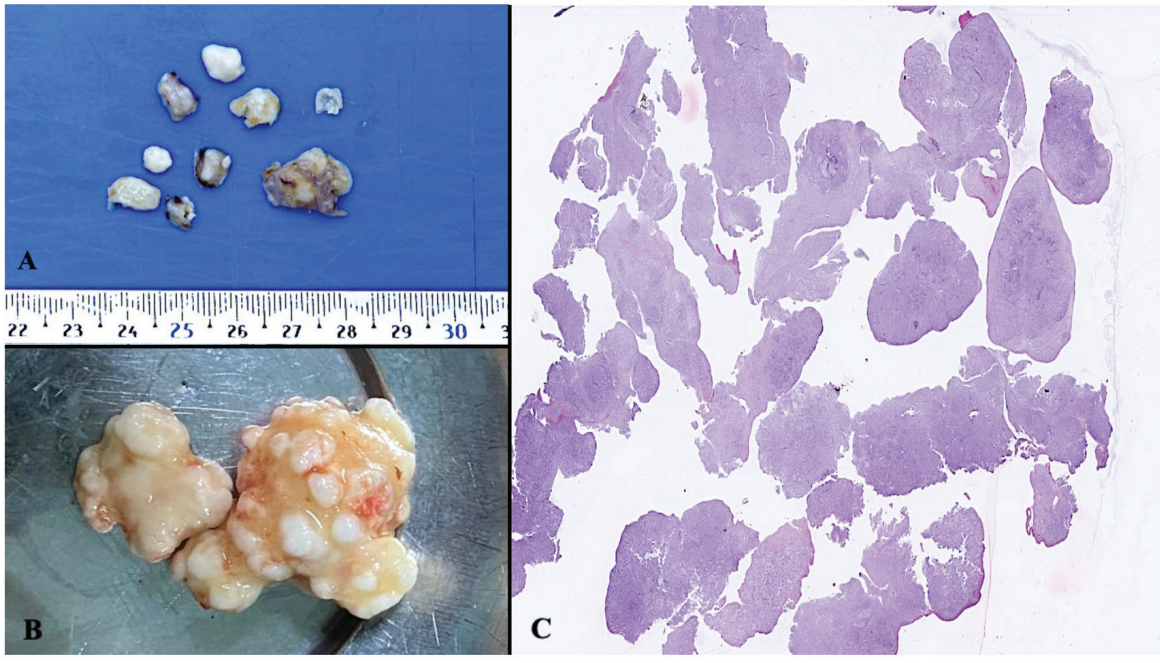


Figure 2. (a, b) Examples of free joint bodies from synovial chondromatosis. Histological confirmation is essential for differentiating these bodies. (c) Histologically confirmed synovial chondromatosis (H&E, ×2)

H&E: Hematoxylin and eosin

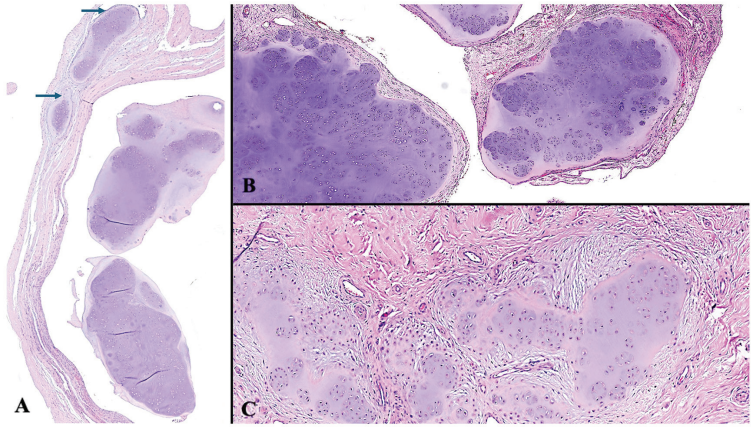


Figure 3. (a, b) Synovial chondromatosis with well-formed nodules in the joint space or just beneath the synovial lining (H&E, ×4 and ×10). (c) Deeper cartilaginous nodules embedded within the synovium (H&E, ×10)

H&E: Hematoxylin and eosin

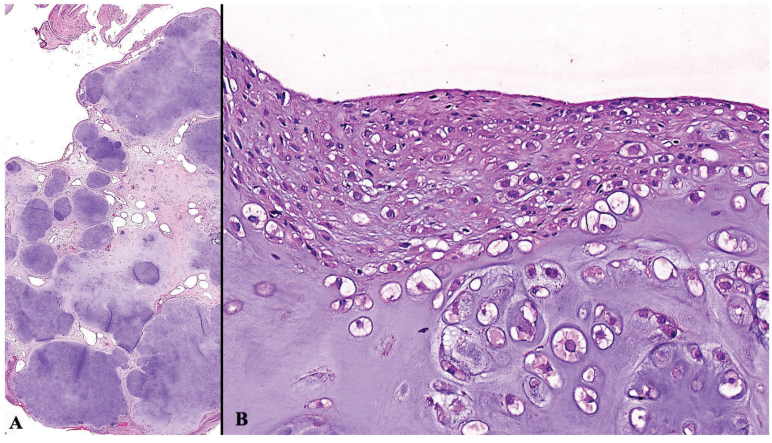


Figure 4. (a) Marked peripheral hypercellularity with lower cellularity in the central regions of the nodules (H&E, ×10). (b) Disorganized peripheral cells with increased cellularity, raising suspicion for chondrosarcoma (H&E, ×20)

H&E: Hematoxylin and eosin

DISCUSSION

SC is a benign neoplastic or metaplastic joint disorder characterized by the formation of multiple cartilaginous or osteocartilaginous nodules within the synovium, tendon sheath, bursa, or joint space. First described by Laennec in 1813 as intra-articular loose bodies originating from subsynovial tissues,[3,4,15] SC was later more thoroughly characterized by Fisher in 1920.[16]

SC predominantly affects large joints, with the knee involved in approximately 70% of cases; followed by the hip (20%), then the shoulder, elbow, ankle, and wrist.[2,4,9] SC is typically monoarticular and can affect any synovial joint. In this clinicopathological review, spanning 34 years, of 114 SC cases, the knee was the most commonly involved joint, consistent with previous literature, although the proportion in our series (56.14%) was slightly lower than previously reported.[2-4,9] Notably, we observed a higher-than-expected incidence in the ankle, diverging from existing literature.[17] While shoulder involvement has been infrequently reported,[10,18] ankle localization is even rarer,[17,19] and there is little to no documented evidence of primary or secondary SC in these locations.[1,18,20] No cases of extra-articular SC were identified in our cohort, contrary to earlier reports,[7] possibly due to incomplete radiological records in some patients.

Approximately 70% of SC nodules undergo endochondral ossification; the remainder, particularly those embedded in the synovium, remain unossified (Figure 3c). These nodules can detach into the joint, where they may undergo mineralization or persist in the synovium. Histologically, some nodules displayed densely packed and disorganized cellular architecture (Figure 3b, 4a). Although nuclear features were generally uniform-with fine chromatin and

small nucleoli-occasional atypical or binucleated nuclei were observed (Figure 4b). We found no association between histological features, including nuclear atypia, and clinical parameters such as age, sex, lesion location, or recurrence in our research.

Radiologically, SC is characterized by intra- or extra-articular cartilage nodules, which may become dislodged into the joint space and are often referred to as "joint mice" (Figure 1a-d). Plain radiographs are typically diagnostic, showing multifocal, rounded, mineralized nodules within or around the joint. However, in roughly 20% of cases, mineralization is not evident due to its time-dependent nature.[3] Non-mineralized nodules may mimic joint effusions, leading to misdiagnosis. In such cases, CT and MRI provide valuable information regarding disease extent and involvement of adjacent structures.[2] MRI findings depend on the degree of synovial proliferation and the presence of calcified or uncalcified nodules.

Milgram[20] proposed a widely accepted three-phase progression of SC: (1) active synovitis without loose bodies; (2) a transitional phase with active synovial proliferation and cartilaginous loose bodies; and (3) multiple ossified bodies with inactive synovitis. These loose bodies may cause mechanical erosion of articular surfaces, contributing to joint damage. Milgram's[20] staging is widely accepted today and supports the view that synovial activity varies throughout the disease course.

Treatment primarily involves the removal of loose bodies to relieve symptoms and prevent joint deterioration. Synovectomy is often employed more extensively than other procedures.[2,3] Treatment strategy and prognosis may vary depending on the affected joint.[2] If untreated, SC can lead to progressive joint and periarticular damage. Therefore, early surgical intervention-via arthroscopic or open removal of loose bodies, with or without synovectomy-is recommended. The necessity of synovectomy remains debated, as each approach carries distinct risks for recurrence and complications.[2,3] While local excision is often curative, recurrence is possible. Malignant transformation (i.e., secondary synovial chondrosarcoma) is considered rare and is usually detected through the recurrence of SC.[2-4,6,7,10] Conversely, it has been demonstrated that two lesions can be detected simultaneously in some patients.[10] Due to the small number of cases, the literature lacks an explanation for the underlying mechanism. Recurrence rates are estimated at 15-20%, especially in tenosynovial variants, which account for 15-20% of all cases.[7,11,20] In our cohort, the recurrence rate was lower than previously reported,[7] which we attribute to the routine practice of total synovectomy. No correlation was found between recurrence and cellularity, ossification, or other histological features. This supports the notion that complete synovial excision significantly reduces recurrence and may prevent malignant transformation.

Once dislodged into the joint, these nodules-"joint mice"-can absorb nutrients from synovial fluid and continue to grow, potentially causing secondary joint surface degeneration. In line with previous studies, nearly all cases in our series featured ossified or calcified nodules.

Histopathologically, SC is characterized by multiple cartilaginous nodules or loose bodies within the synovium. The nodules may be composed of hyaline cartilage with low (Figure 3a, c) or typically high cellular content, particularly at the periphery, and often exhibit cytological atypia (Figure 4b). Features such as multinucleated cells, myxoid matrix generation, nuclear crowding, and pleomorphism were observed. In some cases, the synovial lining over the nodules was not consistently present, and areas of increased cellularity were observed (Figure 3b). If the nuclear pleomorphism and hyperchromasia are evaluated without knowledge of lesion location or radiologic context, they may be diagnosed as low-grade chondrosarcomas.

Primary SC is thought to be more likely neoplastic in origin and typically occurs in the third or fourth decade of life, without identifiable underlying joint pathology (clinically, radiologically, or histopathologically).[3,4,21] In contrast, secondary SC is more common, usually appearing in the fifth or sixth decade and associated with conditions such as synovitis, trauma, osteoarthritis, or neuropathic arthropathy.[2,3,21] Case reports of SC in athletes are also available in the literature and support the repetitive trauma origin.[22] In suspected primary cases, a neoplastic mechanism may be more plausible, while secondary cases are generally attributed to synovial metaplasia due to existing joint pathology. Primary SC tends to feature more numerous and uniformly sized nodules, whereas secondary SC typically involves fewer, larger, and irregularly shaped bodies.[8,23] Some reports estimate that 70-95% of SC cases are primary, while others suggest a higher prevalence of secondary SC.[3] Distinguishing between the two remains a subject of ongoing debate and is often clinically challenging. Regardless of type, SC leads to the formation of multiple chondroid nodules and osteochondral loose bodies, and results in degeneration of the joint space.[2]

Villacin et al.[1] proposed differentiation between primary and secondary SC based on their histopathological characteristics. In our classification, which takes into account the distinction based on histopathological nuclear atypia, we found no association between nuclear atypia and clinical outcomes such as recurrence during the follow-up period. We evaluated the cases histopathologically with

archive record information, and because we could not access the radiological records of all cases pre-2000, it was not possible to determine whether there were underlying degenerative joint findings in our series. For this reason, our inability to distinguish primary from secondary in our cases represents a limitation of our series.

Furthermore, even in presumed primary SC, degenerative changes may arise over time, thereby diminishing the diagnostic value of radiological evidence alone. No significant correlation was observed between histological/cytological features and clinical parameters in our cohort. In conclusion, our findings suggest that the commonly proposed histological and clinical parameters for distinguishing primary from secondary SC-such as the presence of joint degeneration, metaplastic cartilage, cellular atypia, calcification patterns in nodules, ossification type 1, nodule size uniformity, and number or age-may lack diagnostic reliability.[6,8]

Other joint pathologies may also present with loose bodies and synovial proliferation, including crystal deposition diseases, osteochondral fractures, osteochondritis dissecans, neurotrophic arthritis, rheumatoid arthritis, degenerative arthritis, and tuberculous arthritis. Additionally, synovial soft tissue tumors-such as synovial hemangioma, arborizing lipoma, and diffuse-type tenosynovial giant cell tumor-should be considered in the differential diagnosis. While these entities may overlap clinically, they are generally distinguishable by histopathological examination. Although rare, malignant lesions such as intraosseous low-grade chondrosarcoma extending into the joint or synovial sarcoma involving adjacent bone, should also be included in the differential diagnosis. SC is typically non-invasive, with repeated recurrences being the primary concern regarding potential malignant transformation. H&E staining alone is insufficient to distinguish SC from malignancy; clinical history and imaging are essential for accurate diagnosis.[3]

The etiology of SC remains poorly understood and is believed to involve activation of the synovial membrane.[4,6]

It is still unclear whether SC represents a reactive metaplastic process or a true neoplastic transformation.[3,6] Although synovial metaplasia is the more widely accepted mechanism, recent genetic studies have provided evidence suggesting a neoplastic origin.[5,13] If primary SC is considered a neoplastic entity, the identification of genetic alterations, particularly involving the FN1 and ACVR2A genes,[5] in excised SC tissue may help distinguish it from secondary SC, which is currently viewed as a more metaplastic process. ACVR2A encodes a receptor for Activin A, BMP-4, and BMP-6, all of which are involved in skeletal development.[24] Testing for IDH1 and IDH2 mutations may further assist in differentiating SC from chondrosarcoma, a malignant process, either through immunohistochemistry or through identifying the mutated gene. In addition to histopathological and molecular assessments, radiologic evaluation plays a critical role in making this distinction. Together, these tools can guide more aggressive treatment strategies and follow-up protocols for cases that exhibit genetic alterations.

Study Limitations

The inability to apply molecular and genetic tests is a limitation of our study.

CONCLUSION

SC is a rare, benign condition that can lead to significant joint dysfunction. It is widely recognized to occur in both primary and secondary forms, most commonly affecting the knee. Differentiating between these forms remains challenging due to overlapping clinical features and limitations in radiological and histopathological criteria. Early diagnosis and appropriate surgical intervention are critical to preventing disease progression and joint damage.

Although treatment outcomes are relevant, our study did not yield clinical or histological predictors of recurrence. Further research is warranted to elucidate the pathogenesis of SC, particularly the molecular mechanisms underlying both primary and secondary forms. While primary SC may involve neoplastic processes, secondary SC is more often associated with mechanical joint pathology; however, these pathways may intersect. A comprehensive understanding of the clinical, pathological, molecular, and therapeutic dimensions of SC is essential for the effective management of this uncommon but impactful disease.

Ethics

Ethics Committee Approval: The study was approved by the Ege University Medical Research Ethics Committee (decision number: 22-3.1T/61, date: 24.03.2022).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

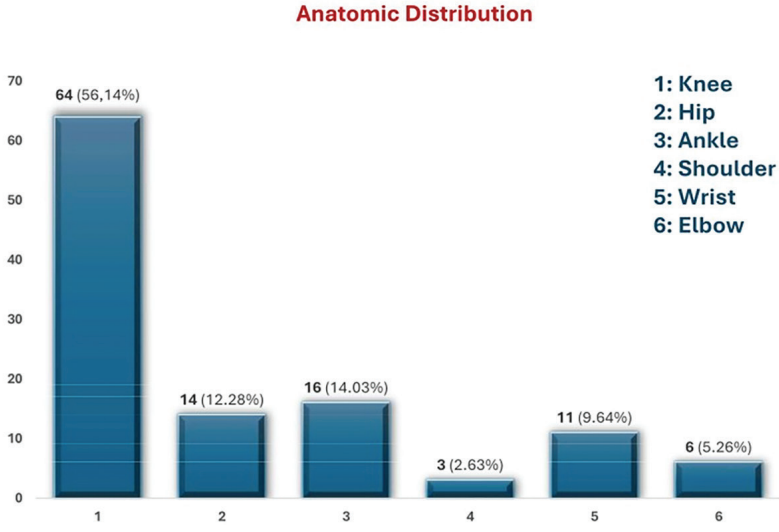
Surgical and Medical Practices: H.K., E.K.B., Concept: Ö.Ö., M.S., B.D., Design: Ö.Ö., B.D., Data Collection or Processing: Ö.Ö., İ.T., E.K.B., E.B., Analysis or Interpretation: Ö.Ö., M.S., B.D., G.B., Literature Search: Ö.Ö., Writing: Ö.Ö., B.D., G.B.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Villacin AB, Brigham LN, Bullough PG. Primary and secondary synovial chondrometaplasia: histopathologic and clinicoradiologic differences. Hum Pathol. 1979;10:439-51.
2. Adelani MA, Wupperman RM, Holt GE. Benign synovial disorders. J Am Acad Orthop Surg. 2008;16:268-75.
3. Neumann JA, Garrigues GE, Brigman BE, Eward WC. Synovial chondromatosis. JBJS Rev. 2016;4:e2.
4. Evans S, Boffano M, Chaudhry S, Jeys L, Grimer R. Synovial chondrosarcoma arising in synovial chondromatosis. Sarcoma. 2014;2014:647939.
5. Agaram NP, Zhang L, Dickson BC, et al. A molecular study of synovial chondromatosis. Genes Chromosomes Cancer. 2020;59:144-51.
6. McCarthy C, Anderson WJ, Vlychou M, et al. Primary synovial chondromatosis: a reassessment of malignant potential in 155 cases. Skeletal Radiol. 2016;45:755-62.
7. Maurice H, Crone M, Watt I. Synovial chondromatosis. J Bone Joint Surg Br. 1988;70:807-11.
8. Murphey MD, Vidal JA, Fanburg-Smith JC, Gajewski DA. Imaging of synovial chondromatosis with radiologic-pathologic correlation. Radiographics. 2007;27:1465-88.
9. Davis RI, Hamilton A, Biggart JD. Primary synovial chondromatosis: a clinicopathologic review and assessment of malignant potential. Hum Pathol. 1998;29:683-8.
10. Jahoda J, Lu A, Cavenah T, et al. High-grade chondrosarcoma arising from synovial chondromatosis: case report and review of the literature. Journal of Medical Case Reports. 2024.
11. McKenzie G, Raby N, Ritchie D. A pictorial review of primary synovial osteochondromatosis. Eur Radiol. 2008;18:2662-9.
12. Ko E, Mortimer E, Fraire AE. Extraarticular synovial chondromatosis: review of epidemiology, imaging studies, microscopy and pathogenesis, with a report of an additional case in a child. Int J Surg Pathol. 2004;12:273-80.
13. Amary F, Perez-Casanova L, Ye H, et al. Synovial chondromatosis and soft tissue chondroma: extraosseous cartilaginous tumor defined by FN1 gene rearrangement. Mod Pathol. 2019;32:1762-71.
14. Hiemcke-Jiwa LS, Bovée JVMG. Chapter 45-biology of cartilage tumor family. Bone Cancer: Bone Sarcomas and Bone Metastases - From Bench to Bedside. 2021;649-62.
15. Halstead AE. IV. floating bodies in joints. Ann Surg. 1895;22:327-42.
16. Fisher AGT. A study of loose bodies composed of cartilage or of cartilage and bone occuring in joints. With special reference to their pathology and etiology. British Journal of Surgery. 1920;8:493-523.
17. Galat DD, Ackerman DB, Spoon D, Turner NS, Shives TC. Synovial chondromatosis of the foot and ankle. Foot Ankle Int. 2008;29:312-7.
18. Utashima D, Matsumura N, Suzuki T, Iwamoto T, Ogawa K. Clinical results of surgical resection and histopathological evaluation of synovial chondromatosis in the shoulder: a retrospective study and literature review. Clin Orthop Surg. 2020;12:68-75.
19. Zmerly H, Mauela M, Akkawi I. Ankle synovial chondromatosis in anterior and posterior compartments. A case report. Acta Biomed. 2020;91:267-70.
20. Milgram JW. The classification of loose bodies in human joints. Clin Orthop Relat Res. 1977;(124):282-91.
21. Wittkop B, Davies AM, Mangham DC. Primary synovial chondromatosis and synovial chondrosarcoma: a pictorial review. Eur Radiol. 2002;12:2112-9.
22. Jadawala VH, Deshpande S, Ahmed S, Goel S, Suneja A. Synovial chondromatosis in a young athlete: a report of a rare case. Cureus. 2024;16:e53173.
23. Ji JH, Shafi M, Jeong DS. Secondary synovial chondromatosis of the shoulder. Knee Surg Sports Traumatol Arthrosc. 2015;23:2624-27.
24. Hopyan S, Nadesan P, Yu C, Wunder J, Alman BA. Dysregulation of hedgehog signalling predisposes to synovial chondromatosis. J Pathol. 2005;206:143-50.



Supplementary Figure 1. The distribution of synovial chondromatosis cases

DOI: 10.4274/forbes.galenos.2025.65668
Forbes J Med 2025;6(2):196-9

Case Report / Olgu Sunumu

Ventricular Fibrillation Caused by a Low-Rate Pacemaker

Düşük Hızlı Pacemakerin Neden Olduğu Ventriküler Fibrilasyon Olgusu

Hüseyin YILDIZ[1], Ayşe SÜLÜ[2], Osman BAŞPINAR[2]

[1]Medicana Konya Hospital, Clinic of Pediatric Health and Diseases, Konya, Türkiye
[2]Gaziantep University Faculty of Medicine, Department of Pediatric Cardiology, Gaziantep, Türkiye

Cite as: Yıldız H, Sülü A, Başpınar O. Ventricular fibrillation caused by a low-rate pacemaker. Forbes J Med. 2025;6(2):196-9

ABSTRACT

After undergoing surgery for a perimembranous ventricular septal defect and discrete membrane resection, a 7-year-old patient developed complete atrioventricular block and received an epicardial pacemaker. When the epicardial pacemaker reached end-of-life status, a transvenous pacemaker was implanted. Due to the low basal rate setting of this newly implanted transvenous pacemaker, the patient experienced bradycardia-induced acquired long-QT syndrome (LQTS), subsequently resulting in episodes of ventricular fibrillation. This case report discusses the mechanisms leading to acquired LQTS and fibrillation, and emphasizes the importance of an appropriate basal pacing rate, especially in pediatric patients, to avoid life-threatening arrhythmias.

Keywords: Acquired long QT syndrome, pacemaker, bradycardia, ventricular fibrillation

ÖZ

Perimembranöz ventriküler septal defekt ve diskret membran rezeksiyonu ameliyatı geçiren 7 yaşındaki bir hastada, ameliyat sonrasında tam atriyoventriküler blok gelişmiş ve bu nedenle epikardiyal pacemaker implante edilmiştir. Epikardiyal pacemaker'in ömrünün sonuna gelmesi üzerine, transvenöz bir pacemaker takılmıştır. Ancak yeni implante edilen transvenöz pacemaker'in düşük bazal hız ayarı, hastada bradikardi kaynaklı edinsel uzun QT sendromuna (LQTS) neden olmuş ve bu da ventriküler fibrilasyon atakları ile sonuçlanmıştır. Bu olgu sunumunda, edinsel LQTS ve fibrilasyona yol açan olası mekanizmalar ele alınmakta ve özellikle pediatrik hastalarda uygun bazal pace hızının, yaşamı tehdit edebilecek aritmileri önlemedeki önemi vurgulanmaktadır.

Anahtar Kelimeler: Kazanılmış uzun QT sendromu, pacemaker, bradikardi, ventriküler fibriiasyon

Received/Geliş: 17.02.2025
Accepted/Kabul: 12.06.2025
Epub: 17.07.2027
Yayınlanma Tarihi/ Publication Date: 06.08.2025

Corresponding Author/ Sorumlu Yazar:
Osman BAŞPINAR, MD,
Gaziantep University Faculty of Medicine, Department of Pediatric Cardiology, Gaziantep, Türkiye
osmanbaspinar@hotmail.com
ORCID: 0000-0002-9307-0344

INTRODUCTION

Long QT syndrome (LQTS) is a cardiac repolarization disorder that can be congenital or acquired, potentially leading to life-threatening arrhythmias and sudden cardiac death.[1] The QT interval encompasses both depolarization and repolarization phases; however, prolongation of the repolarization phase is the principal contributor to LQTS. Clinical manifestations range from mild symptoms such as dizziness to fatal events, often precipitated by torsade de pointes (TdP). Moreover, TdP may degenerate into ventricular fibrillation (VF), causing sudden cardiac death.[1]

Acquired LQTS is more common than congenital LQTS. Bradyarrhythmias, including those induced by inadequate pacing rates, represent a significant risk factor by prolonging the action potential duration and facilitating early after depolarizations.[2] In the pediatric population, prompt identification and management of acquired LQTS is crucial given the higher incidence of severe or fatal outcomes compared to adults, even in seemingly low-risk clinical settings.[1]

CASE REPORT

A 7-year-old, 17 kg male patient developed complete atrioventricular (AV) block following surgery for a perimembranous ventricular septal defect, and discrete membrane resection. An epicardial pacemaker was previously implanted; however, when the device reached its end-of-life, a transvenous pacemaker was implanted. A single chamber pacemaker generator was attached to the lead, the pacemaker's basal rate was set to 50 beats per minute. During the procedure, the patient's intrinsic heart rate was noted to be very low (35-40 bpm) when the pacemaker was temporarily off.

Eighteen hours post-implantation, the patient experienced brief episodes of VF (Figure 1). Between these episodes, the patient's heart rhythm was paced at 50 bpm, and there were no hemodynamic instabilities. A review of the patient's electrocardiograms after implantation revealed a prolonged corrected QT ranging from 507 ms to 547 ms (Figure 2). The patient started taking propranolol. Pacemaker function was verified and found to be effective. However, as VF episodes persisted, the pacemaker's basal rate was increased from 50 bpm to 90 bpm, after which VF episodes ceased, and a stable paced rhythm was maintained.

Because the patient's QT interval remained prolonged (QTc >500 ms), an implantable cardioverter-defibrillator (ICD) with pacemaker capability was deemed necessary. The transvenous pacemaker was removed, and a single chamber ICD was implanted at a more appropriate basal rate, taking the patient's age-related heart rate range into account. During monitored follow-up, no further VF episodes were observed. The ICD functioned without complications, and the patient's clinical course remained uneventful under regular outpatient follow-up. Informed consent was obtained from patient's parents to publish this very special medical condition.

DISCUSSION

Acquired LQTS occurs more frequently than the congenital form; its exact incidence remains unknown.[1] The most common causes include medications that prolong the QT interval ranging from antiarrhythmic agents (quinidine, procainamide, sotalol, amiodarone) to various non-cardiac drugs (macrolide antibiotics, certain antifungals, some antipsychotics) and electrolyte imbalances (hypokalemia, hypomagnesemia, hypocalcemia).[1-3] Other etiologies include any process causing bradycardia, such as advanced AV block, sick sinus syndrome, hypothyroidism, or hypothermia, each of which increases the risk for early after depolarizations and TdP.[1-4]

In this patient, no potential QT-prolonging drugs or significant electrolyte disturbances were identified, and thyroid function tests were normal. Electrocardiograms did not suggest any form of sick sinus syndrome. However, both the patient's intrinsic rhythm (35-40 bpm) and the pacemaker's set basal rate (50 bpm) were markedly bradycardic relative to the normal pediatric range of 65-135 bpm for children aged 6-10 years. This supports the conclusion that the patient developed bradycardia-induced acquired LQTS secondary to the excessively low pacemaker setting.

Pacemakers can shorten QT duration by maintaining a higher heart rate and preventing pauses that predispose early after

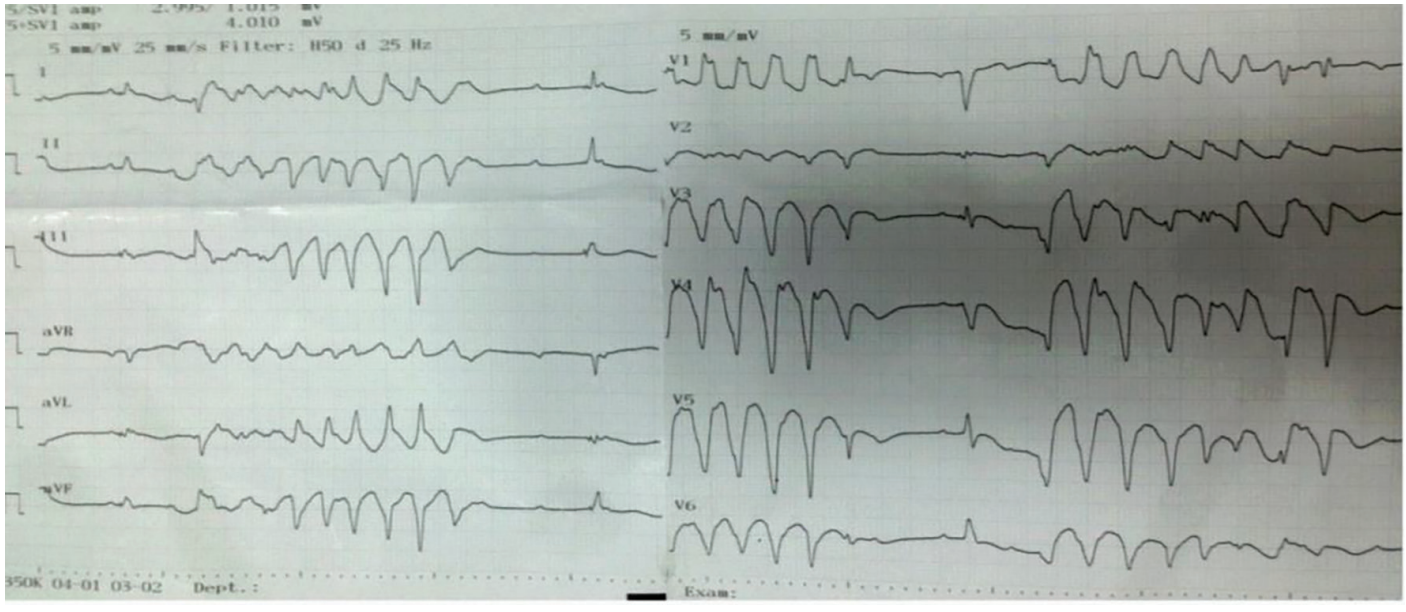


Figure 1. Ventricular fibrillation observed following transvenous pacemaker implantation

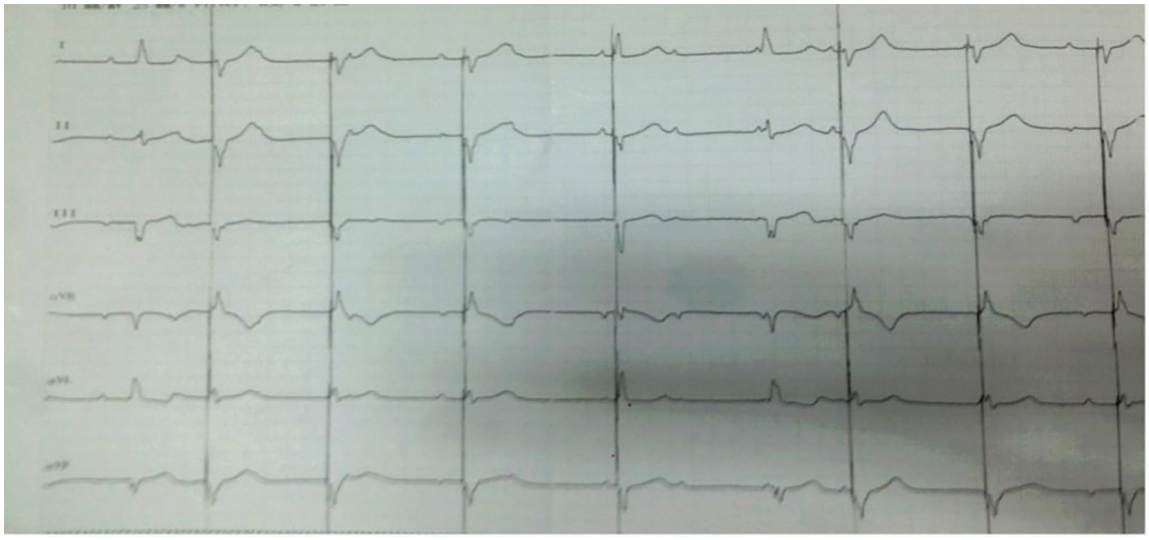


Figure 2. Prolonged QT interval after setting the transvenous pacemaker at a low basal rate

depolarizations. Most pediatric guidelines recommend a basal pacing rate above 80 bpm in patients at risk for LQTS or TdP to ensure more homogeneous repolarization and reduce arrhythmic risk.[5,6] Consistent with these recommendations, increasing the pacemaker rate from 50 bpm to 90 bpm in our patient immediately abolished VF episodes, indicating a clear link between bradycardia and the development of LQTS.

Even in adult patients, cases have been reported where a reduction in the pacemaker's basal rate (hysteresis, oversensing, premature ventricular contractions, postventricular atrial refractory period, medications) led to QTc prolongation and TdP. In these cases, TdP was not reported after adjusting the basal rate to >70 beats per minute. It has been suggested that particularly in patients with an increased risk of TdP, such as those using QT-prolonging medications, the basal rate may be set higher.[5] Similarly our case, Bernstein et al.[6] reported a case of a pediatric patient with congenital AV block who presented with syncope and was found to have QTc prolongation associated with bradycardia, as well as ventricular tachycardia at a rate of 300 beats per minute, potentially consistent with torsades, during pacemaker interrogation. They emphasized the importance of regular QTc monitoring in the follow-up of these patients after pacemaker implantation.[6]

Children with LQTS, whether congenital or acquired, may present with life-threatening events, such as cardiac arrest, more often than adults. Hence, therapy is recommended even for asymptomatic pediatric patients.[3] Beta-blockers reduce the likelihood of arrhythmic events in up to 70% of LQTS patients, but around 30% may still experience breakthrough arrhythmias or sudden death despite optimal medical therapy. For those with a history of cardiac arrest or recurrent arrhythmic events despite beta-blockade, ICD implantation is the first-line option.[3] ICDs provide both anti-bradycardia pacing and arrhythmia termination, significantly improving survival in pediatric patients at high risk for lethal arrhythmias.[3-5]

CONCLUSION

To extend pacemaker battery life, clinicians may be tempted to program excessively low basal rates, but this can pose a significant risk factor for acquired LQTS and subsequent malignant arrhythmias in pediatric patients. Maintaining a basal pacing rate appropriate for the patient's age can help prevent the development of bradycardia-induced LQTS and VF episodes. In this case, raising the basal pacemaker rate effectively eliminated VF attacks. Consequently, we recommend carefully considering the minimum basal heart rate for pediatric pacemaker programming, ensuring it remains within a physiologically safe range to minimize arrhythmic complications.

Ethics

Informed Consent: Informed consent was obtained from patient's parents to publish this very special medical condition.

Footnotes

Authorship Contributions

Surgical and Medical Practices: O.B., Concept: O.B., Design: O.B., Data Collection or Processing: H.Y., A.S., Analysis or Interpretation: H.Y., A.S., Literature Search: H.Y., A.S., Writing: H.Y.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Jackman WM, Friday KJ, Anderson JL, Aliot EM, Clark M, Lazzara R. The long QT syndromes: a critical review, new clinical observations and a unifying hypothesis. Prog Cardiovasc Dis. 1988;31:115-72.
2. Ishida S, Takahashi N, Nakagawa M, Fujino T, Saikawa T, Ito M. Relation between QT and RR intervals in patients with bradyarrhythmias. Br Heart J. 1995;74:159-62.
3. Zhu W, Bian X, Lv J. From genes to clinical management: a comprehensive review of long QT syndrome pathogenesis and treatment. Heart Rhythm 02. 2024;5:573-86.
4. Bhattad PB, Jha A, Wholey R. Bradycardia-induced torsades de pointes in atrioventricular block. Cureus. 2023;15:37507.
5. Ashworth SW, Levsky ME, Marley CT, Kang CS, Bradycardia-associated torsade de pointes and the long-QT syndromes: a case report and review of the literature. Mil Med. 2005;170:381-6.
6. Bernstein BS, Silver ES, Liberman L. QT prolongation and torsades de pointes in a patient with heart block and a pacemaker. Cardiol Young. 2016;26:161-3.