Nobel Med 2026; 22(2): 125-130

PES PLANUS TANI VE TEDAVİSİNDE CHATGPT-5.2'NİN DOĞRULUĞUNUN DEĞERLENDİRİLMESİ: YAPAY ZEKA TABANLI BİR ÇALIŞMA

Muhammed Taha Demir

İstanbul Aydın Üniversitesi, Sağlık Bilimleri Fakültesi, Fizyoterapi ve Rehabilitasyon. İstanbul, Türkiye

ÖZET

Amaç: Bu çalışmanın amacı, ChatGPT-5.2 (OpenAI, ABD) modelinin pes planus tanı, değerlendirme ve tedavi konularındaki yanıt doğruluğunu sistematik olarak incelemek ve yapay zekâ modellerinin ortopedik klinik uygulamalardaki olası katkısını ortaya koymaktır.

Materyal ve Metot: Pes planus ile ilgili tanı, klinik değerlendirme, radyolojik bulgular ve tedavi yaklaşımlarını kapsayan beş seçenekli, çoktan seçmeli güncel literatür bilgilerini kapsayan 30 soru hazırlandı. Sorular, biri ayak cerrahisi alanında deneyimli olmak üzere üç ortopedi uzmanı tarafından çok aşamalı değerlendirme süreci ile oluşturularak ChatGPT-5.2 modeline yöneltildi. Beş seçenekli sorular için %20 rastlantısal başarı olasılığı varsayılarak doğruluk oranı binom testi ile analiz edildi ve %95 güven aralığı Wilson yöntemi ile hesaplandı. Soru kategorileri arası doğruluk oranları Fisher's exact test ile analiz edildi. Model yanıtları üç uzman tarafından bağımsız değerlendirildi ve gözlemciler arası uyum Fleiss' kappa katsayısıyla incelendi.

Bulgular: ChatGPT-5.2 modeli 30 sorunun 24'üne doğru yanıt vererek %80 doğruluk oranı göstermiştir (95% GA: %62,7–%90,5). %20'lik rastlantısal başarı düzeyi varsayımı altında, doğruluk oranı tek yönlü binom testi ile anlamlı bulunmuştur (p<0,001). Kategoriler arası analizde tanım, anatomi ve klinik değerlendirme sorularında yüksek performans; tedavi ve cerrahi endikasyonlara ilişkin sorularda ise daha düşük doğruluk oranı saptanmıştır. Uzmanların cevapları değerlendirmesindeki uyum mükemmel düzeyde bulunmuştur (Fleiss' κ=0,89). Yanlış yanıtların yarısı yüksek klinik risk düzeyindedir ve bunların tamamı tedavi/cerrahi kategorisinde yer almaktadır. ChatGPT-5.2 performansı kategorilere göre anlamlı farklılık göstermiştir (p=0,048).

Sonuç: ChatGPT-5.2, pes planus hakkında ortopedik bilgi sağlanmasında ve hasta eğitimi süreçlerinde yardımcı bir araç olabilir. Bununla birlikte, klinik karar verme süreçlerinde tek başına kullanılmamalı ve uzman hekim değerlendirmesinin yerini almamalıdır.

Anahtar kelimeler: Yapay zeka, ChatGPT, pes planus.

İLETİŞİM İÇİN: Muhammed Taha Demir Ataköy 7-8-9-10. kısım mah. Çobançeşme E-5 yan yol caddesi. Ataköy Towers, A Blok, No 79, Bakırköy, İstanbul, Türkiye
mtahademir@yahoo.com

MTD https://orcid.org/0000-0003-0593-0816

GÖNDERİLDİĞİ TARİH: 02 / 03 / 2026 • **KABUL TARİHİ:** 05 / 03 / 2026

EVALUATION OF THE ACCURACY OF CHATGPT-5.2 IN THE DIAGNOSIS AND TREATMENT OF PES PLANUS: AN ARTIFICIAL INTELLIGENCE–BASED STUDY

ABSTRACT

Objective: *This study aimed to systematically assess the accuracy of ChatGPT-5.2 (OpenAI, USA) in responding to questions on the diagnosis, evaluation, and treatment of pes planus, and to investigate the potential role of artificial intelligence models in orthopaedic clinical practice.*

Material and Method: *A total of 30 five-option multiple-choice questions covering the diagnosis, clinical evaluation, radiological findings, and treatment approaches of pes planus were prepared based on current literature. The questions were developed through a multi-stage expert evaluation process by three orthopaedic specialists, one of whom had specific experience in foot and ankle surgery, and were subsequently submitted to the ChatGPT-5.2 model. Assuming a 20% probability of random success for five-option questions, accuracy was analyzed using a binomial test, and the 95% confidence interval was calculated using the Wilson method. Differences in accuracy rates across question categories were analyzed using Fisher's exact test. Model responses were independently evaluated by three experts, and interobserver agreement was assessed using Fleiss' kappa coefficient.*

Results: *The ChatGPT-5.2 model correctly answered 24 of 30 questions, yielding an overall accuracy of 80% (95% CI: 62.7%–90.5%). Under the assumption of a 20% random success rate, the observed accuracy was found to be significant using a one-sided binomial test ($p<0.001$). Category-based analysis demonstrated high performance in definition, anatomy, and clinical evaluation questions, whereas lower accuracy was observed in treatment and surgical indication items. Inter-rater agreement among experts was at a near-perfect level (Fleiss' $\kappa=0.89$). Half of the incorrect responses were classified as high clinical risk, all of which were within the treatment/surgical category. ChatGPT-5.2 performance differed significantly across categories ($p=0.048$).*

Conclusion: *ChatGPT-5.2 may serve as a useful supplemental tool for orthopaedic information and patient education regarding pes planus; however, it should not replace specialist clinical judgment.*

Keywords: *Artificial intelligence, chatGPT, pes planus.*

GİRİŞ

Yapay zeka (Artifical Intelligence, AI), son yıllarda sağlık alanında hızla gelişen ve klinik uygulamalarda giderek daha fazla yer bulan bir teknolojidir. Özellikle büyük dil modelleri (LLM) arasında yer alan ChatGPT-5.2 (OpenAI, ABD), kullanıcıların doğal dilde sorduğu sorulara anlamlı ve tutarlı yanıtlar verebilme kapasitesi ile dikkat çekmektedir.[1] Bu tür modeller, hem hastalar hem de sağlık profesyonelleri tarafından hastalıklar hakkında bilgi edinmek, tanı süreçlerini anlamak ve tedavi seçeneklerini değerlendirmek amacıyla giderek daha sık kullanılmaktadır.[2,3]

AI tabanlı sistemler, kullanıcıya özgü, bağlamsal ve daha detaylı yanıtlar sağlayabilmektedir.[3] Bununla birlikte, bu sistemlerin sağladığı bilgilerin doğruluğu, güvenilirliği ve klinik karar verme sürecindeki rolü ile ilgili önemli tartışmalar devam etmektedir.[4] Yapılan çalışmalar, ChatGPT modellerinin birçok ortopedik konuda doğru bilgiler sağlayabildiğini göstermiştir.[5,6]

Toplumda görülme sıklığı nadir olmayan pes planus (PP), medial longitudinal arkın azalması veya kaybı ile karakterize yaygın bir ayak deformitesidir.[7,8] Bu deformite çocukluk çağında sıklıkla fizyolojik ve asemptomatik olmakla birlikte bazı bireylerde ağrı, fonksiyonel kısıtlılık ve ilerleyici deformiteye yol açabilir.[9-11] PP tanısı klinik muayene ve radyolojik değerlendirme ile konulmakta olup, tedavi seçenekleri konservatif yaklaşımlardan cerrahi rekonstrüksiyona kadar geniş bir yelpazeyi kapsamaktadır.[12-14] Cerrahi endikasyonların doğru belirlenmesi, hasta sonuçlarını doğrudan etkileyen önemli bir faktördür.[13,15] Yapılan çalışmalarda ChatGPT modellerinin klinik karar verme sürecinde tek başına kullanılmasının uygun olmadığını göstermektedir.[5,6,16,17] Özellikle ayak ve ayak bileği patolojileri gibi çok faktörlü değerlendirme gerektiren durumlarda, AI sistemlerinin doğruluğunun objektif olarak değerlendirilmesi ilerde bu sistemlerin kullanılabilirliklerinin yaygınlaşması adına büyük önem taşımaktadır.

Bu çalışmanın amacı, ChatGPT-5.2 modelinin PP'de tanı, klinik değerlendirme ve tedavi ile ilgili çoktan seçmeli sorulara verdiği yanıtların doğruluğunu değerlendirmek ve AI tabanlı LLM'nin ortopedik klinik uygulamalardaki potansiyel rolünü ortaya koymaktır.

MATERYAL ve METOT

PP ile ilgili tanı, klinik değerlendirme, radyolojik bulgular ve tedavi yaklaşımlarını kapsayan toplam

30 adet çoktan seçmeli soru hazırlanmıştır. Soru hazırlama süreci, çok aşamalı uzman değerlendirme yaklaşımı doğrultusunda yürütülmüş; sorular biri ayak cerrahisi alanında deneyimli, diğer ikisi genel ortopedi pratiği yürüten üç uzman tarafından kapsam, doğruluk ve klinik temsil gücü açısından karşılıklı değerlendirilerek oluşturulmuştur. Ayak cerrahisi alanında deneyimli uzmanın sürece dahil edilmesi, özellikle cerrahi teknikler ve endikasyonlara ilişkin soruların bilimsel doğruluğunu ve güncelliğini sağlamak amacıyla planlanmıştır. Genel ortopedi pratiği yürüten uzmanların katkısı ise soruların yalnızca belirli bir uzmanlık alanına özgü kalmaması, genel ortopedik klinik pratiği temsil edecek şekilde dengeli ve uygulanabilir olması amacıyla sağlanmıştır. Soruların içerik geçerliliği, kapsam uygunluğu, bilimsel doğruluğu ve klinik temsil gücü üç uzman arasında görüş birliği sağlandıktan sonra nihai hale getirilmiştir. Sorular güncel ortopedi literatürü temel alınarak oluşturulmuştur.[7-9,11-15,18] Zorluk düzeyi, temel, orta ve ileri düzey klinik bilgi gereksinimi dikkate alınarak planlanmış ve üç uzman arasında sağlanan konsensüs doğrultusunda dengelenmiştir. Sorular dört ana kategoriye ayrılmıştır: Tanım ve anatomi (*n*=8), klinik değerlendirme ve tanı (*n*=8), radyolojik değerlendirme (*n*=7), tedavi ve cerrahi endikasyonlar (*n*=7). Her soru beş seçenekli olarak hazırlanmış ve yalnızca bir doğru cevap içerecek şekilde düzenlenmiştir. Hazırlanan sorular, ChatGPT-5.2 (OpenAI) modeline standart bir kullanıcı arayüzü üzerinden yöneltilmiştir. Tüm sorular aynı oturum içerisinde, ek klinik yönlendirme, ipucu veya açıklama verilmeden sorulmuştur. Her soru için ChatGPT-5.2 modelinden yalnızca bir cevap seçmesi istenmiştir. Değerlendirme Ekim 2025 tarihinde gerçekleştirilmiştir. ChatGPT-5.2 tarafından verilen yanıtlar kaydedilmiş ve her bir yanıt önceden belirlenmiş doğru cevaplar ile karşılaştırılmıştır. Uzmanlar verilen cevapları bağımsız olarak değerlendirdi.

ChatGPT-5.2 modelinin performansı; toplam doğru-yanlış cevap sayısı, toplam doğruluk oranı (%) ve kategori bazlı doğruluk oranı (%) parametreleriyle değerlendirilmiştir.

Yanlış yanıtlanan sorular klinik etki ve potansiyel risk düzeyi açısından değerlendirilmiştir. Her bir yanlış yanıt; cerrahi karar sürecine etkisi, hasta yönetiminde gecikmeye yol açma potansiyeli ve olası klinik sonuçları dikkate alınarak üç ortopedi uzmanı tarafından sınıflandırılmıştır. Risk düzeyi düşük, orta ve yüksek olarak belirlenmiştir.

İstatistiksel Analiz

İstatistiksel analizler IBM SPSS Statistics for Windows (Version 26.0, IBM Corp., Armonk, NY, USA) yazılımı kullanılarak gerçekleştirildi. Modelin doğruluk oranının rastlantısal başarı düzeyinden farklı olup olmadığını değerlendirmek amacıyla, beş seçenekli sorular için rastlantısal doğru cevap olasılığı %20 olarak kabul edilmiş; doğruluk oranının %95 güven aralığı Wilson yöntemi ile hesaplanmış ve gözlenen başarı oranı tek yönlü binom testi ile analiz edilmiştir. Cevaplar üç ortopedi uzmanı tarafından bağımsız olarak değerlendirilmiş ve gözlemciler arası uyum Fleiss' kappa katsayısı ile değerlendirilmiştir. Kappa değerleri Landis ve Koch sınıflamasına göre yorumlanmıştır. Soru kategorileri bazlı doğruluk oranları arasındaki fark, örneklem büyüklüğünün küçük olması nedeniyle Fisher's exact test kullanılarak analiz edilmiştir.

Bu çalışmada insan katılımcılar, hasta verileri, radyolojik görüntüler veya kişisel sağlık verileri kullanılmamıştır. Çalışma yalnızca AI modelinin performans değerlendirmesini içerdiğinden etik kurul onayı gerekmemektedir.

BULGULAR

ChatGPT-5.2 modeli 30 sorunun 24'üne doğru yanıt vererek %80,0 doğruluk oranı göstermiştir; Wilson yöntemi ile hesaplanan %95 güven aralığı (%62,7–%90,5) ise, örneklem büyüklüğü dikkate alındığında modelin gerçek doğruluk oranının büyük olasılıkla bu aralıkta yer aldığını göstermektedir. Soruların beş seçenekli olması nedeniyle rastlantısal doğru cevap olasılığı %20 olarak kabul edildiğinde, ChatGPT-5.2 modelinin gözlenen doğruluk oranının rastlantısal tahmin düzeyinden istatistiksel olarak anlamlı derecede yüksek olduğu binom testi ile görülmüştür ($p<0,001$). (Tablo 1)

Kategori bazlı değerlendirme sonuçları Tablo 2'de gösterilmiştir. Tanım ve anatomi kategorisinde 8

Tabo 1. ChatGPT-5.2 modelinin genel performans değerlendirmesi

Parametre	Değer
Toplam soru sayısı	30
Doğru cevap sayısı	24
Yanlış cevap sayısı	6
Doğruluk oranı (%)	80,0
%95 güven aralığı	62,7 – 90,5
Rastlantısal doğruluk oranı (%)	20
Binom testi	$p<0,001$

Tablo 2. ChatGPT-5.2 modelinin kategori bazlı doğruluk oranları

Kategori	Toplam soru sayısı	Doğru cevap sayısı	Yanlış cevap sayısı	Doğruluk oranı (%)
Tanım ve anatomi	8	7	1	87,5
Klinik değerlendirme ve tanı	8	7	1	87,5
Radyolojik değerlendirme	7	6	1	85,7
Tedavi ve cerrahi endikasyonlar	7	4	3	57,1
Toplam	30	24	6	80,0

Kategoriler arası doğruluk oranları Fisher exact test ile karşılaştırılmış ve anlamlı bir fark görülmüştür (p=0,048).

Tablo 3. AI (Yapay zeka) tarafından Yanlış Yanıtlanan Soruların Klinik Etki ve Risk Düzeyi Analizi

Soru No	Kategori	Cerrahi Kararı Etkiliyor mu?	Klinik Risk Düzeyi	Klinik Değerlendirme
2	Tanım ve anatomi	Hayır	Düşük	Tanı kavramsal düzeyde; doğrudan tedavi kararını etkilemez
14	Klinik değerlendirme ve tanı	Potansiyel	Orta	Tanısal yaklaşımda gecikmeye yol açabilir
18	Radyolojik değerlendirme	Potansiyel	Orta	Yanlış görüntüleme tercihi cerrahi planlamayı geciktirebilir
24	Tedavi ve cerrahi endikasyonlar	Evet	Yüksek	Cerrahi gerekliliğin yanlış değerlendirilmesi hasta yönetimini doğrudan etkileyebilir
27	Tedavi ve cerrahi endikasyonlar	Evet	Yüksek	Cerrahi endikasyon kararında sapma riski
28	Tedavi ve cerrahi endikasyonlar	Evet	Yüksek	Uygun tedavi seçiminin değişmesine yol açabilir

AI: Yapay zeka, **LLM:** büyük dil modelleri, **PP:** pes planus..

sorunun 7'sine doğru yanıt verilmiş ve doğruluk oranı %87,5 olarak bulunmuştur. Klinik değerlendirme ve tanı kategorisinde 8 sorunun 7'si doğru yanıtlanmış ve doğruluk oranı %87,5 olarak hesaplanmıştır. Radyolojik değerlendirme kategorisinde 7 sorunun 6'sına doğru yanıt verilmiş olup doğruluk oranı %85,7 olarak bulunmuştur. Tedavi ve cerrahi endikasyonlar kategorisinde doğruluk oranı daha düşük olup, 7 sorunun yalnızca 4'ü doğru yanıtlanmış ve doğruluk oranı %57,1 olarak hesaplanmıştır.

Cevapların değerlendirilmesinde uzmanlar arası uyum analizi sonucunda Fleiss' kappa katsayısı 0,89 olarak hesaplanmış ve Landis-Koch sınıflamasına göre neredeyse mükemmel düzeyde uyum saptanmıştır.

Yanlış yanıtlanan 6 soru, gerçek klinik pratikte yol açabileceği olası sonuçlar ve klinik risk düzeyi bakımından ayrıntılı olarak analiz edilmiştir (Tablo 3). Bu soruların 1'i düşük risk, 2'si orta risk ve 3'ü yüksek risk düzeyinde sınıflandırılmıştır. Yüksek riskli yanlış yanıtların tamamı tedavi ve cerrahi endikasyonlar kategorisinde yer almıştır.

Kategoriler arasında doğruluk oranları karşılaştırıldığında istatistiksel olarak anlamlı bir fark saptanmıştır (Fisher exact test, p=0,048). Bu bulgu, modelin tüm konu başlıklarında benzer performans göstermediğini ve en düşük doğruluk oranının tedavi ve cerrahi endikasyonlar kategorisinde gözlendiğini ortaya koymaktadır.

TARTIŞMA

Bu çalışmada, ChatGPT-5.2 modelinin PP ile ilgili çoktan seçmeli sorulardaki genel doğruluk oranı %80 olarak bulunmuştur (%95 GA: %62,7–%90,5; p<0,001). Modelin gözlenen doğruluk oranının rastlantısal tahmin düzeyinin (%20) anlamlı derecede üzerinde olması, performansın tesadüfi olmadığını ve ChatGPT-5.2 modelinin PP konusunda iyi düzeyde bilgi doğruluğuna sahip olduğunu göstermektedir. Soru kategorilerinin doğruluk oranları karşılaştırıldığında istatistiksel olarak anlamlı fark saptanmış (Fisher exact test, p=0,048) ve en düşük doğruluk oranının tedavi ve cerrahi endikasyonlar kategorisinde olduğu görülmüştür. Elde edilen doğruluk oranı, ortopedi alanında ChatGPT modellerinin performansını değerlendiren önceki çalışmalarla uyumludur.[4-6,16,17] Bu çalışmalar, LLM'lerin temel ortopedik bilgi sağlama konusunda başarılı olduğunu, ancak özellikle klinik karar verme gerektiren daha kompleks konularda performansın olumsuz olabileceğini göstermiştir. Kung JE ve ark., ChatGPT'nin ortopedik sınav sorularında orta-iyi düzeyde doğruluk gösterdiğini ve klinik karar destek aracı olarak kullanılmadan önce dikkatli değerlendirilmesi gerektiğini belirtmiştir.[5] Benzer şekilde ChatGPT'nin birçok tıbbi konuda doğru yanıtlar verebildiği, ancak bazı kritik klinik sorularda hatalar yapabildiği gösterilmiştir.[16] Jaleel ve ark. tarafından yapılan sistematik derleme ve meta-analizde, ChatGPT-4.0 ve 3.5'in tıp eğitimini destekleme ve klinik karar verme süreçlerine katkı sağlama potansiyeline sahip olduğu, ancak etkili tıbbi uygulama için gerekli kapsamlı klinik becerilerin yerini alamayacağı sonucuna varılmıştır.[17]

PP, klinik değerlendirme ve tedavi açısından heterojen bir patoloji olup; etiyoloji, deformite derecesi ve hastaya özgü faktörlere bağlı olarak farklı tedavi yaklaşımları gerektirmektedir.[7-9,11-14] Bu çok boyutlu ve hasta-özel değişkenlere dayalı yapı, cerrahi endikasyonlara yönelik sorularda daha düşük doğruluk oranı saptanmasını açıklayabilir; zira genellenmiş bilgi kalıplarına dayalı yanıt üretimi, bireyselleştirilmiş klinik karar verme süreçlerinde sınırlı kalabilmektedir. Bu bağlamda yapılan klinik

etki analizi, yanlış yanıtların bir kısmının cerrahi karar sürecini ve hasta yönetimini doğrudan etkileyebilecek nitelikte olduğunu göstermiştir. Özellikle tedavi ve cerrahi endikasyonlar kategorisinde yer alan yanlış yanıtların yüksek risk düzeyinde sınıflandırılması, modelin cerrahi karar verme süreçlerinde dikkatli kullanılmasının gerekliliğini ortaya koymaktadır.

Yanıtların üç ortopedi uzmanı tarafından bağımsız olarak değerlendirilmiş olması ve gözlemciler arası uyumun yüksek düzeyde bulunması (Fleiss' κ=0,89), elde edilen sonuçların güvenilirliğini güçlendirmektedir. Bu metodolojik yaklaşım, çalışmanın yalnızca model doğruluk oranını değil, aynı zamanda değerlendirme sürecinin güvenilirliğini de ortaya koymasını sağlamıştır. Bununla birlikte, ChatGPT-5.2 modelinin PP ile ilgili temel ortopedik bilgiler, klinik değerlendirme ve konservatif tedavi yaklaşımları konusundaki yüksek doğruluk oranı, bu tür AI sistemlerinin hasta eğitimi ve genel klinik bilgi sağlanması amacıyla güvenilir bir yardımcı araç olarak kullanılabileceğini desteklemektedir.[2,6,18] Ancak cerrahi tedavi ve klinik karar verme gerektiren durumlarda gözlenen daha düşük doğruluk oranı ve yüksek riskli hata potansiyeli, bu sistemlerin bağımsız bir klinik karar verici olarak değil, hekim tarafından kullanılan destekleyici bir araç olarak değerlendirilmesi gerektiğini göstermektedir. Gelecekte daha gelişmiş AI modelleri ve daha geniş klinik veri setleri ile yapılacak çalışmalar, bu sistemlerin ortopedik klinik uygulamalardaki potansiyel rolünü daha net ortaya koyacaktır.

Çalışmanın Sınırlılıkları

Soru sayısının 30 ile sınırlı olması, elde edilen doğruluk oranlarının genellenebilirliğini kısmen sınırlandırabilir. Tek bir AI modeli ve belirli bir zaman noktasındaki sürüm performansını değerlendirmektedir; model güncellemeleri ile sonuçlar değişkenlik gösterebilir. Ayrıca değerlendirme çoktan seçmeli soru formatı üzerinden yapılmış olup, gerçek hasta senaryoları ve dinamik klinik karar süreçlerini tam olarak yansıtmayabilir.

SONUÇ

Bu bulgular, AI tabanlı LLM'lerin bilgi temelli konularda güvenilir destek sağlayabileceğini; ancak bireyselleştirilmiş klinik değerlendirme ve cerrahi karar verme süreçlerinde performanslarının sınırlı kalabileceğini göstermektedir. Özellikle tedavi ve cerrahi endikasyonlara ilişkin sorularda saptanan düşük doğruluk oranı ve bazı hataların yüksek klinik risk potansiyeli taşıması, bu modellerin tek başına karar verici olarak kullanılmasının uygun olmadığını ortaya koymaktadır. Klinik karar verme sürecinde uzman hekim değerlendirmesi temel belirleyici olmaya devam etmelidir.

Gelecekte daha geniş soru setleri, farklı ortopedik patolojiler ve farklı AI modellerini içeren çalışmalar, bu sistemlerin klinik güvenilirliği ve uygulanabilirliğini daha net ortaya koyacaktır.

*Yazarlar herhangi bir çıkar ilişkisi içinde bulunmadıklarını bildirmiştir.

KAYNAKLAR

1. OpenAI. ChatGPT technical report. OpenAI; 2024. Erişim adresi: https://openai.com. Erişim tarihi: 10 Şubat 2026.
2. Gaudiani MA, Castle JP, Abbas MJ, et al. ChatGPT-4 generates more accurate and complete responses to common patient questions about anterior cruciate ligament reconstruction than Google's search engine. Arthrosc Sports Med Rehabil 2024; 6: 100939.
3. Kim SE, Lee JH, Kim HJ, et al. Performance of ChatGPT on solving orthopedic board-style questions: a comparative analysis of ChatGPT-3.5 and ChatGPT-4. Clin Orthop Surg 2024; 16: 1-9.
4. Lubitz M, Latario L. Performance of two artificial intelligence generative language models on the Orthopaedic In-Training Examination. Orthopedics 2024; 47: e146-e150.
5. Kung JE, Marshall C, Gauthier C, Gonzalez TA, Jackson JB. Evaluating ChatGPT performance on the Orthopaedic In-Training Examination. JBJS Open Access 2023; 8: e23.00056.
6. Ghanem D, Covarrubias O, Raad M, LaPorte D, Shafiq B. ChatGPT performs at the level of a third-year orthopaedic surgery resident on the Orthopaedic In-Training Examination. JBJS Open Access 2023; 8: e23.00103.
7. İnce N, Özyıldırım B, İnce H, ve ark. İstanbul'daki (Silivri) öğretmenlerde mesleksel maruziyete bağlı hastalıkların araştırılması. Nobel Med 2012; 8: 35–41.
8. Bednarczyk E, Sikora S, Kossobudzka-Górska A, Jankowski K, Hernandez-Rodriguez Y. Understanding flat feet: An in-depth analysis of orthotic solutions. J Orthop Rep 2024; 3: 100250.
9. Arangio GA, Salathé EP. Medial displacement calcaneal osteotomy reduces the excess forces in the medial longitudinal arch of the flat foot. Clin Biomech 2001; 16: 535-539.
10. Vergillos-Luna M, Martínez-Nova A, Sánchez-Rodríguez R, Escamilla-Martínez E, Gómez-Martín B. Pediatric flatfoot: when do we need surgical referral? J Clin Med 2023; 12: 3809.

11. Myerson MS, Thordarson DB, Johnson JE, et al. Classification and nomenclature: progressive collapsing foot deformity. Foot Ankle Int 2020; 41: 1271-1276.

12. Henry JK, Shakked R, Ellis SJ. Adult-acquired flatfoot deformity. Foot Ankle Orthop 2019; 4: 2473011418820847.

13. Piraino JA, Theodoulou MH, Ortiz J, et al. American college of foot and ankle surgeons clinical consensus statement: appropriate clinical management of adult-acquired flatfoot deformity. J Foot Ankle Surg 2020; 59: 347-355.

14. Polichetti C, Tarallo L, Mugnai R, Adani R, Catani F. Adult acquired flatfoot deformity: imaging review. Diagnostics 2023; 13: 225

15. Oerlemans LNT, van der Steen MC, Steverink JG, Dekker R, Hijmans JM. Foot orthoses for flexible flatfoot: systematic review and meta-analysis. BMC Musculoskelet Disord 2023; 24: 16.

16. Gilson A, Safranek CW, Huang T, et al. How well does ChatGPT do when taking the medical licensing exams? The implications of large language models for medical education and knowledge assessment. PLOS Digit Health 2023; 2: 0000198.

17. Jaleel A, Aziz U, Farid G, et al. Evaluating the potential and accuracy of ChatGPT-3.5 and 4.0 in medical licensing and in-training examinations: systematic review and meta-analysis. JMIR Med Educ 2025; 11: 68070.

18. Yüce A, Yerli M, Misir A, Çakar M. Enhancing patient information texts in orthopaedics: how OpenAI's ChatGPT can help. J Exp Orthop 2024; 11: 70019.