T.C.

KARADENİZ TEKNİK ÜNİVERSİTESİ

TIP FAKÜLTESİ

ACİL TIP ANA BİLİM DALI

KRİTİK HASTA TANISINDA YAPAY ZEKA PERFORMANSI: CHATGPT, GEMİNİ, CLAUDE VE ACİL SERVİS HEKİMLERİNİN KARŞILAŞTIRILMASI

UZMANLIK TEZİ

Dr. İbrahim GÜNAYDIN

TRABZON-2025

T.C.

KARADENİZ TEKNİK ÜNİVERSİTESİ

TIP FAKÜLTESİ

ACİL TIP ANA BİLİM DALI

KRİTİK HASTA TANISINDA YAPAY ZEKA PERFORMANSI: CHATGPT, GEMİNİ, CLAUDE VE ACİL SERVİS HEKİMLERİNİN KARŞILAŞTIRILMASI

Uzmanlık Tezi

Dr. İbrahim GÜNAYDIN

Tez Danışmanı: Doç. Dr. Sinan PASLI

TRABZON-2025

TEŞEKKÜR

Acil Tıp uzmanlık eğitimim sürecinde, tez çalışmam boyunca değerli bilgi, destek ve rehberliğini esirgemeyen, hocalığı ve abiliğiyle her zaman yanımda olan Doç. Dr. Sinan PASLI'ya,

Eğitim sürecim boyunca bilgi ve tecrübelerini paylaşmaktan hiçbir zaman kaçınmayan, desteğini her daim hissettiğim saygıdeğer hocalarım Prof. Dr. Abdulkadir GÜNDÜZ, Prof. Dr. Süleyman TÜREDİ, Doç. Dr. Özgür TATLI, Doç. Dr. Yunus KARACA, Doç. Dr. Aynur ŞAHİN, Dr. Öğr. Melih İMAMOĞLU, Dr. Öğr. Üyesi Vildan ÖZER ve Dr. Öğr. Üyesi Abdul Samet ŞAHİN'e,

Acil Tıp serüvenine birlikte başladığımız, bilgi, emek ve dostluklarıyla bu süreci daha anlamlı kılan kıymetli arkadaşlarım Uzm. Dr. Muhammet Samet BEKAR ve Uzm. Dr. Süleyman Cem AKIL'a,

Tez sürecimde ve sosyal yaşamımda desteğini hiçbir zaman esirgemeyen, bilgi birikimiyle bana yol gösteren Uzm. Dr. Mutlu YILMAZ ve Uzm. Dr. Tacettin KARAMAN'a,

Eğitimim boyunca bilgi ve tecrübelerinden faydalanmaktan büyük mutluluk duyduğum Uzm. Dr. Alireza MANEVİ, Uzm. Dr. Adem GÜLSOY, Uzm. Dr. Kürşat GÜL, Dr. Alperen Kasımhan ÜNLÜER, Uzm. Dr. Mehmet Erdi YILMAZ, Uzm. Dr. Emre KOÇ, Uzm. Dr. İbrahim TEMİZKAN, Uzm. Dr. Özlem BÜLBÜL, Uzm. Dr. Ceren SELİM ve Uzm. Dr. Hülya KARAKOÇ'a,

Veri toplama sürecinde özverili katkılarından dolayı Dr. Yunus Emre YILMAZ, Dr. Hakan ŞAHİN, Dr. Sinan AKPUNAR, Dr. Songül KURŞUN, Dr. Esra YANIK ve Dr. Nurullah Samet YILMAZ'a,

Birlikte çalışmaktan büyük mutluluk duyduğum, dostluklarıyla güzel anılara vesile olan tüm KTÜ Acil Servis personeline,

Hayatımın her döneminde sevgi, sabır ve emekleriyle yanımda olan, desteklerini hiçbir zaman esirgemeyen canım annem ve babama, her zaman yanımda hissettiğim çok sevdiğim kardeşime,

Tıp yolculuğumun her basamağında yanımda olan, sevgisi, anlayışı ve desteğiyle hayatıma anlam katan, yaşamımı güzelleştiren sevgili eşim Dr. Kiraz TEKİN GÜNAYDIN'a,

Tüm kalbimle saygı, sevgi ve şükranlarımı sunarım.

Dr. İbrahim GÜNAYDIN

ÖZET

KRİTİK HASTA TANISINDA YAPAY ZEKA PERFORMANSI: CHATGPT, GEMİNİ, CLAUDE VE ACİL SERVİS HEKİMLERİNİN KARŞILAŞTIRILMASI

Amaç: Bu çalışmanın amacı, acil servise başvuran kritik hastalarda, hekimin tanısal sürecini ChatGPT, Claude ve Gemini'nin oluşturduğu tanısal çıktılarıyla karşılaştırmaktır.

Gereç ve yöntem: Acil servise başvuran kritik hastaların vital bulgu ve anamnez bilgilerine göre hekim tarafından, henüz fizik muayene ve tetkik-görüntüleme aşamasına geçilmeden önce en olası 5 adet ön tanı belirlendi. Aynı veriler AI modellerine kimliksizleştirilmiş olarak girildi. Modellerden en olası 5 ön tanı oluşturması istendi. İkinci aşamada hekim, fizik muayene, laboratuvar ve görüntüleme sonuçlarıyla ön tanılarını gözden geçirdi ve 3 ayırıcı tanı olarak güncelledi. Bu verilerle AI modellerinden 3 ayırıcı tanı oluşturması istendi. Son aşamada toplanan tüm veriler ışığında hastanın klinik durumuna neden olan primer kesin tanısı netleştirildi ve bu karar altın standart karar olarak belirlendi. AI modellerinden de tüm veriler ışığında hastanın kesin tanısının belirlenmesi istendi. Her bir model tarafından üretilen ön tanı, ayırıcı tanı ve kesin tanı çıktıları altın standart karar ile karşılaştırıldı. İstatistiksel analizler, R programlama dili kullanılarak gerçekleştirildi.

Bulgular: Çalışma 180 hasta ile tamamlandı. Dahil edilen hastaların %56 sı erkek, %44'ü kadındı. Kesin tanı doğruluk oranları incelendiğinde, GPT-4o'nun doğruluk oranı %67,2 GPT-5'in %65,6, Claude'un %63,3 ve Gemini'nin %59,4 olarak saptandı. Gruplar arasında anlamlı istatistiksel fark saptanmadı (p=0.16). Altın standart tanı ile kesin tanı uyum düzeyleri değerlendirildiğinde, Cohen's κ katsayısı GPT-4o için 0,656, GPT-5 için 0,638, Claude için 0,616 ve Gemini için 0,575 olarak bulundu.

Sonuç: Yapay zekâ modellerinin kesin tanı doğruluk oranlarının %60-70 seviyelerinde olması, altın standart karar ile genel olarak iyi düzeyde uyum göstermesi bu sistemlerin acil servis ortamında klinik karar vermede yardımcı olabilme potansiyelini desteklemektedir.

Anahtar Kelimeler: Yapay zeka, Büyük dil modeli, Claude, Gemini, ChatGPT

ABSTRACT

PERFORMANCE OF ARTİFİCİAL İNTELLİGENCE İN CRİTİCAL PATİENT DİAGNOSİS: COMPARİSON OF CHATGPT, GEMİNİ, CLAUDE AND EMERGENCY PHYSİCİANS

Introduction: This study aimed to compare the diagnostic process of emergency physicians with the diagnostic outputs generated by ChatGPT, Claude, and Gemini in critically ill patients presenting to the emergency department(ED).

Methods: For critically ill patients presenting to the ED, physicians identified five preliminary diagnoses based on vital signs and medical history. The same anonymized data were entered into AI models, which were asked to generate five preliminary diagnoses. After physical examination, laboratory, and imaging results became available, both physicians and AI models refined these into three differential diagnoses. Finally, the physician's final diagnosis, explaining the patient's clinical condition, was accepted as the gold standard. Each AI model was then asked to determine the final diagnosis using the complete dataset. The preliminary, differential, and final diagnoses generated by each model were compared with the gold standard. Statistical analyses were performed using R.

Results:

The study included 180 patients (56% male, 44% female). Regarding the accuracy of final diagnoses, GPT-4o achieved an accuracy rate of 67.2%, GPT-5 achieved 65.6%, Claude achieved 63.3%, and Gemini achieved 59.4%. No statistically significant difference was found among the groups (p=0.16).When evaluating the agreement between the AI models' definitive diagnoses and the gold-standard diagnosis, Cohen's κ coefficients were calculated as 0.656 for GPT-4o, 0.638 for GPT-5, 0.616 for Claude, and 0.575 for Gemini.

Conclusion

The finding that AI models achieved diagnostic accuracy rates of 60%-70% and showed overall substantial agreement with the gold standard decision supports their potential to assist clinical decision-making in the ED setting.

Keywords: Artificial intelligence, ChatGPT, Claude, Gemini, Large language model

İÇİNDEKİLER

TABLOLAR DİZİNİ

ŞEKİLLER DİZİNİ

KISALTMALAR DİZİNİ

AI	**:**	Yapay zekâ (Artificial intelligence)
ANOVA	**:**	Varyans analizi (Analysis of variance)
AUC	**:**	Eğri altı alan (Area under curve)
AV	**:**	Atriovenrtiküler
BERT	**:**	Bidirectional Encoder Representations from Transformers
BT	**:**	Bilgisayarlı Tomografi
CDR	**:**	Clinical decision rules
CDSS	**:**	Clinical decision support system
CI	**:**	Güven aralığı (Confidence interval)
CNN	**:**	Konvolüsyonel sinir ağları
DL	**:**	Derin öğrenme
EKG	**:**	Elektrokardiyografi
GİS	**:**	Gastrointestinal sistem
GPT	**:**	Generative Pre-trained Transformer
HIPAA	**:**	ABD hasta veri gizliliği yasası (Health insurance portability and accountability act)
HVCAF	**:**	Hızlı ventrikül cevaplı atrial fibrilasyon
KBY	**:**	Kronik böbrek yetmezliği
KİBAS	**:**	Kafa içi basınç artışı sendromu
KKDS	**:**	Klinik karar destek sistemi
KKY	**:**	Konjestif kalp yetmezliği
KOAH	**:**	Kronik obstrüktif akciğer hastalığı
LLM	**:**	Büyük dil modeli (Large language model)
ML	**:**	Makine öğrenimi
MLP	**:**	Multilayer perceptron
MRG	**:**	Manyetik rezonans görüntüleme
n	**:**	Sayı
NLP	**:**	Doğal dil işleme (Natural language processing)
NSTEMİ	**:**	ST yükselmesiz miyokard infarktüsü
p	**:**	Olasılık değeri (p-value)

PET	**:**	Pozitron emisyon tomografisi
PTE	**:**	Pulmoner tromboemboli
Q-Q	**:**	Nicelik-Nicelik grafiği (Quantile-Quantile)
SPSS	**:**	Statistical package fort he social sciences
STEMİ	**:**	ST yükselmeli miyokard infarktüsü
SVO	**:**	Serebrovasküler olay
SVT	**:**	Supraventriküler taşikardi
VES	**:**	Ventriküler eskstrasistol
VT	**:**	Ventriküler taşikardi
YZ	**:**	Yapay zeka
XAI	**:**	Açıklanabilir yapay zeka
α	**:**	Anlamlılık düzeyi
%	**:**	Yüzde

1. GİRİŞ

Acil servis, tanı ve tedavi kararlarının zaman baskısı, klinik belirsizlik ve yüksek hasta yoğunluğu altında yürütüldüğü, sağlık hizmetlerinin en kritik alanlarından biridir. Özellikle kırmızı alan, yaşamı tehdit eden akut klinik durumların değerlendirildiği ve erken tanı ile hızlı müdahalenin hastanın prognozunu doğrudan etkilediği bir birimdir (1-3). Bu hasta grubunda tanısal gecikme; mortalite, morbidite ve kaynak kullanımında belirgin olumsuz sonuçlara yol açmaktadır.

Acil tıpta karar verme süreci çoğu zaman eksik veriyle, sınırlı değerlendirme süresi içinde gerçekleşir. Klinik yaklaşım, sezgisel örüntü tanıma (Sistem 1) ile analitik değerlendirme (Sistem 2) arasında dinamik bir denge üzerine kuruludur (4-6). Ancak zaman baskısı, dikkat bölünmesi ve bilişsel yük arttıkça karar süreci sezgisel düşünmeye daha fazla kayar; bu durum çapa etkisi, onaylayıcı önyargı ve aşırı güven gibi bilişsel sapmaların tanısal doğruluğu olumsuz etkilemesine zemin hazırlayabilir (7-9). Tanısal hataların mortalite artışıyla ilişkili olduğu ve hasta güvenliğini tehdit ettiği çeşitli çalışmalarda ortaya konmuştur (10-13). Bu nedenle acil serviste karar süreçlerini yalnızca bireysel deneyime bırakmak, kritik hastanın yönetiminde klinik risk oluşturmaktadır.

Son yıllarda büyük dil modelleri (LLM) olarak tanımlanan yapay zekâ sistemleri, klinik öykü, semptom kümeleri ve serbest metin verilerini analiz edebilme kapasiteleriyle dikkat çekmektedir (14-17). ChatGPT, Gemini ve Claude gibi modeller; anamnezi işleme, olası tanı önerileri oluşturma ve klinik düşünceyi yapılandırma süreçlerinde potansiyel bir karar destek aracı olarak değerlendirilmektedir. Bununla birlikte, bu modellerin acil tıp pratiğinde tanısal doğruluk, bağlam duyarlılığı, açıklanabilirlik ve hasta güvenliği açısından gerçek klinik koşullarda nasıl performans gösterdiği halen net değildir (18-21). Bu belirsizlik, yapay zekânın klinik karar süreçlerinde destekleyici bir rol mü yoksa yönlendirici bir mekanizma mı üstlenebileceği sorusunu önemli kılmaktadır.

Bu çalışmanın amacı, acil servise başvuran kırmızı triyaj kodlu hastalarda, primer hekimin tanısal sürecinin üç aşamasını (ön tanı, ayırıcı tanı ve kesin tanı) ChatGPT-4o, ChatGPT-5, Claude Pro Opus 4.1 ve Gemini Pro 2.5'in oluşturduğu

tanısal çıktılarla karşılaştırarak, ön tanı ve ayırıcı tanıda hekim çıktılarıyla eşleşme düzeyini, kesin tanıya ulaşmadaki performansını ve hekim ile uyumunu değerlendirmektir.

Bu çalışma, kritik hastada yapay zekâ uygulamalarının karar süreçlerinde tamamlayıcı bir araç olarak kullanılabilirliğine ilişkin değerlendirme yapmayı hedeflemektedir. Elde edilen bulguların, büyük dil modellerinin klinik tanı süreçlerinde hangi koşullarda ve ne ölçüde destekleyici rol üstlenebileceğine dair daha nesnel ve dengeli bir bakış açısı sağlayacağı öngörülmektedir.

2. GENEL BİLGİLER

2.1. Acil Serviste Kritik Hasta ve Tanısal Zorluklar

2.1.1. Kritik Hastanın Tanımı

Kritik hastalık kavramı, tıp literatüründe uzun süredir yer almakla birlikte, evrensel olarak kabul görmüş tekil bir tanıma halen ulaşılamamıştır. Bu tanımsal belirsizlik, özellikle farklı sağlık sistemlerinde ve klinik bağlamlarda kullanılan ölçütlerin çeşitliliğinden kaynaklanmaktadır (1, 2). Geçmişte yoğun bakım ihtiyacı üzerinden tanımlanmaya çalışılan bu kavram, zamanla yetersiz kalmış ve hastalığın fizyopatolojik özelliklerine ek olarak, müdahale edilebilirliğini ve sistem kaynaklarını da içerecek şekilde yeniden ele alınmıştır (2).

Güncel yaklaşım, kritik hastalığı üç temel kavramsal bileşen üzerinden açıklamaktadır: hayati organ disfonksiyonu, yüksek mortalite riski ve uygun tıbbi müdahaleyle bu durumun geri döndürülebilir olması (1). Ancak bu üçlü yapı, yalnızca hastanın klinik durumunu değil, aynı zamanda sağlık sisteminin müdahale kapasitesini de içerecek şekilde değerlendirilmelidir. Bu noktada kavramın sadece bireysel hastalık özelliklerine değil, aynı zamanda bağlama duyarlı olması gerektiği vurgulanmaktadır. Özellikle kaynak kısıtlı ülkelerde, aynı klinik tablo farklı şekilde sınıflandırılabilir; çünkü "kritik" olarak tanımlanma, çoğu zaman müdahale edilebilirlik düzeyiyle ilişkilidir (3).

Maslove ve arkadaşları, kritik hastalık kavramının sağlık hizmet organizasyonu ve kaynak planlaması açısından yeniden tanımlanmasının önemine dikkat çekerken; Kayambankadzanja ve ekibi, tanımın evrenselleştirilebilirliği için bağlamsal değişkenlerin tanıma entegre edilmesini önermektedir (1, 2). Kritik hastalığın sabit bir kategoriden ziyade, dinamik ve duruma göre değişkenlik gösteren bir klinik durum olduğu görüşü, bu çerçevede önem kazanmaktadır. Klinik gerçekliğin, sadece fizyolojik parametrelerle değil, sağlık sisteminin yapısal kapasitesiyle birlikte değerlendirilmesi gerekliliği açıktır.

Acil servis özelinde kritik hasta, genellikle ani gelişen organ yetmezliği, yaşamı tehdit eden fizyolojik bozulma ve hızlı müdahale gerektiren durumlarla

başvuran bireylerdir. Bu bağlamda kritiklik, yalnızca hastalığın ciddiyetiyle değil, zaman faktörü, ekip organizasyonu ve müdahale edilebilirlik düzeyiyle birlikte tanımlanır (3). Acil serviste karşılaşılan her hasta, aynı objektif bulgulara sahip olsa bile, sistem koşullarına göre farklı şekilde değerlendirilmekte ve önceliklendirme algoritmaları içinde konumlandırılmaktadır.

2.1.2. Acil Serviste Tanı Süreci ve Zorlukları

Acil servisler, tanı koyma sürecinin hem klinik belirsizlik hem de zaman baskısı altında yürütüldüğü, yüksek tempolu ve dinamik ortamlardır. Bu ortamda hekimler, sınırlı zaman dilimi içerisinde çok sayıda hastayı değerlendirme, triyaj yapma, doğru ve hızlı karar alma yükü altındadır. Tanı süreci genellikle kısa bir anamnez, hızlı fizik muayene, acil laboratuvar ve görüntüleme sonuçlarının yorumlanması gibi bir dizi adımı içerir. Ancak bu basamakların hiçbiri, ideal koşullarda çalışmaz. Hasta yoğunluğu, bilgiye erişim kısıtlılığı, iletişim zorlukları ve kaynak sınırlılıkları gibi faktörler, sürecin doğruluğunu doğrudan etkilemektedir (10, 11).

Zaman baskısı, acil servislerde tanısal karar sürecini belirleyici şekilde şekillendiren başlıca faktörlerden biridir. Hekimler çoğu zaman semptomların tüm yönlerini irdelemeye fırsat bulamadan, olası hayatı tehdit eden durumları dışlama odaklı kararlar vermek zorunda kalmaktadır. Bu süreçte kararlar, genellikle eksik bilgiyle ve sınırlı veri setine dayalı olarak alınır. Tudela ve arkadaşları, bu ortamda yapılan tanı hatalarının önemli bir bölümünün ya zamansal kısıtlar altında verilmiş erken kararlar ya da hasta hakkında yetersiz bilgiye dayanan varsayımsal çıkarımlardan kaynaklandığını ortaya koymuştur (11). Tanı hataları, sadece yanlış tanı konmasıyla sınırlı kalmamakta; aynı zamanda gecikmiş tanı ve hastanın durumunun yeterince değerlendirilmemesi gibi alt kategorilere de ayrılmaktadır (12).

Zaman baskısının tanı kalitesi üzerindeki olumsuz etkisi, hasta güvenliğini doğrudan tehdit eder. Hautz ve çalışma arkadaşlarının 2019 tarihli çok merkezli analizinde, acil serviste tanı hatası yapılan hastalarda hem hastanede kalış süresinin hem de mortalitenin anlamlı ölçüde arttığı gösterilmiştir (10). Özellikle non-spesifik semptomlarla başvuran ve ilk bakışta acil kabul edilmeyen hastalarda, zamanla yarışan tanı süreci çoğu zaman hastalığın seyrini kötüleştirebilmektedir. Bu durum, sadece

bireysel hasta sonuçlarını değil, sistem genelinde kaynakların verimsiz kullanımını da beraberinde getirmektedir.

Kelen ve arkadaşları, 2023 tarihli çalışmalarında, tanı hatalarının nedenlerine ilişkin yaygın kabul gören açıklamaların çoğunun klinik gerçeklikten kopuk olduğuna dikkat çekmiştir. Onlara göre, acil servislerdeki tanısal başarısızlıkların temelinde sistematik olarak yapılandırılmamış karar süreçleri, bilgi yükü fazlalığı ve bilişsel tükenmişlik gibi nedenler yatmaktadır (13). Bu nedenle tanı hatalarının yalnızca bireysel performansla açıklanması yerine, yapısal ve organizasyonel düzeyde değerlendirilmesi gerektiği önerilmektedir. Bu yaklaşım, hataların tekrar etmesini önleyici stratejilerin de daha etkili biçimde geliştirilebilmesine imkân tanır.

Acil servis ortamında tanı sürecinin doğasına dair bu tablo, karar destek sistemlerine olan ihtiyacı da açık biçimde ortaya koymaktadır. Hekimlerin zamanla yarıştığı, veri analizine yeterli süre ayıramadığı, nadir ya da atipik tablolarla sık karşılaştığı bu bağlamda, yapay zekâ tabanlı çözümler yalnızca teknolojik bir yenilik değil, aynı zamanda klinik güvenliği artırmaya yönelik bir zorunluluk olarak değerlendirilmelidir. Bu sistemlerin başarısı ise, gerçek dünya kliniklerinin karmaşıklığını ve aciliyetini içeren verilerle geliştirilmesine bağlıdır (10-13).

2.1.3. Tanı Hataları ve Sonuçları

Acil servis ortamı, yüksek hasta yoğunluğu, zaman kısıtı ve bilgi eksikliği gibi çoklu stresörlerin bir arada bulunduğu, karar verme açısından oldukça zorlu klinik alanlardan biridir. Bu bağlamda tanı koyma süreci, yalnızca klinik bilgiye değil, aynı zamanda karar vericinin zihinsel yüküne, sistemin işleyişine ve mevcut kaynaklara da bağlı olarak şekillenir. Bu değişkenler altında gelişen tanı hataları, tıbbın en ciddi hasta güvenliği sorunları arasında yer almakta ve hem bireysel hasta düzeyinde hem de sağlık sistemi genelinde geri döndürülemez sonuçlar doğurabilmektedir (22).

Klinik kararlar çoğu zaman, hekimin sınırlı bilgiyle kısa sürede karar vermesini gerektiren koşullarda alınır. Bu durumlarda karar süreci, daha çok hızlı, sezgisel ve otomatik düşünmeye dayanır. Bu düşünme tarzı, klinisyenin geçmiş deneyimlerine, benzer olgulara ve ilk izlenimlerine dayanarak hızla karar vermesini sağlar. Ancak bu hızlı karar alma süreci, sistematik analizden yoksun olduğu için çeşitli bilişsel önyargılara açık hale gelir (7). Özellikle acil servis koşullarında sık karşılaşılan bilişsel

önyargılar arasında aşırı güven (hekimin kendi yargılarına aşırı güven duyması), doğrulama önyargısı (ilk hipotezi destekleyen bilgileri seçici biçimde algılama), erişilebilirlik önyargısı (en kolay akla gelen tanıya yönelme) ve çapa etkisi (ilk izlenime saplanma ve alternatif tanıları dışlama) yer alır. Bu önyargılar, hastanın şikâyetlerinin yanlış yorumlanmasına, farklı tanı olasılıklarının göz ardı edilmesine ve tanının erken sonlandırılmasına neden olabilir (7).

Tanı hatalarının yalnızca bireysel bilişsel süreçlerden değil, aynı zamanda sağlık sistemine özgü yapısal sınırlılıklardan da kaynaklandığı açıktır. Acil servis pratiğinde sık karşılaşılan zaman kısıtı, yetersiz hasta öyküsü, tanısal testlere erişim güçlüğü, laboratuvar ve görüntüleme gecikmeleri ile ekip içi iletişim aksaklıkları, tanı sürecini doğrudan olumsuz etkilemektedir. Bu tür sistemik kısıtlar altında çalışan klinisyen, çoğu zaman sınırlı klinik veri ve artmış bilişsel yük altında hızlı karar vermek zorunda kalmakta; bu durum da yanlış, eksik ya da gecikmiş tanı riskini belirgin ölçüde artırmaktadır. Lin ve arkadaşlarının kapsamlı kohort analizinde, acil serviste potansiyel tanı hatası ile hastaneye yatırılan olguların 30 günlük mortalite oranlarında artış ve evde geçirilen sağlıklı gün sayısında anlamlı azalma saptanmıştır (22). Özellikle spinal apse (%15,6) ve menenjit (%10,9) gibi düşük prevalansa sahip ancak prognozu ağır hastalıkların tanısında hata oranlarının dikkat çekici düzeyde yüksek olduğu rapor edilmiştir.

Bu tanı hatalarının sonuçları yalnızca morbidite ve mortaliteyle sınırlı kalmaz; aynı zamanda sağlık sisteminde yatış sürelerinin uzaması, tekrar başvurular, gereksiz tetkikler ve artan maliyet gibi zincirleme etkiler yaratır. Acil serviste hatalı biçimde taburcu edilen hastalarda, daha sonra hastaneye yatış oranlarının ve mortalitenin anlamlı düzeyde yüksek olduğu gösterilmiştir (22). Bu durum, tanı sürecinin yalnızca bireysel kararlarla değil, organizasyonel yapıyla da doğrudan ilişkili olduğunu ortaya koymaktadır.

Bu çok boyutlu ve dinamik tanı ortamında, yapay zekâ ve makine öğrenimine dayalı karar destek sistemleri, klinik karar alma süreçlerini yapılandırmak ve tanı güvenilirliğini artırmak adına önemli bir potansiyel sunmaktadır. Brasen ve arkadaşlarının gerçekleştirdiği çok merkezli ve geniş örneklemli bir çalışmada, acil servise başvuran 9.190 hastanın girişimsel olmayan biyokimyasal parametreleri temel

alınarak geliştirilen 19 farklı makine öğrenimi algoritması, çeşitli klinik sonuçların öngörülmesinde yüksek doğruluk düzeylerine ulaşmıştır (23). Söz konusu algoritmalar; 30 günlük mortalite, yoğun bakım ihtiyacı ve güvenli taburculuk gibi prognostik açıdan kritik sonlanımlar için %87 ile %91 arasında değişen alan altı eğri (AUC) değerleriyle güçlü öngörü performansı sergilemiştir. Özellikle 30 günlük mortalite tahmini (AUC: %91,3) ve güvenli taburculuk (AUC: %87,3) algoritmalarının başarımı, bu sistemlerin erken risk sınıflaması ve kaynak yönetimi açısından klinik uygulamaya entegre edilebileceğini göstermektedir (23).

Ancak bu sistemlerin etkili biçimde kullanılabilmesi, yalnızca teknolojik doğruluğuna değil, klinik karar sürecine doğru entegrasyonuna da bağlıdır. Hekimlerin kararlarını tamamlayıcı şekilde yapılandırılmış, güvenilir ve yorumlanabilir yapay zekâ destekli sistemler; özellikle yüksek iş yükü altında çalışan acil servis hekimlerinin tanısal doğruluğunu artırabilir. Bu bağlamda, yapay zekâ uygulamaları yalnızca bireysel bilişsel önyargıların değil, aynı zamanda yapısal sistem sorunlarının da bir ölçüde önüne geçebilecek stratejik araçlar olarak değerlendirilmektedir.

2.1.4. Klinik Belirsizlikle Başa Çıkma ve Tanısal Karar Stratejileri

Acil servis, hızlı ve doğru karar almanın yaşamsal önem taşıdığı; aynı zamanda veri eksikliği, semptom karmaşası ve sınırlı kaynaklar nedeniyle belirsizliğin kaçınılmaz olduğu bir klinik ortamdır. Tanısal sürecin bu bağlamda şekillenmesi, hekimin yalnızca tıbbi bilgi düzeyine değil, aynı zamanda belirsizlikle başa çıkabilme becerilerine ve zihinsel kaynak yönetimine de bağlıdır. Klinik belirsizlik, yalnızca bilgi eksikliği değil, eldeki bilgilerin yorumlanmasında yaşanan güvensizlik ya da çelişki durumlarını da içerir. Bu ortamda alınan kararlar, çoğunlukla kesinlikten uzak, olasılıklar arasında yapılan seçimlere dayanır (24).

Zaman baskısı, bu belirsizlik ortamını daha da derinleştiren bir faktör olarak öne çıkar. Kısıtlı süre içerisinde karar vermeye zorlanan hekim, ayrıntılı değerlendirmeler yerine daha hızlı sonuçlara ulaşabileceği kestirme düşünme yollarına yönelebilir. Literatürde, bu durumun karar sürecinde veri toplamayı azaltarak hipotez çeşitliliğini sınırladığı ve tanısal değerlendirmeyi yüzeysel hâle getirebildiği belirtilmektedir (25). Zaman baskısının, özellikle klinik sunumu belirsiz veya atipik

olan olgularda, fark edilmeyen veya geç tanınan durumların artmasına zemin hazırladığı görülmektedir.

Zamanla sınırlı bu karar ortamında, hekimin karşı karşıya kaldığı en büyük zorluklardan biri de artan bilişsel yükle başa çıkabilmektir. Vardiya boyunca karşılaşılan çok sayıda hasta, sık görev geçişleri, dikkat bölünmeleri ve bilgiye ulaşımın parçalı olması, karar kalitesini tehdit eden zihinsel yükün başlıca kaynaklarıdır. Bilişsel yük kuramı çerçevesinde, bilgi işleme kapasitesi zorlandığında karar süreci daha yüzeysel hâle gelir; değerlendirme derinliği azalır ve alternatif hipotez üretimi sınırlanır (26). Bu durum, tanıların yeterince sorgulanmadan hızla sonuçlandırılmasına neden olabilir.

Klinik belirsizlik, zaman kısıtı ve bilişsel yükün kesişiminde şekillenen bu karar ortamı, bilişsel sapmaların tetiklenmesi açısından oldukça elverişlidir. Tanısal önyargılar, bu bağlamda bireysel dikkatsizlikten ziyade sistematik olarak ortaya çıkan, yinelenebilir zihinsel örüntülerdir. Bu önyargılar tamamen ortadan kaldırılamasa da, karar sürecine metabilişsel kontrol noktaları yerleştirilmesi ve yapılandırılmış sorgulama tekniklerinin uygulanmasıyla etkileri azaltılabilir (27). Özellikle yüksek belirsizlik altında çalışan klinisyenler için bu tür stratejiler, karar kalitesini korumada önemli araçlar sunar.

Son olarak, belirsizliğe verilen tepkinin hekimden hekime değiştiği, karar süreçlerinde bireysel farklılıkların önemli rol oynadığı bilinmektedir. Belirsizlikle yüksek düzeyde başa çıkabilen hekimler daha esnek kararlar alabilirken; düşük tolerans gösterenlerde erken karar verme, gereksiz tetkik isteme veya müdahale eğilimleri daha belirgin hâle gelebilir (24). Bu bireysel farklılıkların yalnızca bilişsel değil, aynı zamanda davranışsal sonuçları da bulunmaktadır. Bu nedenle, belirsizlikle başa çıkma kapasitesinin artırılması hem bireysel farkındalık çalışmalarıyla hem de organizasyonel destek yapılarıyla teşvik edilmelidir.

2.2. Klinik Karar Verme Süreci

Kritik hasta yönetimi, yalnızca doğru tanının konulmasını değil, aynı zamanda hızla değişen klinik tablolar karşısında zamanında, etkili ve bağlama duyarlı kararların alınmasını gerektirir. Bu süreç, hekimin klinik verileri yorumlamasını, belirsizlikle baş etmesini ve çoğu zaman sınırlı bilgiyle yüksek riskli seçimler yapmasını zorunlu kılar.

Özellikle acil servis pratiği, karar alma sürecinin yalnızca bilgiye değil; aynı zamanda dikkat, zaman yönetimi, deneyim ve bilişsel dayanıklılığa dayandığı bir ortam sunar. Platts-Mills ve arkadaşları, acil tıpta karar sürecinin doğasında yer alan belirsizliğin hekimler için bilişsel ve duygusal bir stres kaynağı oluşturduğunu, bu durumun klinik yaklaşımı doğrudan etkileyebildiğini ortaya koymaktadır (24).

2.2.1. Klinik Akıl Yürütme Modelleri

Klinik karar verme süreci, özellikle acil servis ortamında, yüksek zaman baskısı, bilgi eksikliği ve belirsizlik gibi etmenler altında yürütülen karmaşık bir bilişsel faaliyettir. Bu süreci açıklamak amacıyla geliştirilen ve günümüzde yaygın kabul gören yaklaşımlardan biri olan çift sistemli akıl yürütme kuramı (dual process theory), tanı koyma sürecinin birbirinden farklı ama etkileşimli iki bilişsel sistem aracılığıyla işlediğini öne sürmektedir. Norman ve arkadaşları, bu modeli tanısal uzmanlık gelişiminin bilişsel temellerini açıklamada temel bir çerçeve olarak sunmakta; birinci sistemin hızlı, sezgisel ve örüntü tanımaya dayalı; ikinci sistemin ise yavaş, analitik ve kurallara dayalı olduğunu belirtmektedir (4). Klinik uygulamada bu sistemler birbirine karşıt değil, aksine birbirini tamamlayıcı biçimde çalışmaktadır.

Sezgisel sistem (Sistem 1), özellikle deneyimle gelişen örüntü tanıma kapasitesi sayesinde, sık karşılaşılan veya tipik olguların tanınmasında önemli avantajlar sunar. Acil servis gibi yüksek tempolu ortamlarda bu sistemin hızı, klinik verimlilik açısından kritik önemdedir. Adams ve arkadaşlarının genç acil servis hekimleriyle yaptığı niteliksel çalışmada, klinik akıl yürütmenin genellikle hızlı, sezgisel kararlarla başladığı, ancak klinik belirsizlik ya da alışılmadık tabloyla karşılaşıldığında analitik sistemin devreye girdiği gösterilmiştir (5). Bu gözlem, klinik karar verme sürecinin doğrusal değil, duruma göre şekillenen dinamik bir yapı taşıdığını ortaya koymaktadır.

Analitik sistem (Sistem 2), özellikle sezgisel kararların yetersiz kaldığı ya da çelişkili olduğu durumlarda devreye girer. Bu sistem, hipotez üretme, alternatif tanıları değerlendirme ve sistematik düşünme gibi yürütücü işlevleri içerir. Ancak Al-Azri, acil servis ortamında analitik sistemin sınırlı sıklıkla kullanılabildiğini, çünkü zaman baskısı, dikkat bölünmeleri ve kognitif yük nedeniyle sezgisel sistemin daha baskın

olduğunu belirtmektedir (6). Bu durum, hızlı tanı koyma avantajı sunsa da özellikle atipik vakalarda tanı hatası riskini artırır.

Tanısal hata riski, yalnızca bireysel bilişsel kısıtlılıklardan değil; aynı zamanda bilgi temelli akıl yürütmenin yeterince gelişmemiş olmasından da kaynaklanmaktadır. Norman ve arkadaşları, her iki sistemin performansının da bilgi organizasyonuna dayandığını vurgulamakta; dolayısıyla sistemler arası dengesizlikten çok, bilgi temelli örüntülerin yetersizliğini temel problem olarak tanımlamaktadır (4). Bu nedenle tanı hatalarını azaltma girişimleri, yalnızca sezgisel süreçlerin bastırılması ya da analitik düşünmenin artırılmasıyla değil, klinik bilgi altyapısının güçlendirilmesiyle anlam kazanmaktadır.

Bu bağlamda, eğitim stratejilerinin odağı da yeniden düşünülmelidir. Pelaccia ve arkadaşları, bilişsel hata teorilerini esas alan geleneksel öğretim yaklaşımlarının sınırlı etki sağladığını, bunun yerine örüntü tanıma becerisini ve bilgi temelli akıl yürütmeyi geliştiren yöntemlerin daha etkili olduğunu savunmaktadır (28). Karma vaka simülasyonları, yapılandırılmış geribildirim oturumları ve metabilişsel farkındalık çalışmaları hem sezgisel hem de analitik sistemlerin birlikte etkin kullanılmasına olanak sağlayan eğitim araçları arasında yer almaktadır. Bu yaklaşım, karar verme sürecinde yalnızca hata oranlarını azaltmayı değil, aynı zamanda uzmanlık gelişimini de desteklemeyi amaçlamaktadır.

2.2.2. Bilişsel Önyargılar ve Hata Kaynakları

Klinik karar sürecinde meydana gelen tanı hatalarının önemli bir kısmı, fark edilmeden işleyen bilişsel önyargıların sonucudur. Bu önyargılar, hekimin değerlendirme sürecini kısıtlayan sistematik düşünme sapmalarıdır ve çoğunlukla karar zincirinin erken evrelerinde şekillenerek tüm tanısal yaklaşımı etkileyebilir. Özellikle yüksek hasta yoğunluğu, bilgi sınırlılığı ve zaman kısıtı gibi durumlarda, hekimler belirli tanı hipotezlerine zihinsel olarak erken bağlanma eğilimi gösterir. Hansen, bu durumun tanıya ilişkin değerlendirme süreçlerini daralttığını ve alternatif olasılıkların ihmal edilmesine yol açtığını vurgular (8).

Bilişsel önyargılar çoğu zaman örtük biçimde ortaya çıkar; hekim farkında olmadan dikkatini belirli verilere yoğunlaştırır, diğerlerini dışlar. Örneğin, çapa etkisi (anchoring), ilk alınan bilgiye gereğinden fazla ağırlık verilmesine neden olurken;

onaylayıcı önyargı (confirmation bias), mevcut tanı hipotezine uymayan verilerin göz ardı edilmesine yol açar. Kunitomo ve arkadaşlarının Japonya'da gerçekleştirdiği çok merkezli analiz, tanı hatası yapılan olguların %96'sında en az bir bilişsel önyargının etkili olduğunu göstermiştir. Özellikle gastrointestinal, kardiyovasküler ve nörolojik sistem hastalıklarında bu hata türleri daha sık görülmektedir (7).

Tanısal önyargılar yalnızca bireysel karar süreçlerinde değil, aynı zamanda ekip içi iletişimde de etkili olur. Ng ve arkadaşları, acil serviste hasta devri (handover) sırasında tanı kararlarının nadiren yeniden değerlendirildiğini; önceki hekimin oluşturduğu tanısal çerçevenin çoğunlukla sorgulanmaksızın kabul edildiğini belirtmektedir. Bu durum, "diagnostic momentum" adı verilen bir bilişsel örüntüye neden olur. Tanı devralan hekim, önceki değerlendirmeye bağlı kalarak yeni semptomları farklı bir perspektiften yorumlama fırsatını kaybeder (9).

Bilişsel önyargıların bazıları klinik deneyime rağmen azalmaz; hatta deneyimli hekimlerde daha da sabitlenmiş örüntüler hâlinde devam edebilir. Özellikle aşırı güven (overconfidence), hekimin kendi yargısına olan güveni nedeniyle dış görüş alma veya tanı hipotezlerini sorgulama eğilimini zayıflatabilir. Bu tür önyargılar, vaka karmaşıklığının yüksek olduğu durumlarda daha belirgin riskler oluşturur. Örnekleme yanlılığı (availability bias) da sık karşılaşılan bir diğer sapmadır; hekim yakın zamanda yaşadığı dikkat çekici bir vakaya zihinsel olarak yönelerek, mevcut durumu objektif biçimde değerlendiremeyebilir (8).

Bu tür bilişsel sapmaların etkisini azaltmaya yönelik önerilen stratejiler arasında, bireysel farkındalık çalışmaları kadar, iletişim protokollerinin yapılandırılması da yer alır. Ng ve arkadaşları, hasta devri gibi kritik karar noktalarında sistematik kontrol listelerinin ve ikinci görüş mekanizmalarının tanı kararlarını daha sağlam temellere oturtabileceğini savunmaktadır (9). Bu öneriler, bilişsel önyargıların tamamen ortadan kaldırılamayacağını kabul etmekle birlikte, karar sürecine yapısal denetim noktaları yerleştirmenin hata riskini azaltabileceğini göstermektedir.

2.2.3. Acil Tıpta Karar Verme Dinamikleri

Acil tıpta karar verme süreçleri, klasik tanısal algoritmalardan çok, bağlama duyarlı, çok katmanlı bir bilişsel organizasyonu yansıtır. Bu süreç yalnızca hasta özelliklerine değil; aynı zamanda olayın aciliyeti, karar anındaki belirsizlik düzeyi,

klinik ortamın baskısı ve hekimin deneyim profiline göre şekillenir. Kovacs ve Croskerry, acil tıpta karar almanın salt bilgiye dayalı bir analiz değil; aynı zamanda zaman, risk ve sonuçlar arasında hızlı denge kurma becerisi olduğunu vurgular (29). Bu yönüyle karar verme, her klinik temasın teknik ve durumsal bağlamına gömülü bir süreç hâline gelir.

Triyaj, bu karar ortamının en kritik ve hızlı geri bildirim gerektiren aşamalarından biridir. Egoda Kapuralalage ve arkadaşları, triyaj kararlarının sıklıkla sınırlı bilgiyle ve anlık izlenime dayalı olarak verildiğini, bu nedenle bilişsel önyargıların daha kolay tetiklendiğini belirtmektedir. Özellikle "öncelik seviyesi" belirleme süreci, vaka yoğunluğu arttıkça sezgisel karar kalıplarına daha fazla bağımlı hâle gelir. Çalışma, doğru kategorilendirilemeyen hastaların daha geç müdahale aldığı ve bu durumun klinik sonuçlar üzerinde doğrudan etkili olduğunu ortaya koymaktadır (30).

Kritik olaylara yönelik karar verme ise triyajdan farklı olarak, daha geniş bilgi ve zaman penceresi gerektirir; ancak kararın potansiyel sonucu daha ciddidir. Reale ve arkadaşları, yüksek riskli olaylarda karar vericilerin sadece klinik bilgiye değil, aynı zamanda olayla ilişkili stres düzeyi, olay tipi (örn. kardiyak arrest vs. travma) ve kararın doğuracağı olası sonuçların ağırlığına göre pozisyon aldığını göstermiştir. Bu tür durumlarda, karar kalitesini belirleyen temel unsur çoğu zaman bilgi miktarından çok, olayın zihinsel temsili ve hekimin içsel kararlılık düzeyidir (31).

Bu zorlu karar ortamında klinik karar kuralları (clinical decision rules – CDR), bilgiye dayalı standartlaştırılmış destek sağlayarak sübjektifliğin etkisini azaltmayı hedefler. Ancak Chan ve arkadaşları, CDR'lerin etkin kullanımının yalnızca algoritmanın doğruluğuna değil, aynı zamanda hekimin o anda kurala ne kadar güvendiğine ve klinik bağlamın CDR'yi ne ölçüde taşıyabildiğine bağlı olduğunu belirtmektedir. Bu bağlamda, klinik yargı ile algoritmik önerilerin dengeli entegrasyonu ne salt protokollere bağımlılık ne de kuralsız sezgiyle ilerleme anlamına gelir; karar verme becerisinin bilgi, bağlam ve güven dengesiyle yönetilmesi gerekir (32).

2.2.4. Klinik Karar Destek Sistemlerine Duyulan İhtiyaç

Acil servis ortamı, klinik karar verme süreçlerinin en kırılgan olduğu alanlardan biridir. Hekimler, sınırlı bilgiyle, yüksek hasta akışı ve zaman baskısı altında hızlı ve doğru kararlar vermekle yükümlüdür. Bu koşullarda, erken kötüleşme riski taşıyan hastaların ayırt edilmesi, klinik öngörünün sınırlarını zorlayan bir beceriye dönüşmektedir. Karar verme süreci yalnızca klinik bilgiye değil; aynı zamanda dikkat, hafıza ve çevresel koşullar gibi değişkenlere de bağlıdır. Bu değişkenliğin yol açtığı belirsizlik ortamı, hata riskini artırmakta; aynı zamanda tutarlı ve sürdürülebilir karar almayı güçleştirmektedir (33).

Klinik karar destek sistemleri (KKDS), bu karmaşık karar ortamına istikrar kazandırma potansiyeline sahip teknolojik araçlardır. Özellikle acil serviste triyaj, erken kötüleşme öngörüsü ve kaynak önceliklendirmesi gibi yüksek bilişsel yük oluşturan noktalarda, karar süreçlerini standartlaştırıcı etkileri ön plana çıkmaktadır. Gelişmiş makine öğrenimi modelleri, klinik ve fizyolojik verilerden risk tahmini yaparak müdahale gereksinimi yüksek olan hastaları erken dönemde belirlemeye katkı sağlayabilmektedir (34). Bu yaklaşım, özellikle deneyim düzeyi değişken olan hekim grupları için klinik dengeyi sağlayıcı bir rol üstlenir.

Ancak yalnızca öngörü doğruluğu değil, sistemin klinik iş akışına entegrasyonu ve kullanıcı tarafından kabul edilebilirliği de belirleyici faktörlerdir. KKDS'lerin etkili olabilmesi için önerilerinin zamanında, bağlamsal olarak uygun ve güvenilir olması gerekir. Kullanıcılar açısından bu sistemlerin “yardımcı” mı yoksa “yönlendirici” mi olduğu sorusu, benimsenme düzeyini doğrudan etkiler. Ayrıca, hasta başı karar alma hızını yavaşlatmadan çalışabilen, uyarı yorgunluğu yaratmayan ve hekimin mesleki yargısını destekleyen sistemler daha yüksek kabul görmektedir (35).

Klinik karar destek sistemlerine olan ihtiyaç, yalnızca teknolojik bir gelişme beklentisi değil; mevcut karar ortamının yapısal sınırlılıklarına verilen stratejik bir yanıttır. Erken tanı, hasta güvenliği ve operasyonel verimlilik gibi hedefler doğrultusunda, KKDS'ler acil servisin dinamik yapısına entegre edilebildiğinde hem klinik kalitenin hem de karar sürecinin öngörülebilirliğinin artırılması mümkün hâle gelmektedir (36).

2.3. Yapay Zekâ Kavramı ve Tıpta Kullanım Alanları

Modern tıpta karar verme süreçleri, artan veri yoğunluğu ve klinik karmaşıklık nedeniyle geleneksel yöntemlerin ötesinde çözümlere ihtiyaç duymaktadır. Bu bağlamda yapay zekâ, büyük ve çok boyutlu sağlık verilerini anlamlandırarak tanı, tedavi ve izlem süreçlerinde hekime bilişsel destek sunan güçlü bir araç hâline gelmiştir. Yalnızca tanısal doğruluğu artırmakla kalmayıp; klinik yükü dengeleyen, erken kötüleşmeyi öngörmeye yardımcı olan ve karar sürekliliğini sağlayan mekanizmaların temelinde yer almaktadır. Günümüzde kullanılan yapay zekâ yaklaşımları, öğrenmeye dayalı algoritmalardan tıbbi görüntü analizine, doğal dil işleme tekniklerinden tahmine dayalı modellemeye kadar uzanan geniş bir uygulama alanına sahiptir. Bu teknolojik dönüşüm, hasta güvenliği ve hizmet kalitesinde iyileşme sağlarken, hekimin karar verme kapasitesini de belirleyici noktalarda stratejik olarak desteklemektedir.

2.3.1. Yapay Zekânın Tanımı ve Tarihsel Gelişimi

Yapay zekâ (YZ), insan zekâsına özgü bilişsel süreçlerin —örneğin öğrenme, akıl yürütme, karar verme ve problem çözme— dijital sistemler aracılığıyla simülasyonunu ifade eder. Bu kavram ilk kez 1956 yılında Dartmouth Konferansı'nda akademik bir alan olarak tanımlanmış; ancak kökeni, Alan Turing'in 1950 yılında ileri sürdüğü "Bir makine düşünebilir mi?" sorusuna kadar uzanır. Bu tarihsel başlangıç, YZ'nin yalnızca hesaplamaya değil, aynı zamanda düşünme biçimlerinin modellenmesine de dayandığını göstermektedir (37).

YZ'nin tıpta erken dönem kullanım örneklerinden biri, 1970'lerde geliştirilen kural tabanlı uzman sistem MYCIN'dir. Adını *mycin* son eki taşıyan antibiyotiklerden esinlenerek alan bu sistem, "Mycology + Medicine" kavramlarını temsil eden sembolik bir ad taşır. MYCIN, bakteriyel enfeksiyonları tanımlamak ve uygun antibiyotik tedavisini önermek amacıyla geliştirilmiş, mantıksal çıkarıma dayalı bir danışma sistemiydi. Kan ve beyin omurilik sıvısı kültürlerinden gelen verileri değerlendirerek, sabit kurallar aracılığıyla önerilerde bulunabiliyordu. Ancak sabit yapısı, değişken klinik senaryolara uyum sağlama kapasitesini sınırlamıştı (37). Bu sınırlılıklar, ilerleyen yıllarda makine öğrenimi ve derin öğrenme gibi veri temelli yaklaşımlarla aşılmıştır. Günümüzde YZ, yalnızca tanı doğruluğunu artırmakla

kalmayıp, hasta izleminde, tedavi planlamasında ve karar destek sistemlerinde çok boyutlu katkılar sunarak sağlık hizmetlerinin dönüşümünde merkezi bir rol üstlenmektedir (38).

2.3.2. Makine Öğrenimi, Derin Öğrenme ve Doğal Dil İşleme (NLP) Kavramları

Makine öğrenimi (ML), dijital sistemlerin açık kurallarla programlanmadan veriler üzerinden örüntü tanıma ve öğrenme yeteneği kazanmasını sağlayan matematiksel bir yaklaşımdır. Klinik pratikte, özellikle kritik hastaların tanısında kullanılan ML modelleri; yaş, vital bulgular, laboratuvar değerleri ve semptom kümeleri gibi çok sayıda değişkeni bir arada işleyerek yüksek riskli durumların daha erken fark edilmesine olanak tanır (39).

Derin öğrenme (DL), ML'nin çok katmanlı yapay sinir ağları yapısına dayanan gelişmiş bir formudur. Bu yapı sayesinde DL, kompleks ilişkiler içeren büyük veri setlerini analiz edebilir ve örüntü tanıma kapasitesiyle geleneksel yöntemlerin ötesinde doğruluk sağlayabilir. Kritik hasta değerlendirmesinde, DL modelleri klinik belirti kümelerini çözümleyerek, bozulma eğilimi taşıyan hastaları erken dönemde ayırt etmede güçlü performans sergilemektedir (40). Özellikle klinik verilerin zamana duyarlı, eksik ve çok boyutlu olduğu acil servis ortamlarında DL uygulamaları, sezgisel yargının ötesine geçebilecek analitik destek sunar (39).

Doğal dil işleme (NLP), sağlık kayıtlarında serbest metin şeklinde yer alan tanımlayıcı verilerin (örneğin klinik öykü, hekim gözlemleri, konsültasyon notları) analiz edilmesini mümkün kılar. NLP sistemleri, bu yapılandırılmamış verileri sınıflandırarak, kritik hasta tanısı açısından değerli olabilecek semptom dizilimlerini, zaman içindeki değişimleri ve hastalığa özgü anlatımları tanımlayabilir (41). NLP, özellikle DL ile entegre kullanıldığında, hasta ile ilgili çok yönlü bilgi akışını sayısallaştırarak klinik karar desteğini güçlendirir (40).

Bu üç yöntem, yapay zekâ temelli sistemlerin yalnızca klinik veriyi işlemesine değil, aynı zamanda karmaşık tanı problemlerine karşı çok boyutlu ve dinamik çözümler üretebilmesine olanak tanır. Kritik hastalıkların tanısında insan yargısının ötesine geçen erken uyarı sistemleri oluşturulabilmesi, ML, DL ve NLP'nin birlikte çalıştığı bu bilişsel çerçeveyle mümkün olmaktadır (42).

2.3.3. Tıpta Yapay Zekâ Uygulamaları

YZ, modern tıbbın dönüşümünde giderek daha belirleyici bir rol oynamaktadır. Özellikle klinik karar süreçlerinin karmaşıklığı, veri hacmindeki artış ve hızlı müdahale gerektiren durumlar, YZ tabanlı teknolojilerin sağlık alanındaki uygulama potansiyelini artırmıştır. Günümüzde bu uygulamalar, üç ana eksende incelenebilir: tıbbi görüntü analizi, tahmine dayalı modelleme ve klinik karar destek sistemleri.

- **Görüntü Analizi**

Tıbbi görüntüleme, YZ uygulamalarının en olgunlaşmış ve etkili olduğu alanlardan biridir. Bilgisayarlı tomografi (BT), manyetik rezonans görüntüleme (MRG) ve pozitif emisyon tomografisi (PET) gibi yüksek çözünürlüklü tekniklerle elde edilen görüntülerin analizinde, özellikle derin öğrenme algoritmaları etkili biçimde kullanılmaktadır. Konvolüsyonel sinir ağları (CNN), görsel verilerdeki karmaşık yapıları tespit etme becerisi sayesinde, tümör, damar patolojileri veya retinopati gibi durumların erken evrede saptanmasına olanak tanımaktadır (43). Bu sistemler, yalnızca anormallikleri saptamakla kalmayıp, aynı zamanda tedavi planlarının kişiselleştirilmesine de katkı sağlar. Görüntü analizinde açıklanabilir yapay zekâ (explainable AI – XAI) modelleri, klinisyenlerin algoritmik kararların altında yatan nedenleri anlamasına imkân tanıyarak sistemlere olan güveni artırmaktadır (44, 45).

- **Tahmine Dayalı Modelleme**

YZ'nin en güçlü yönlerinden biri, klinik sonuçların öngörülmesine yönelik modellemeler geliştirebilmesidir. Bu modeller, elektronik sağlık kayıtları, vital bulgular ve laboratuvar değerleri gibi çok boyutlu verileri analiz ederek, sepsis, yoğun bakım ihtiyacı veya kısa vadeli mortalite gibi olumsuz sonuçların erken tahminini mümkün kılar. Yapay sinir ağları, özellikle multilayer perceptron (MLP) gibi derin öğrenme yapıları, bu öngörü modellerinde sıkça tercih edilmektedir. Acil servise hiperglisemik kriz ile başvuran hastalar üzerinde yapılan çalışmalarda, MLP tabanlı modellerin sepsis ve tüm nedenlere bağlı 30 günlük mortalite tahmininde oldukça başarılı sonuçlar verdiği gösterilmiştir (46). Bu öngörücü sistemler, klinik kaynakların

doğru yönlendirilmesine katkı sağlayarak mortalite ve morbidite oranlarının azaltılmasına yardımcı olabilir.

- **Klinik Karar Destek Sistemleri**

YZ ile entegre çalışan klinik karar destek sistemleri (Clinical Decision Support Systems – CDSS), hasta özelinde önerilerde bulunarak tanı doğruluğunu artırma, tedaviye uyumu kolaylaştırma ve sağlık hizmeti kalitesini yükseltme amacı taşır. Bu sistemler, yapılandırılmış veriler kadar serbest metinlerden de bilgi çekebilen NLP tekniklerinden faydalanmaktadır. NLP, hekim notları, epikrizler ve diğer metinsel kaynaklardan anlamlı içerik çıkartarak karar süreçlerini zenginleştirmektedir (44). CDSS uygulamaları ayrıca, erken uyarı sistemlerinin entegrasyonu ile hastalıkların preklinik evrede tanınmasını ve müdahalenin zamanında yapılmasını sağlar. Bunun ötesinde, kişiselleştirilmiş tedavi önerileri sunarak klinik kararlara özgüven kazandırır ve heterojen hasta gruplarında bireyselleştirilmiş bakım sağlar (44).

YZ destekli CDSS sistemlerinin sağlık alanında yaygınlaşmasında karşılaşılan temel zorluklar arasında algoritmaların açıklanabilirliği, veri gizliliği ve sistemlerin klinik iş akışlarına entegrasyonu yer almaktadır. Derin öğrenme tabanlı birçok model, "kara kutu" yapısında olduğundan dolayı, karar mekanizmalarının klinisyenlerce anlaşılması güçleşebilir. Bu nedenle, açıklana bilirlik sağlayan sistemler geliştirilmesi önem kazanmaktadır (44, 47).

Genel olarak değerlendirildiğinde, YZ'nin sağlık hizmetlerinde sunduğu katkılar; tanı süreçlerinin hızlanması, risk öngörülerinin hassaslaşması ve tedavi planlarının kişiselleştirilmesi üzerinden şekillenmektedir. Ancak bu dönüşüm, teknolojik inovasyonların ötesinde, etik, yasal ve sistemsel düzeyde bütüncül bir adaptasyonu da zorunlu kılmaktadır.

2.4. Büyük Dil Modelleri: ChatGPT, Gemini, Claude

Klinik karar alma süreçlerinde artan bilişsel yük, zaman baskısı ve bilgiye anında erişim ihtiyacı, geleneksel bilgi sistemlerinin ötesine geçen çözümleri zorunlu hale getirmiştir. Bu bağlamda, büyük dil modelleri (Large Language Models, LLM); dilin yapısını yalnızca sözdizimsel düzeyde değil, bağlamsal, kavramsal ve hatta klinik sezgi düzeyinde işleyebilen yeni bir yapay zekâ kuşağını temsil eder (14). Bu modeller,

klasik algoritmalardan farklı olarak, eğitim sürecinde insan diline ait örüntüleri devasa veri setlerinden öğrenir; bu sayede soru yanıtlama, özet çıkarma ve klinik öneri üretme gibi görevlerde yüksek derecede esnek ve içerik odaklı yanıtlar üretebilir (15).

Özellikle tıbbi uygulamalarda öne çıkan ChatGPT (OpenAI), Gemini (Google DeepMind) ve Claude (Anthropic) gibi modeller, yalnızca teknik yapı bakımından değil, aynı zamanda bilgi güvenilirliği, açıklanabilirlik ve etik hassasiyet gibi yönlerden de farklılaşmaktadır. Bu modeller, hekimlerin kararlarını birebir değiştirme iddiasında değildir; ancak karar süreçlerini destekleyici, hata riskini azaltıcı ve bilişsel yükü dengeleyici işlevleriyle dikkate değerdir. Dolayısıyla, her bir modelin mimari yapısının ötesinde, klinik güvenlik ve işlevsel uygulanabilirlik bakımından değerlendirilmesi gerekmektedir.

2.4.1. Büyük Dil Modellerinin Yapısı ve Çalışma Prensibi

LLM, doğal dil işleme (NLP) alanında geliştirilen ve metni anlama, üretme, özetleme gibi görevlerde yüksek başarı gösteren derin öğrenme tabanlı yapılardır. LLM'ler, çok büyük hacimli metin verileri üzerinde eğitilerek sözcükler arasındaki bağlam ilişkilerini istatistiksel olarak öğrenir ve bu bilgi üzerinden yanıt üretir. Bu modellerin temelini, "Transformer" adı verilen bir mimari oluşturur. Transformer yapısı, girdi verisini paralel olarak işleyebilme ve uzun metinlerdeki anlam ilişkilerini yakalayabilme yeteneği ile önceki geleneksel dil modelleme yaklaşımlarına kıyasla önemli avantajlar sağlamaktadır (16).

LLM'ler mimari olarak iki temel gruba ayrılır: "Encoder tabanlı" ve "Decoder tabanlı" yapılar. Encoder tabanlı modeller, metni iki yönlü okuyarak anlamlandırır ve özellikle sınıflandırma, bilgi çıkarımı gibi görevlerde kullanılır. Bu gruba ait en bilinen modellerden biri BERT (Bidirectional Encoder Representations from Transformers)'tir. Buna karşılık, decoder tabanlı modeller girdiyi tek yönlü işler ve yeni metin üretme gibi görevlerde daha etkilidir; GPT (Generative Pre-trained Transformer) bu yapının temsilcisidir. Sağlık alanındaki uygulamalarda bu iki yapı farklı amaçlara hizmet eder: bilgi çıkarımı ve özetleme gibi görevlerde encoder modeller tercih edilirken, klinik not yazımı ve diyalog simülasyonu gibi üretken görevlerde decoder modeller kullanılmaktadır (16, 17).

Modelin eğitiminde kullanılan veri kümesinin içeriği, performansı doğrudan etkileyen faktörlerden biridir. Genel amaçlı LLM'ler, geniş internet metinleriyle eğitildiklerinden sağlık terimlerini sınırlı kapsayabilir. Bu nedenle, klinik veri ile yeniden eğitilmiş modeller geliştirilmiştir. Örneğin, ClinicalBERT ve Med-PaLM gibi yapılar, tıbbi literatür ve hasta kayıtları üzerinden eğitilerek klinik dilin özgünlüğünü daha iyi yansıtan sonuçlar üretmektedir. Bu modeller, karar destek sistemlerinde, risk sınıflamasında veya tanı önerilerinde daha isabetli çıktılar verebilmektedir (16, 17).

LLM'lerin çalışma prensibi, girdiye karşılık en olası sözcük dizisini tahmin etmek üzerine kuruludur. Bu süreçte modelin önceki eğitim evresinde öğrenmiş olduğu dil kalıpları ve kavramsal bağlamlar kullanılır. Dikkat mekanizması sayesinde, model hangi kelime ya da kavramlara öncelik vereceğini bağlama göre belirleyebilir. Bu yapılar, uygun şekilde eğitildiklerinde klinik sorulara yanıt verme, literatür tarama, hasta geçmişi analizi gibi görevlerde anlamlı katkılar sunabilmektedir.

2.4.2. ChatGPT: Mimari ve Klinik Kullanım Potansiyeli

ChatGPT, OpenAI tarafından geliştirilen ve LLM ailesine mensup olan üretken yapay zekâ sistemidir. Temelini oluşturan mimari, dil verilerini çok katmanlı olarak işleyebilen "Transformer" yapısına dayanır. Bu yapı, girdideki kelimeler arasındaki anlam ilişkilerini eşzamanlı analiz eden dikkat mekanizması sayesinde bağlamsal olarak tutarlı yanıtlar üretebilmektedir. ChatGPT, GPT-3.5 ve GPT-4 gibi ardışık model varyantlarıyla, milyarlarca parametre içeren geniş bir yapıda eğitilmiş olup, çok çeşitli dil görevlerinde yüksek doğrulukla yanıt üretme kapasitesine sahiptir (48).

Sağlık hizmetlerinde ChatGPT, klinik karar destek sistemlerinden hasta bilgilendirmeye, epikriz yazımı ve eğitim materyali üretiminden triyaj ön değerlendirmelerine kadar geniş bir yelpazede potansiyel uygulama alanına sahiptir. Klinik karar desteğinde, mevcut semptomlara dayalı olası tanı alternatiflerini sunabilir; bu yönüyle özellikle yoğun veri akışı altındaki acil servislerde ikinci görüş desteği sağlayabilir. Raporlama süreçlerinde, kısa sürede yapılandırılmış bilgi üretme kabiliyetiyle hekimlerin idari yükünü azaltabilir. Ayrıca, hasta eğitimi bağlamında tıbbi bilgiyi sadeleştirerek bireylerin anlayabileceği şekilde sunma yetisi, sağlık okuryazarlığını artırma açısından önemlidir (49).

ChatGPT’nin öne çıkan avantajları arasında çok hızlı yanıt üretme kapasitesi, geniş içerik kapsamı, doğal dilde tutarlı diyalog yürütme becerisi ve çok sayıda dilde çalışabilme yetisi sayılabilir. Buna karşın bazı önemli sınırlılıkları da bulunmaktadır. En kritik sorunlardan biri, modelin gerçekte var olmayan veya doğrulanmamış bilgiler üretmesi olarak tanımlanan “halüsinasyon” problemidir. Ayrıca, tıbbi karar destek sistemlerinde kullanımı sırasında hatalı önerilerde bulunabilme riski, etik sorumlulukların netleşmemiş olması ve hasta güvenliği açısından yorumlanabilirlik eksikliği, güvenli entegrasyonun önündeki başlıca engellerdendir (50).

2.4.3. Gemini: Google altyapısı ve özellikleri

Gemini, Google DeepMind tarafından geliştirilen ve PaLM 2 mimarisi üzerine inşa edilen bir LLM’dir. Çok modlu veri işleme kapasitesiyle tanımlanan bu model, metinlerin yanı sıra görsel ve kod temelli girdileri de işleyebilecek şekilde tasarlanmıştır (51). Yapılandırma açısından değerlendirildiğinde, sağlık hizmetlerinde kullanılan diğer modellerle karşılaştırıldığında Gemini'nin farkı, çoklu veri türlerini birlikte analiz edebilme potansiyelinde yatmaktadır. Ancak bu potansiyelin pratikte ne ölçüde işlevsel olduğu, modelin kullanım alanına göre değişkenlik gösterebilmektedir.

Tıbbi bilgi üretimi, metin anlama ve referanslı yanıt oluşturma açısından yapılan karşılaştırmalı değerlendirmelerde, Gemini’nin klinik bağlamı kavrama düzeyi yüksek olmakla birlikte, otomatik referans oluşturma sürecinde hatalı ya da sahte kaynaklar üretme olasılığı bulunmaktadır (51). Bu nedenle oluşturulan çıktılar, doğruluk açısından insan denetimine ihtiyaç duymaktadır. Klinik sorulara verilen yanıtların güvenilirliği Gemini Advanced sürümünde artmakla birlikte, bu sürümde dahi tüm yanıtlar aynı düzeyde tutarlılık göstermemektedir.

Oftalmoloji alanında yapılan değerlendirmelerde, Gemini'nin özellikle bilgi düzeyi ve açıklayıcı anlatım açısından bazı başlıklarda ChatGPT-4 ile karşılaştırılabilir yanıtlar ürettiği, ancak karmaşık karar senaryolarında zaman zaman yetersiz kaldığı bildirilmiştir (52). Görsel içeriğe dayalı sorularda ise performansın belirgin şekilde düştüğü, bazı durumlarda yanıtların bağlamsal bütünlükten uzaklaştığı görülmüştür. Bu durum, modelin teorik çok modlu yapısına rağmen, görsel analiz alanında sınırlılıklarının bulunduğunu ortaya koymaktadır. Benzer şekilde ortopedi alanında, glukokortikoid ilişkili osteoporoz konulu yapılandırılmış değerlendirmelerde

Gemini'nin performansı farklı alt başlıklarda değişkenlik göstermiştir. Özellikle terminolojiye dayalı bilgi çekme sorularında tatmin edici yanıtlar üretse de klinik yönlendirme içeren karmaşık senaryolarda yanıtların eksik, yüzeysel veya bağlam dışı olabildiği bildirilmiştir (53).

Genel olarak Gemini, çoklu veri tiplerini işleyebilme ve tıbbi metinlere yanıt üretebilme kapasitesi açısından dikkat çekici özelliklere sahip olsa da sağlık alanında doğrudan uygulamaya geçmeden önce doğruluk, kaynak güvenilirliği ve bağlam duyarlılığı açısından kapsamlı şekilde değerlendirilmelidir.

2.4.4. Claude: Antropik Yaklaşımlar ve Tıbbi Güvenlik

Claude, Anthropic tarafından geliştirilen ve güvenlik öncelikli tasarımıyla dikkat çeken bir büyük dil modeli ailesidir. Claude 3 Opus ve Claude 3.5 Sonnet sürümleriyle temsil edilen bu yapı, yalnızca yanıt üretme yetkinliğiyle değil; aynı zamanda etik güvenlik ilkeleri, kullanıcı mahremiyetine duyarlılığı ve tıbbi bağlamda veri işleme prensipleriyle de farklılaşmaktadır (54).

Tıbbi belge üretimi bağlamında Claude 3.5 Sonnet, renal yetersizlik tanısı almış hastalar için oluşturulan taburculuk özetlerinde, insan hekimlerle benzer düzeyde içerik kalitesi sunmuştur. Yapılan karşılaştırmada Claude'un doğruluk puanı 90.32 ± 3.75, hekimlerin ise 90.81 ± 3.86 olarak hesaplanmış; istatistiksel olarak anlamlı bir fark saptanmamıştır (55). Kalite puanları açısından da benzer bir denklem söz konusudur. Ancak en belirgin fark üretim hızında gözlenmiştir: Claude 3.5 Sonnet, belgeleri yaklaşık 30 saniyede üretirken, hekimler ortalama 15 dakikalık bir sürede tamamlamıştır. Bu durum, modelin zaman tasarrufu sağlayarak klinik iş yükünü hafifletme potansiyeline işaret etmektedir. Bununla birlikte, bu hızlı üretim sürecinde modelin anlam derinliği ve bireyselleştirilmiş ifade kapasitesi gibi nitel özelliklerinin dikkatle izlenmesi önem taşımaktadır (55).

Tanısal doğruluk bağlamında, radyolojik değerlendirme senaryolarında Claude'un performansı üç farklı veri koşulunda test edilmiştir: yalnızca klinik öykü, klinik öykü + görüntü açıklamaları ve klinik öykü + görüntü dosyaları. Claude 3.5 Sonnet, metin tabanlı bilgiyle çalıştığında %64.9 doğruluk oranına ulaşırken, görsel veri içeren senaryoda bu oran %30.1'e düşmüştür (56). Claude 3 Opus ile karşılaştırıldığında, yalnızca görsel dosya içeren durumda Sonnet'in doğruluk oranı

anlamlı şekilde daha yüksekti (p=0.028). Bu bulgular, Claude'un metin temelli karar süreçlerinde daha kararlı bir performans sunduğunu, ancak görsel veriye dayalı klinik tanılarda hâlen sınırlı kapasiteye sahip olduğunu göstermektedir (56).

Claude'un sağlık sistemiyle gerçek dünya etkileşimleri, 4 milyonu aşkın kullanıcı verisi üzerinde yapılan çapraz kesitsel analizle incelenmiştir. Bu analizde, tüm kullanım senaryoları içinde sağlıkla ilişkili görevlerin oranı yalnızca %2.58 olarak belirlenmiştir. Kullanım örüntüsü, özellikle kayıt yönetimi, hasta eğitimi ve bilgilendirme gibi görevlerde yoğunlaşmaktadır (54). Ancak kullanıcıların kimliklerinin anonim olması nedeniyle, bu etkileşimlerin sağlık profesyonellerine mi yoksa doğrudan hastalara mı ait olduğu netleştirilememektedir. Bu durum, araştırma bağlamında bir metodolojik sınırlılık olarak değerlendirilmiştir; doğrudan bir güvenlik riski tanımlanmamıştır. Yine de bu belirsizlik, tıbbi sorumluluğun sınırlarını tartışmalı hale getirebileceğinden, gelecekteki klinik uygulamalarda denetimli kullanımın gerekliliğini ortaya koymaktadır.

Claude'un tıbbi güvenlik açısından bir diğer önemli yönü, veri işleme süreçlerinde HIPAA uyumlu anonimleştirme protokollerini kullanmasıdır (55). Kişisel sağlık bilgilerine doğrudan erişmeden işlem yapabilmesi, hasta mahremiyeti açısından avantaj sağlamaktadır. Bununla birlikte, modelin ürettiği içeriklerin bağlamsal doğruluğu ve klinik uygunluğu, özellikle karar destek sistemlerinde kullanılmadan önce hekim denetimi altında değerlendirilmelidir.

2.5. Acil Tıpta Yapay Zekâ Kullanımı

Yapay zekâ (YZ) teknolojilerinin acil tıp uygulamalarına entegrasyonu, klinik karar destek sistemlerinin işlevselliğini yeniden tanımlamakta ve hasta yönetim süreçlerine yeni bir boyut kazandırmaktadır. Özellikle acil servis gibi zamanla yarışan, kaynakları sınırlı ve yüksek karar yoğunluğu gerektiren klinik ortamlarda, YZ sistemlerinin hızlı veri işleme, örüntü tanıma ve karar önerisi sunma kapasiteleri, dikkat çekici ölçüde kullanılabilir hâle gelmiştir (57). Triyaj algoritmalarından mortalite öngörüsüne kadar uzanan geniş bir yelpazede kullanılan bu sistemler, klinik süreçleri dönüştürme potansiyeli taşısa da etik, hukuki ve mesleki düzeyde çok katmanlı değerlendirmeleri beraberinde getirmektedir.

Acil tıp pratiğinde kullanılan YZ tabanlı uygulamaların büyük çoğunluğu, karar destek sistemleri olarak konumlandırılmakla birlikte, bazı örneklerde sistem çıktılarının sınıflandırıcı ya da yönlendirici özellik taşıması dikkat çekmektedir (57). Bu durum, karar üretimi ile karar desteği arasındaki ayrımın belirsizleşmesine yol açabilmekte, klinik sorumluluğun dağılımına ilişkin etik ve düzenleyici boşlukları gündeme getirmektedir. Gerke ve arkadaşları, klinik sorumluluğun yalnızca hekimde bırakılmasının, yapay zekâ önerilerinin karar üzerindeki etkisi göz önüne alındığında, mesleki yükümlülük açısından yetersiz bir yaklaşım olduğunu belirtmektedir (18). Benzer biçimde, Jassar ve arkadaşları, YZ sistemlerinin klinik kararlarda kullanılmasına ilişkin yasal çerçevenin belirsizliğini, sağlık hizmeti sunucuları için risk artırıcı bir faktör olarak tanımlamaktadır (19).

Veri güvenliği ve hasta mahremiyeti, YZ sistemlerinin eğitim, entegrasyon ve sürdürülebilirlik süreçlerinin merkezinde yer almaktadır. Pham, sağlık verilerinin toplanması, anonimleştirilmesi, işlenmesi ve saklanması aşamalarında ulusal ve uluslararası veri koruma ilkelerine sıkı uyum gösterilmesinin zorunlu olduğunu vurgulamaktadır (20). Acil servislerde çoğu zaman bilgilendirilmiş onam süreçleri dışında toplanan yüksek hassasiyetli klinik veriler hem etik açıdan hem de hasta-hekim güven ilişkisi bağlamında özel dikkat gerektirmektedir. Ahun ve arkadaşları, pandemi sürecinde kullanılan YZ tabanlı triyaj sistemlerinde hekimlerin bu sistemlere duyduğu güvenin, büyük ölçüde veri işleme süreçlerinin şeffaflığı ve hasta mahremiyetine duyulan saygıyla ilişkili olduğunu ortaya koymaktadır (21).

Model şeffaflığı ve algoritmik açıklanabilirlik, YZ sistemlerinin klinik uygulamalardaki geçerliliği ve güvenilirliği açısından belirleyici unsurlardır. Weiner ve arkadaşları, karar süreçlerinin hekim tarafından izlenemediği black box modellerin, klinik sorumlulukların gerekçelendirilmesini zorlaştırdığını ve mesleki uygulama ile etik yükümlülükler arasında gerilim yarattığını ifade etmektedir (58). Petersson ve arkadaşları ise acil serviste mortalite öngörüsüne yönelik geliştirilen YZ sistemlerinin etik açıdan kabul edilebilirliği için algoritmanın işleyişinin açıklanabilir, çıktıların doğrulanabilir ve klinik bağlama duyarlı olması gerektiğini belirtmektedir (59). Ayrıca, sistemlerin eğitildiği veri kümelerinin kapsayıcılığı ve çeşitliliği, algoritmanın farklı hasta gruplarına eşit doğrulukla yanıt üretme kapasitesini belirleyen temel parametrelerden biridir (57).

Acil tıp uygulamalarında yapay zekâ (YZ) sistemlerinin kullanımı, yalnızca teknik altyapı özellikleriyle değil; karar süreçlerindeki konumlandırılması, mesleki sorumluluklarla olan ilişkisi, veri güvenliği uygulamaları ve algoritmik şeffaflık düzeyiyle birlikte değerlendirilmektedir. Klinik uygulamaya entegre edilen bu sistemlerin, etik ilkeler, hukuki düzenlemeler ve mesleki standartlarla eşgüdümlü biçimde tasarlanması, güvenli, izlenebilir ve hasta merkezli bir işleyişin sağlanmasına yönelik temel gereklilikler arasında yer almaktadır. Hekimle sistem arasındaki etkileşim biçimi, modelin açıklanabilirlik kapasitesi ve hasta verilerinin işlenme süreçleri, sistemin klinik güvenilirliği üzerinde doğrudan belirleyici unsurlar olarak tanımlanmaktadır (18-21, 57-59).

3. GEREÇ ve YÖNTEM

3.1. Araştırmanın Tipi, Yeri, Zamanı ve İzni

Bu prospektif ve gözlemsel çalışma 1-30 Ağustos 2025 tarihleri arasında Karadeniz Teknik Üniversitesi Tıp Fakültesi Farabi Hastanesi Acil Tıp Anabilim Dalı'nda gerçekleştirildi. 2025/12 protokol numarası ile Karadeniz Teknik Üniversitesi Rektörlüğü Tıp Fakültesi Bilimsel Araştırmalar Etik Kurulu'ndan onay alındıktan sonra yürütüldü.

3.2. Araştırmanın evreni

Çalışmanın evrenini acil servis kırmızı alanına başvuran tüm hastalar oluşturdu. Çalışmanın gerçekleştirildiği acil servis günlük yaklaşık 350-400 hastanın başvurduğu bir kliniktir. Yeşil, sarı, kırmızı alan ve travma alanı olmak üzere 4 alana ayrılmıştır. Kırmızı alan acil müdahale gerektiren (İskemik-hemorajik inme, miyokard infarktüs, gis kanama, pulmoner emboli, septik şok gibi) hastaların tanı, tedavi ve takiplerinin yürütüldüğü alandır ve günlük ortalama 15 hasta başvurusu olmaktadır.

3.3. Örneklem seçimi

Acil servise 1 ay süre boyunca ayaktan başvuran veya 112 ekipleri tarafından olay yerinden getirilen ve triyaj ekibi tarafından ilk başvurusunda kırmızı alana yönlendirilen 18 yaş üzeri hastalar çalışma örneklemini oluşturdu.

3.3.1. Dışlama kriterleri

18 yaş altı hastalar, başka bir merkezden bir ön tanı veya tanı konularak sevk edilen hastalar, kesin tanısı acil serviste netleştirilemeyen hastalar, kesin tanı konulamadan exitus olanlar, veri eksikliği olan hastalar ile çalışmaya katılmaya gönüllü olmayan hastalar çalışma dışı bırakıldı.

3.4. Verilerin toplanması

1. Aşama (Hekim): Acil servis kırmızı alanında değerlendirilmeye alınan ve vital bulgu ölçümleri yapılan hastaların anamnezi bu alanda nöbetçi olan ve 3 yıl ve üzeri deneyimi olan hekim tarafından sorgulandı. Bu vital bulgu ve anamnez bilgilerine göre aynı hekim tarafından, henüz fizik muayene ve tetkik-görüntüleme aşamasına geçilmeden önce en olası 5 adet ön tanı belirlendi ve oluşturulan veri

toplama formuna kaydedildi (**Bkz. EK 1**) Anamnez sorgusu "demografik bilgileri, semptomu, başlangıç zamanı, özgeçmiş, soygeçmiş, kullandığı ilaçlar (antihipertansif, oral antidiyabetik, antiaritmik, antikoagülan, antiplatelet, antiepileptik, antidepresan, diüretik, tiroid ilacı, inhaler, diğer) bilgilerini içermekteydi.

1. Aşama (Yapay zeka modelleri): Hekim tarafından toplanan ve çalışma formuna kaydedilen 1. Aşama verileri çalışma yürütücüsü tarafından aynı gün içerisinde ChatGPT-4o, ChatGPT-5, Claude Pro Opus 4.1 ve Gemini Pro 2.5 modellerine kimliksizleştirilmiş olarak metin formatında girildi. Modellerden en fazla 5 adet olmak üzere en olası ön tanıları oluşturması istendi. Elde edilen veriler hastayı değerlendiren hekimin kör olacağı şekilde veri toplama formuna kaydedildi.

2. Aşama (Hekim): İkinci aşamada hastayı değerlendiren hekim fizik muayenesini gerçekleştirdi ve ihtiyaç duyduğu laboratuvar ve görüntüleme tetkiklerini istedikten ve sonuçlandıktan sonra ön tanılarını gözden geçirdi ve 3 ayırıcı tanı olarak güncelledi. Elde edilen veriler veri toplama formuna kaydedildi.

2. Aşama (Yapay zeka modelleri): Hekim tarafından toplanan ve çalışma formuna kaydedilen 2. Aşama verileri çalışma yürütücüsü tarafından aynı gün içerisinde YZ modellerine kimliksizleştirilmiş olarak metin formatında girildi. Modellerden en fazla 3 adet olmak üzere en olası ayırıcı tanıları belirlemesi istendi. Elde edilen veriler hastayı değerlendiren hekimin kör olacağı şekilde veri toplama formuna kaydedildi.

3. Aşama (Hekim-Uzman): Toplanan tüm veriler ışığında hastayı değerlendiren hekim ve aynı gün nöbetçi olan bir acil tıp uzmanı tarafından hastanın klinik durumuna neden olan primer kesin tanısı netleştirildi ve bu karar altın standart karar olarak belirlendi. Elde edilen veriler veri toplama formuna kaydedildi.

3. Aşama (Yapay zeka modelleri): Yapay zeka modellerinden, elde edilen tüm veriler ışığında hastanın klinik durumuna neden olan primer kesin tanısının belirlenmesi istendi. Elde edilen veriler hastayı değerlendiren hekimin kör olacağı şekilde veri toplama formuna kaydedildi.

Toplanan tüm veriler Microsoft Excel programına aktarılarak veri seti oluşturuldu.

3.5. İstatistiksel analiz

İstatistiksel analizler, R programlama dili (sürüm 4.3.1; R Foundation for Statistical Computing, Viyana, Avusturya) kullanılarak Visual Studio Code (Microsoft Corporation, Redmond, WA, USA) geliştirme ortamında gerçekleştirildi.

Veri düzenleme ve görselleştirme için *tidyverse* ve *ggplot2* paketleri, analizler için *rstatix*, *psych, PMCMRplus* ve *DescTools* paketleri kullanıldı.

Sayısal verilerin normal dağılıma uygunluğu Kolmogorov–Smirnov testi ve görsel incelemeler (histogram, Q–Q grafiği) kullanılarak belirlendi. Dağılım özelliklerine göre, normal dağılan veriler ortalama ± standart sapma, normal dağılmayanlar ise medyan (minimum–maksimum) olarak ifade edildi. Nominal veriler sayı (n) ve yüzde (%) olarak ifade edildi.

3.5.1. Ön tanı analizleri

Hekim tarafından belirlenen 5 ön tanı ile YZ modelleri tarafından üretilen 5 ön tanıların birbiriyle eşleşme % leri belirlendi. Eşleşme yüzdeleri her bir uyum sayısı için (0-5 uyum) hesaplandı. Ayrıca her bir model tarafından üretilen ortalama ön tanı sayısı ve hekim ön tanısı ile eşleşme ortalaması hesaplandı ve gruplar arası farklılık olup olmadığı Repeated measures ANOVA (Friedman) ile belirlendi. Gruplar arası anlamlı farklılık saptanması durumunda, ikili karşılaştırmalar Durbin–Conover post-hoc testi ile yapıldı ve Bonferroni düzeltmesi uygulanarak anlamlılık düzeyi $\alpha = 0.008$ olarak kabul edildi.

3.5.2. Ayırıcı tanı analizleri

Hekim tarafından belirlenen 3 ayırıcı tanı ile YZ modelleri tarafından üretilen 3 ayırıcı tanının birbiriyle eşleşme % leri belirlendi. Eşleşme yüzdeleri her bir uyum sayısı için (0-3 uyum) hesaplandı. Ayrıca her bir model tarafından üretilen ortalama ayırıcı tanı sayısı ve hekim ayırıcı tanısı ile eşleşme ortalamsı hesaplandı ve gruplar arası farklılık olup olmadığı Repeated measures ANOVA (Friedman) ile belirlendi. Gruplar arası anlamlı farklılık saptanması durumunda, ikili karşılaştırmalar Durbin–Conover post-hoc testi ile yapıldı ve Bonferroni düzeltmesi uygulanarak anlamlılık düzeyi $\alpha = 0.008$ olarak kabul edildi.

3.5.2. Primer kesin tanı analizleri

Hekim tarafından belirlenen primer kesin tanı ile modeller tarafından belirlenen primer kesin tanılar arasındaki doğruluk oranları hesaplandı ve gruplar arası fark olup olmadığı Cochran Q testi ile belirlendi.

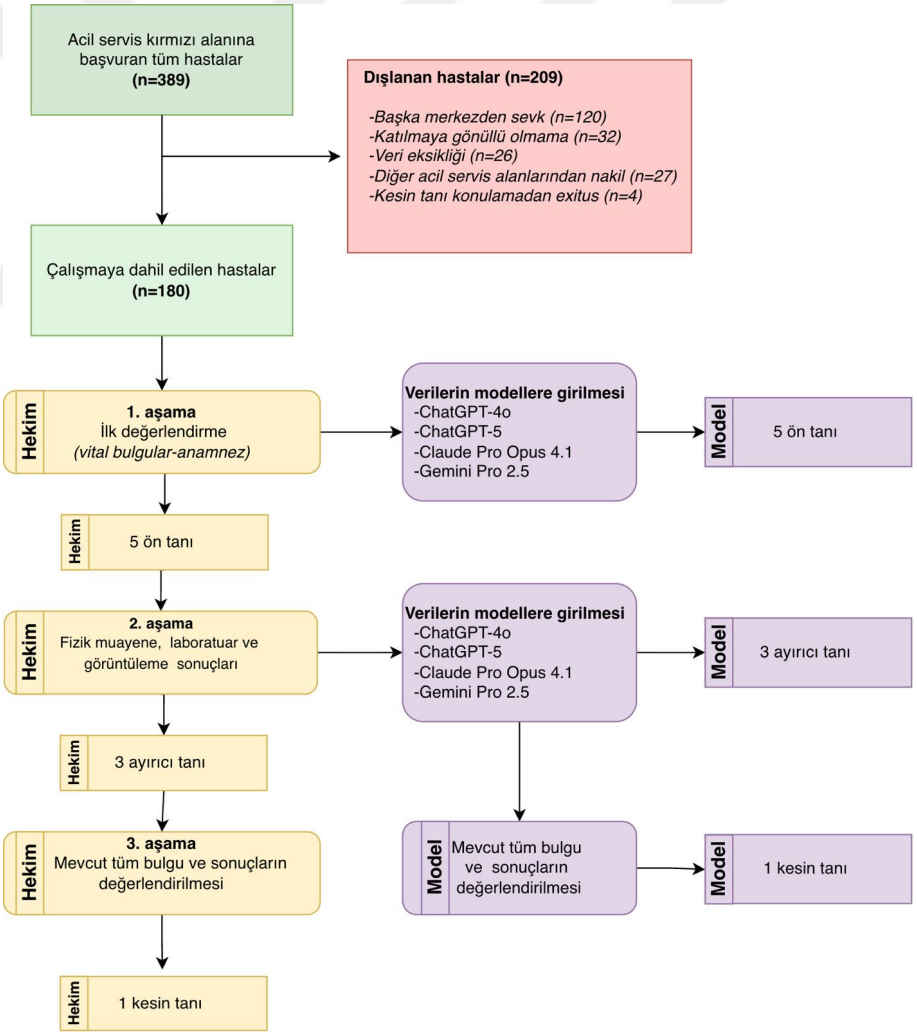
Ayrıca modeller ile hekim arasındaki kesin tanı uyumu Cohen's kappa katsayısı ile ölçüldü. Kappa değerleri 0–0.20 arası zayıf, 0.21–0.40 düşük, 0.41–0.60 orta, 0.61–0.80 iyi ve 0.81–1.00 mükemmel uyum olarak yorumlandı.

İstatistiksel analizlerde %95 güven aralıkları raporlandı. Analizlerde anlamlılık düzeyi $\alpha = 0.05$ olarak kabul edildi; çoklu karşılaştırmalar için Bonferroni düzeltmesi uygulanarak anlamlılık düzeyi $\alpha = 0.008$ olarak belirlendi.

4.BULGULAR

4.1. Hasta Demografik ve Klinik Özellikleri

Çalışmaya 389 hasta dahil edildi. Dışlama kriterleri sonrası 180 hasta ile çalışma tamamlandı. Çalışmaya ait akış şeması **Şekil 1**'de gösterilmiştir. Dahil edilen hastaların %56,1'i (n = 101) erkek, %43,9'u (n = 79) kadındı. Erkek hastalarda ortalama yaş 63,9 ± 15,7 yıl, kadınlarda ise 71,1 ± 15,2 yıl olarak belirlendi. Vital parametreler incelendiğinde, median nabız sayısı 85 atım/dk (45–200), sistolik kan basıncı 136 mmHg (53–220), diastolik kan basıncı 80 mmHg (30–172), solunum sayısı 18/dk (12–45), vücut ısısı 36,5 °C (35,5–38,7) ve SpO_2 değeri %96 (40–100) olarak kaydedildi.



Şekil 1. Çalışmaya ait akış şeması

Komorbiditeler arasında en sık görülen hastalıklar hipertansiyon (%23,6), koroner arter hastalığı (%12,6) ve diabetes mellitus (%10,9) idi. Diğer sık eşlik eden durumlar arasında atriyal fibrilasyon (%8,7), malignite (%6,0), kronik böbrek hastalığı (%5,4) ve serebrovasküler olay (%5,2) yer aldı. İlaç kullanımı değerlendirildiğinde, hastaların %25,7'si antihipertansif, %17,0'ı antiplatelet, %16,4'ü antiaritmik, %8,3'ü antikoagülan ve %7,7'si diüretik tedavi kullanıyordu. Diğer medikasyonlar arasında oral antidiyabetikler (%7,2), inhaler ajanlar (%4,0), antitiroid ilaçlar (%3,6) ve antiepileptikler (%3,2) yer aldı. Hastaların demografik verileri, vital parametreleri, komorbid hastalıkları ve ilaç kullanım oranları **Tablo 1**'de gösterilmiştir.

Tablo 1. Hastaların demografik özellikleri, vital parametreleri, komorbiditeleri ve ilaç kullanımları

Cinsiyet	**n (%)**	**Yaş (mean±sd)**
Erkek	101 (56.1)	63.9 ± 15.7
Kadın	79 (43.9)	71.1 ± 15.2
Vital bulgular	**Median (min-max)**	
Nabız sayısı/dk	85 (45 - 200)	
Sistolik kan basıncı (mmHg)	136 (53 - 220)	
Diastolik kan basıncı (mmHg)	80 (30 - 172)	
Solunum sayısı /dk	18 (12 - 45)	
Vücut ısısı (ºC)	36.5 (35.5 - 38.7)	
SPO_2	96 (40 - 100)	
Komorbid hastalıklar	**n (%)**	
Hipertansiyon	122 (23.6%)	
Koroner arter hastalığı	65 (12.6%)	
Diabetes mellitus	56 (10.9%)	
Atrial fibrilasyon	45 (8.7%)	
Malignite	31 (6.0%)	
Kronik böbrek hastalığı	28 (5.4%)	
Serebrovasküler olay	27 (5.2%)	
Kalp yetmezliği	23 (4.5%)	
Hipotiroidi	22 (4.3%)	
KOAH	21 (4.1%)	
Koroner bypass	18 (3.5%)	
Astım	10 (1.9%)	
Kalp kapak hastalığı	10 (1.9%)	

Psikiyatrik hastalık	6 (1.2%)
Vasküler hastalık	6 (1.2%)
Epilepsi	3 (0.6%)
Diğer	15 (2.9%)
Medikasyon	
Antihipertansif	121 (25.7%)
Antiplatelet	80 (17.0%)
Antiaritmik	77 (16.4%)
Antikoagulan	39 (8.3%)
Diüretik	36 (7.7%)
Oral antidiyabetik	34 (7.2%)
Diğer	23 (4.9%)
İnhaler	19 (4.0%)
Anti-tiroid ajan	17 (3.6%)
Antiepileptik	15 (3.2%)
Antidepresan	9 (1.9%)

4.2. Acil Servise Başvuru Şikayetleri

Hastaların acil servise başvuru nedenleri incelendiğinde, en sık görülen semptomun göğüs ağrısı (%16,9) olduğu belirlendi. Bunu nefes darlığı (%16,0), dizartri veya afazi (%9,3) ve parezi (%7,2) izledi. Tüm başvuru semptomlarının dağılımı **Tablo 2**'de gösterilmiştir.

Tablo 2. Hastaların acil servise başvuru şikayetleri

Başvuru şikâyeti	**n (%)**
Göğüs ağrısı	40 (16.9%)
Nefes darlığı	38 (16.0%)
Dizartri/afazi	22 (9.3%)
Parezi	17 (7.2%)
Senkop	13 (5.5%)
Çarpıntı	13 (5.5%)
Hematokezya	13 (5.5%)

Başvuru şikâyeti	n (%)
Presenkop	12 (5.1%)
Bilinç değişikliği	12 (5.0%)
Melena	10 (4.2%)
Karın ağrısı/kusma	10 (4.2%)
Baş ağrısı	6 (2.5%)
Genel durum bozukluğu	6 (2.4%)
Halsizlik	5 (2.1%)
Baş dönmesi	5 (2.1%)
Diğer	5 (2.1%)
Hemoptizi	3 (1.3%)
Görme bozukluğu	3 (1.3%)
Ateş	2 (0.8%)
Nöbet	2 (0.8%)
Diğer	5 (2.0%)

4.3. Kesin Tanıların Dağılımı

Hastaların kesin tanıları incelendiğinde, en sık görülen tanının iskemik serebrovasküler olay (SVO) (%13,3) olduğu belirlendi. Bunu üst gastrointestinal sistem (GİS) kanaması (%7,8) ve unstabil angina pektoris (%7,2) izledi. Akciğer ödemi (%5,6), sepsis (%5,0), pnömoni (%5,0) ve NSTEMI (%4,4) diğer sık görülen tanılar arasında yer aldı. Daha az sıklıkta geçici iskemik atak (%3,9), septik şok (%3,9), alt GİS kanaması (%3,9) ve KOAH atağı (%2,2) saptandı. Kardiyak ritim bozuklukları (SVT, bradikardi, AV nod blok, VES) ve hipertansif kriz olguları toplamın yaklaşık %7'sini oluşturdu.

Nadir tanılar arasında intraparankimal hemoraji, nöbet, mekanik ağrı, HVCAF, STEMI, anaflaksi, ilaç ilişkili komplikasyonlar (hipotansiyon, GİS irritasyonu, kardiyotoksisite) ve metabolik bozukluklar (hipoglisemi, hiperkalemi) yer aldı.

Ayrıca tekil olgular arasında beyin ödemi, KİBAS, plevral efüzyon, GİS perforasyonu, pankreatit benzeri biliyer kolik, akut kolesistit ve retinal arter tıkanıklığı tanıları kaydedildi. Tüm hastaların kesin tanı dağılımı **Tablo 3**'te gösterilmiştir.

Tablo 3. Anamnez, fizik muayene ve tetkik sonuçlarına göre hastaların kesin tanıları

Kesin Tanı	**n (%)**
İskemik SVO	24 (13.3%)
Üst GİS kanama	14 (7.8%)
Unstabil angina pektoris	13 (7.2%)
Akciğer ödemi	10 (5.6%)
Sepsis	9 (5.0%)
Pnömoni	9 (5.0%)
NSTEMI	8 (4.4%)
Geçici iskemik atak	7 (3.9%)
Septik şok	7 (3.9%)
Alt GİS kanama	7 (3.9%)
KOAH atak	4 (2.2%)
SVT	4 (2.2%)
İntraparankimal hemoraji	4 (2.2%)
Hipertansif kriz	4 (2.2%)
Nöbet	4 (2.2%)
Semptomatik bradikardi	4 (2.2%)
Mekanik ağrı	3 (1.7%)
HVCAF	3 (1.7%)
STEMI	3 (1.7%)
Anaflaksi	2 (1.1%)
İlaca bağlı hipotansiyon	2 (1.1%)
SAK	2 (1.1%)
Vertigo	2 (1.1%)
GİS perforasyonu	2 (1.1%)
KİBAS	1 (0.6%)

AV nod blok	1 (0.6%)
Gastroenterit	1 (0.6%)
PTE + Pnömoni	1 (0.6%)
Solunum yolu obstruksiyonu	1 (0.6%)
VES	1 (0.6%)
Dispepsi	1 (0.6%)
Plevral efüzyon	1 (0.6%)
Beyin ödemi	1 (0.6%)
VT	1 (0.6%)
Hipoglisemi	1 (0.6%)
Alveolar hemoraji	1 (0.6%)
Kardiyotoksisite	1 (0.6%)
İlaca bağlı GİS irritasyonu	1 (0.6%)
Üriner sistem enfeksiyonu	1 (0.6%)
Retina dekolmanı	1 (0.6%)
DKMP'ye bağlı dispne	1 (0.6%)
Anksiyete	1 (0.6%)
Hiperkalemi	1 (0.6%)
Sinüzoidal pause	1 (0.6%)
Kabızlık	1 (0.6%)
Biliyer kolik	1 (0.6%)
Akut kolesistit	1 (0.6%)
Anemi	1 (0.6%)
Pnömoni + Akciğer ödemi	1 (0.6%)
Retinal arter tıkanıklığı	1 (0.6%)
İskemik SVO + PTE	1 (0.6%)
Deli bal zehirlenmesi	1 (0.6%)
Astım atak	1 (0.6%)

4.4. Yapay Zekâ Modellerinin Ön Tanı Uyumu

Yapay zekâ modellerinin ürettiği ön tanı sayıları incelendiğinde, ortalama değerler GPT-4o için 4,80 ± 0,62, GPT-5 için 4,71 ± 0,49, Claude için 4,66 ± 0,66 ve Gemini için 4,91 ± 0,43 olarak belirlendi. Bu ön tanıların hekim ön tanılarıyla eşleşme ortalamaları sırasıyla GPT-4o'da 2,26 ± 0,94, GPT-5'te 2,13 ± 0,90, Claude'da 2,37 ± 1,02 ve Gemini'de 2,29 ± 0,96 düzeyindeydi. Gruplar arasındaki fark istatistiksel olarak anlamlıydı (p = 0,014). Post-hoc analizinde, GPT-5 ile Claude arasındaki fark anlamlı bulundu (p = 0,002). Diğer modeller arasında eşleşme başarısı açısından anlamlı bir farklılık yoktu.

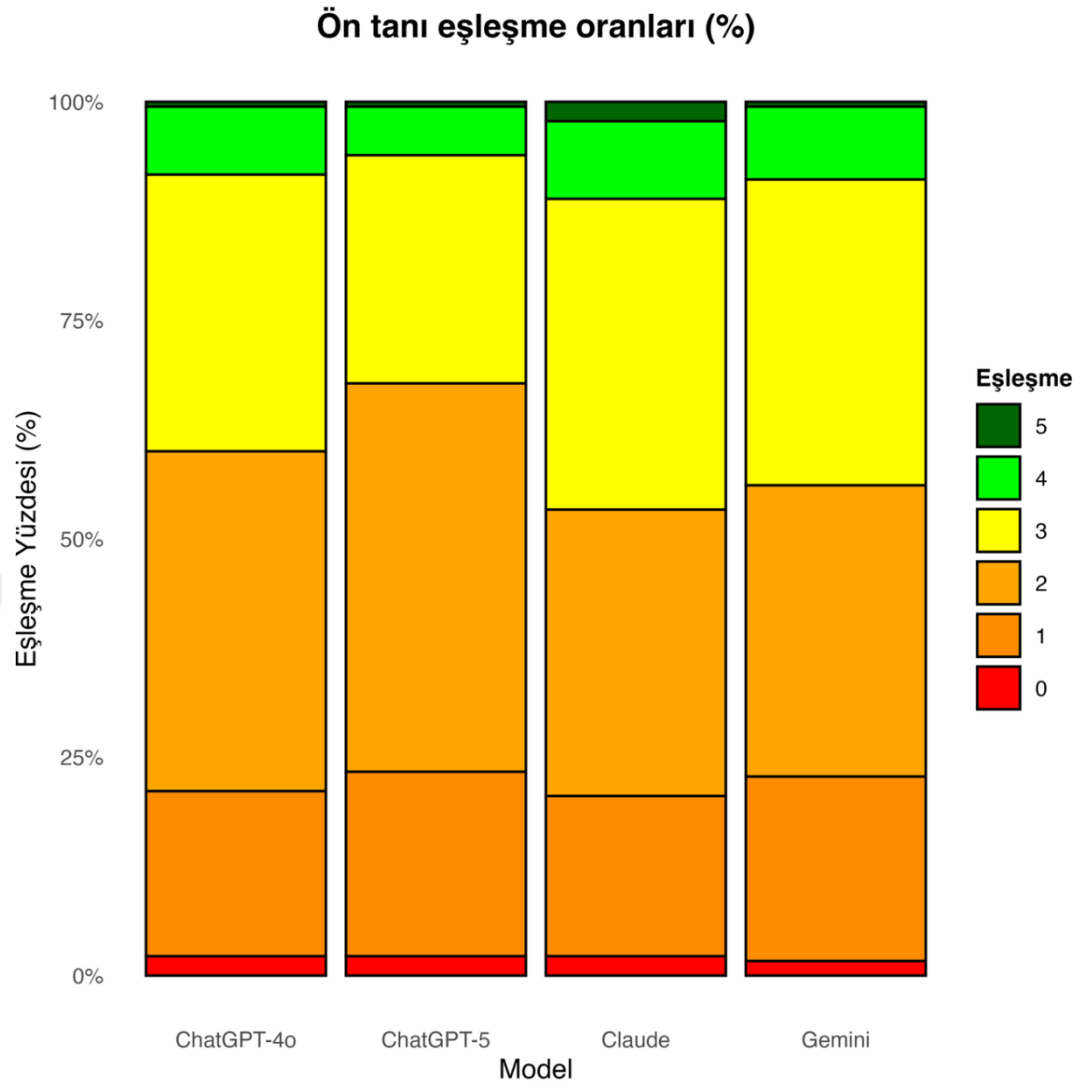
Eşleşme dağılımı değerlendirildiğinde, hekim ön tanısıyla hiç eşleşme olmayan olguların oranı GPT-4o, GPT-5 ve Claude modellerinde 4 olgu (%2,2), Gemini'de ise 3 olgu (%1,7) düzeyindeydi. Bir tanı eşleşmesi GPT-4o'da 34 olgu (%18,9), GPT-5 ve Gemini'de 38 olgu (%21,1), Claude'da ise 33 olgu (%18,3) oranında gözlendi. İki tanı eşleşmesi GPT-4o'da 70 olgu (%38,9), GPT-5'te 80 olgu (%44,4), Claude'da 59 olgu (%32,8) ve Gemini'de 60 olgu (%33,3) oranındaydı. Üç tanı eşleşmesi GPT-4o'da 57 olgu (%31,7), GPT-5'te 47 olgu (%26,1), Claude'da 64 olgu (%35,6) ve Gemini'de 63 olgu (%35,0) olarak saptandı. Dört ve üzeri tanı eşleşmesi nadir olup, GPT-4o'da 15 olgu (%8,4), GPT-5'te 11 olgu (%6,2), Claude'da 20 olgu (%11,1) ve Gemini'de 16 olgu (%8,9) oranlarında görüldü (**Şekil 2**). Yapay zekâ modelleri arasındaki ön tanı ortalama sayıları Friedman testi ile karşılaştırıldı ve fark istatistiksel olarak anlamlı bulundu ($\chi^2(3) = 10,7$; $p < 0,05$). Farkın kaynağı Durbin–Conover post-hoc analizi (Bonferroni düzeltmeli $\alpha = 0,008$) ile belirlendi. Elde edilen veriler **Tablo 4'te** gösterildi.

Tablo 4. ChatGPT, Claude ve Gemini'nin ürettiği ön tanılar ile hekim ön tanıları arasındaki eşleşme sonuçları

	[a]GPT-4o	[b]GPT-5	[c]Claude	[d]Gemini	P değeri
Üretilen ön tanı sayısı (mean ± sd)	4.8±0.62	4.71±0.49	4.66±0.66	4.91±0.43	
Hekim ön tanısıyla eşleşme (mean ± sd)	2.26±0.94	2.13±0.9	2.37±1.02	2.29±0.96	*p = 0.014 †p^{b-c}=0.002
Hekim ön tanısı ile eşleşme sayısı	**n (%)**	**n (%)**	**n (%)**	**n (%)**	
0	4 (2.2%)	4 (2.2%)	4 (2.2%)	3 (1.7%)	
1	34 (18.9%)	38 (21.1%)	33 (18.3%)	38 (21.1%)	
2	70 (38.9%)	80 (44.4%)	59 (32.8%)	60 (33.3%)	
3	57 (31.7%)	47 (26.1%)	64 (35.6%)	63 (35.0%)	
4	14 (7.8%)	10 (5.6%)	16 (8.9%)	15 (8.3%)	
5	1 (0.6%)	1 (0.6%)	4 (2.2%)	1 (0.6%)	

*Repeated measures ANOVA (Friedman): $\chi^2(3) = 10.7$

†Friedman test with Durbin–Conover post hoc analysis (Bonferroni-adjusted α = 0.008)



Şekil 2. ChatGPT, Claude ve Gemini'nin ürettiği ön tanılar ile hekim ön tanıları arasındaki eşleşme düzeyleri

4.5. Yapay Zekâ Modellerinin Ara Tanı Uyumu

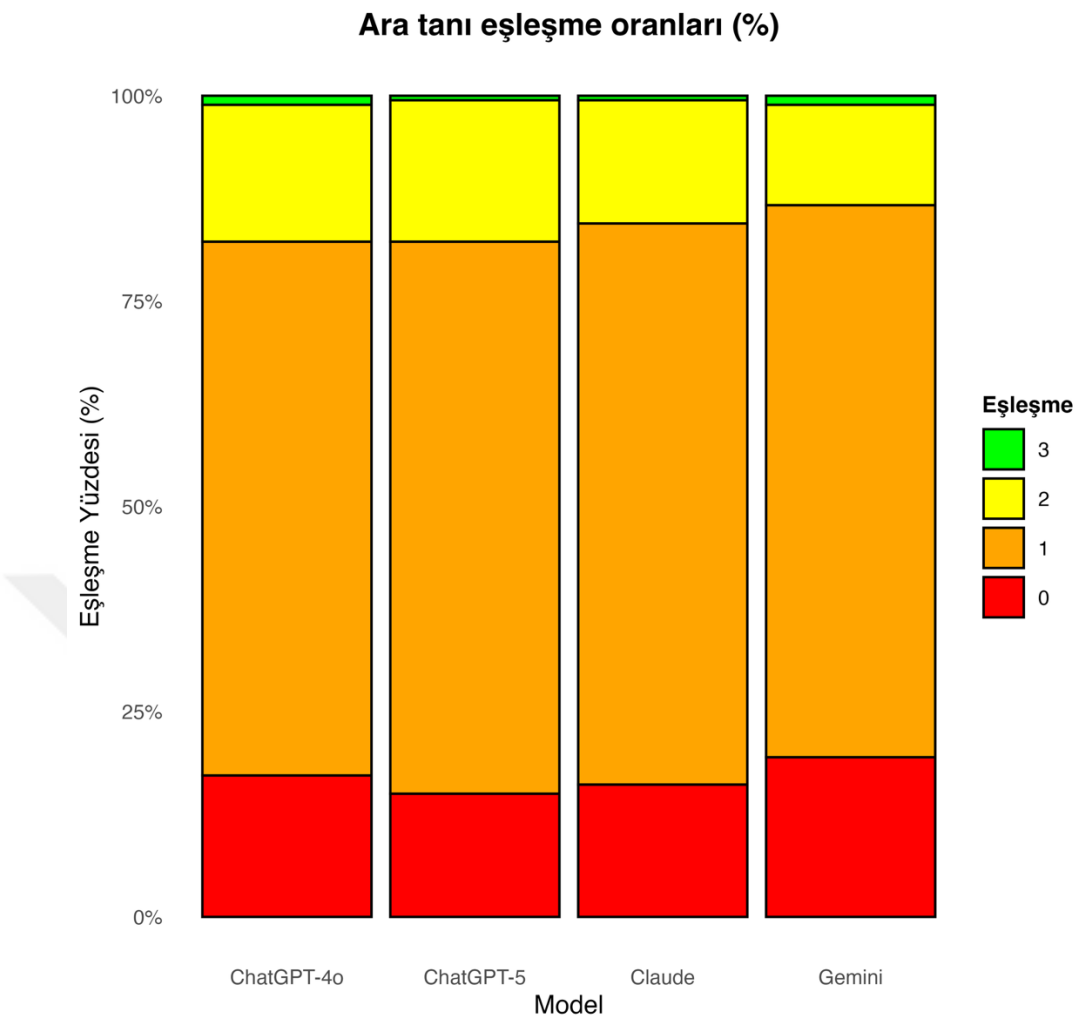
Yapay zekâ modellerinin ürettiği ara tanı sayısı ortalamaları incelendiğinde, GPT-4o için 2,84 ± 0,42, GPT-5 için 2,66 ± 0,58, Claude için 2,86 ± 0,41 ve Gemini için 2,72 ± 0,60 olarak belirlendi. Hekim ara tanılarıyla eşleşme ortalamaları sırasıyla GPT-4o'da 1,02 ± 0,62, GPT-5'te 1,03 ± 0,58, Claude'da 1,00 ± 0,58 ve Gemini'de 0,95 ± 0,60 düzeyindeydi. Gruplar arasındaki fark istatistiksel olarak anlamlı değildi (p = 0,367).

Eşleşme düzeyleri incelendiğinde, hiç eşleşme olmayan olguların oranı GPT-4o'da 31 olgu (%17,2), GPT-5'te 27 olgu (%15,0), Claude'da 29 olgu (%16,1) ve

Gemini'de 35 olgu (%19,4) olarak kaydedildi. Bir tanı eşleşmesi en sık gözlenen düzey olup, GPT-4o'da 117 olgu (%65,0), GPT-5 ve Gemini'de 121 olgu (%67,2), Claude'da 123 olgu (%68,3) oranlarında görüldü. İki tanı eşleşmesi GPT-4o'da 30 olgu (%16,7), GPT-5'te 31 olgu (%17,2), Claude'da 27 olgu (%15,0) ve Gemini'de 22 olgu (%12,2) oranında saptandı. Üç tanı eşleşmesi nadir olup, GPT-4o'da 2 olgu (%1,1), GPT-5 ve Gemini'de 2 olgu (%1,1), Claude'da 1 olgu (%0,6) düzeyindeydi (**Şekil 3**). Yapay zekâ modelleri arasındaki ara tanı uyumu farkı Friedman testi ile değerlendirildi ve istatistiksel olarak anlamlı bulunmadı ($\chi^2(3) = 3,16$; $p > 0,05$). Elde edilen bulgular **Tablo 5'te** gösterildi.

Tablo 5. ChatGPT, Claude ve Gemini'nin ürettiği ara tanılar ile hekim ara tanıları arasındaki eşleşme sonuçları

	GPT-4o	**GPT-5**	**Claude**	**Gemini**	**p değeri**
Üretilen ara tanı sayısı (mean ±sd)	2.84 ± 0.42	2.66 ±0.58	2.86 ± 0.41	2.72 ± 0.6	
Hekim ara tanısıyla eşleşme (mean±sd)	1.02 ± 0.62	1.03±0.58	1 ± 0.58	0.95 ± 0.6	*p=0.367
Hekim ara tanısı ile eşleşme sayısı	**n (%)**	**n (%)**	**n (%)**	**n (%)**	
0	31 (17.2%)	27 (15%)	29 (16.1%)	35 (19.4%)	
1	117 (65.0%)	121(67.2%)	123 (68.3%)	121 (67.2%)	
2	30 (16.7%)	31 (17.2%)	27 (15.0%)	22 (12.2%)	
3	2 (1.1%)	1 (0.6%)	1 (0.6%)	2 (1.1%)	
	*Repeated measures ANOVA (Friedman): $\chi^2(3) = 3.16$				



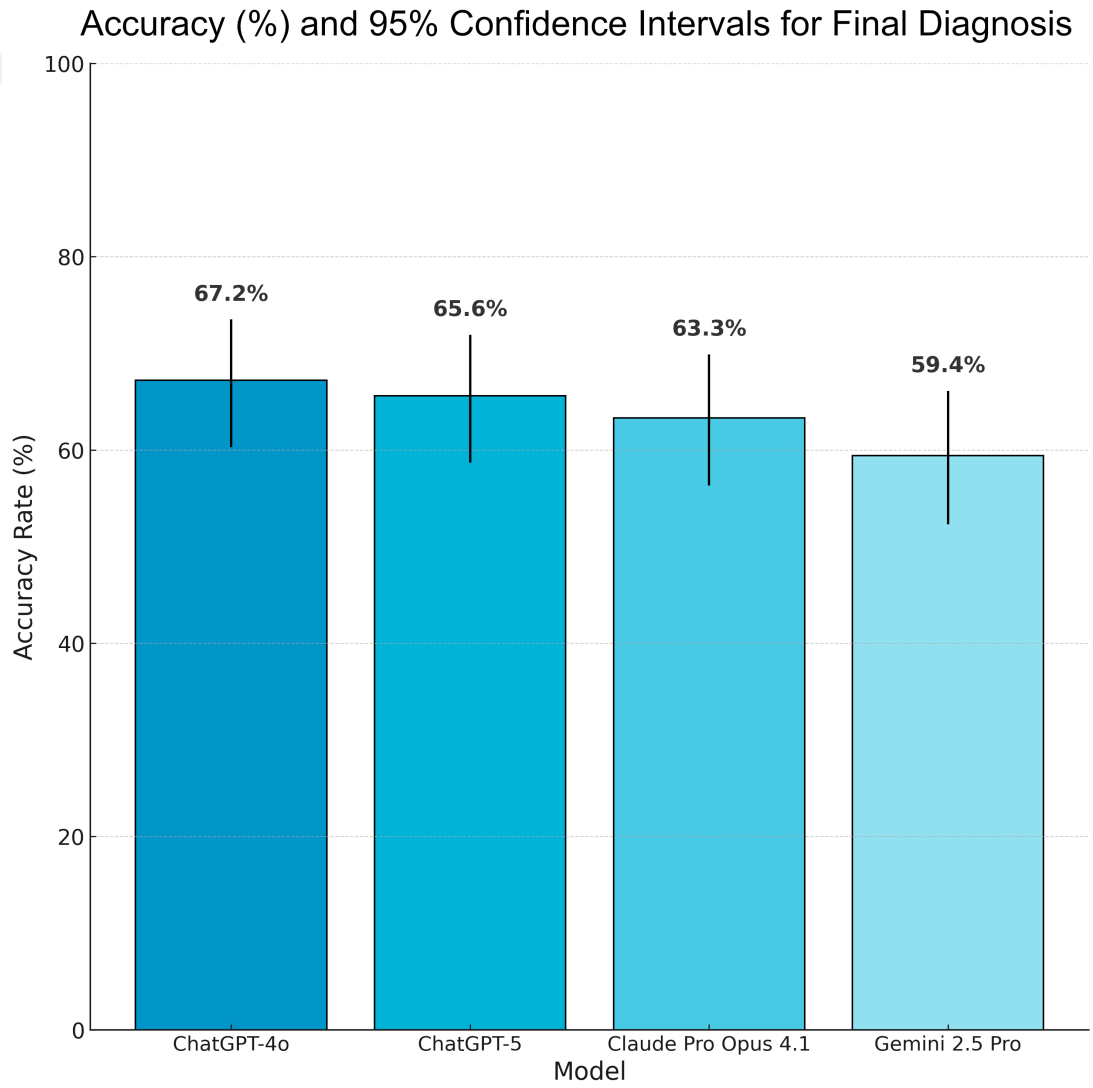
Şekil 3. ChatGPT, Claude ve Gemini'nin ürettiği ara tanılar ile hekim ara tanıları arasındaki eşleşme düzeyleri

4.6. Yapay Zekâ Modellerinin Kesin Tanı Performansı

Yapay zekâ modellerinin kesin tanı doğruluk oranları incelendiğinde, GPT-4o'nun doğruluk oranı %67,2 (n = 121), GPT-5'in %65,6 (n = 118), Claude'un %63,3 (n = 114) ve Gemini'nin %59,4 (n = 107) olarak belirlendi (**Şekil 4**). Gruplar arasındaki fark Cochran's Q testi ile değerlendirildi ve bulgular istatistiksel olarak anlamlı değildi (p = 0,16). Altın standart tanı ile uyum düzeyleri değerlendirildiğinde, Cohen's κ katsayısı GPT-4o için 0,656, GPT-5 için 0,638, Claude için 0,616 ve Gemini için 0,575 olarak bulundu. Tüm modellerde altın standartla uyum istatistiksel olarak anlamlıydı (p < 0,001). Elde edilen bulgular **Tablo 6'da** gösterildi.

Tablo 6. ChatGPT, Claude ve Gemini'nin ürettiği kesin tanıların doğruluğu

	GPT-4o[1]	GPT-5	Claude[2]	Gemini[4]	p değeri
Kesin tanı doğruluğu, n(%) 95% CI	121 (67.2%) [60.3-73.5]	118 (65.6%) [58.7-71.9]	114 (63.3%) [56.3-69.9]	107 (59.4%) [52.3-66.1]	*p=0.16
Altın standart ile uyum (Kappa)	0.656	0.638	0.616	0.575	p<0.001
*Cochran's Q test: Q (3) = 5.1 p<0.05 anlamlılık düzeyi kabul edilmiştir.					



Şekil 4. ChatGPT, Claude ve Gemini'nin ürettiği kesin tanıların doğruluk oranları

5. TARTIŞMA

Bu çalışmada, acil servis kırmızı alanına başvuran kritik hastalarda, popüler büyük dil modellerin (ChatGPT, Claude ve Gemini) tanısal performansları ile hekimlerin tanısal kararlarını çok aşamalı bir yaklaşımla karşılaştırdık. Elde ettiğimiz bulgular, bu modellerin acil tanı sürecinin üç aşamasında (ön tanı, ayırıcı tanı ve kesin tanı) kabul edilebilir bir performans sergilediğini ortaya koymaktadır.

LLM'lerin tanısal performansına ilişkin literatürde bazı çalışmalar bulunsa da, bu çalışmalar çoğunlukla retrospektif tasarıma sahiptir ya da sentetik olarak oluşturulmuş vakalar üzerinden yürütülmüştür. Ayrıca modellerin performansı çoğunlukla oluşturulan ön tanı listelerinde son tanının yer alıp almadığına odaklanmıştır. Bunun aksine, çalışmamızın prospektif olarak yürütülmesi, veri toplama sürecinde seçilim yanlılığını en aza indirerek gerçek klinik pratiğe daha yakın sonuçlar elde edilmesini sağlamıştır. Metodolojik tasarımımız, yalnızca kritik ve komplike hastalarda ayırıcı tanı listesi oluşturma veya son tanı doğruluğunu değerlendirmekle sınırlı kalmayıp, tüm klinik verilerin entegrasyonunu içeren kapsamlı bir yaklaşımı benimsemiştir. Bu kapsamda, hasta yönetiminin ilk değerlendirmeden son tanıya kadar olan tüm aşamaları ayrıntılı biçimde analiz edilmiş olması diğer bir güçlü yanıdır. Bu çok aşamalı yapı, LLM'lerin klinik muhakeme süreçlerinin daha derinlemesine incelenmesine olanak tanımaktadır. Ayrıca, üç farklı popüler LLM performansının analiz edilmiş olması modellerin birbirlerine göre üstünlükleri hakkında karşılaştırmalı veriler sunmaktadır.

Çalışma popülasyonumuz, nispeten ileri yaşta, komorbidite yükü yüksek ve sıklıkla kardiyak, pulmoner veya nörolojik acil durumlar nedeniyle başvuran hastalardan oluşmaktaydı. Bu tür popülasyonlarda hızlı ve doğru tanı koymak, hasta yönetimi ve hayat kurtarıcı müdahalelerin zamanında uygulanabilmesi açısından kritik öneme sahiptir. Çalışmamızın klinik açıdan karmaşık ve kritik olgulara odaklanması, büyük dil modellerinin yüksek riskli hasta gruplarındaki tanısal performansını değerlendirme açısından literatüre önemli bir katkı sağlamaktadır. Bu yönüyle çalışmamız, modellerin yalnızca rutin olgularda değil, aynı zamanda acil servislerde klinik karar vermenin en güç olduğu durumlarda da potansiyel bir karar destek aracı olarak kullanılabilirliğini ortaya koyması bakımından değerlidir.

Literatürde yapılan çalışmalara bakıldığında, Levine ve ark. tarafından GPT-3 modeliyle yürütülen gözlemsel bir çalışmada, 48 sentetik klinik vinyet kullanılmış ve modelin doğru tanıyı ilk üç ayırıcı tanı arasında belirtme oranı %88 olarak bildirilmiştir. Ancak bu çalışmada modele sunulan bilgiler oldukça kısadır ve yalnızca anamnez ve vital bulgu bilgisine dayanmaktadır. Fizik muayene ve laboratuvar/görüntüleme gibi kritik bulguların dahil edilmemesi klinik açıdan kullanışlılığını ve gerçek vakalara genellenebilirliğini önemli ölçüde sınırlamaktadır (60). Çalışmamızda ise gerçek hasta verilerini kullandık ve tanı sürecini çok aşamalı olarak yapılandırdık. Bu sayede, büyük dil modellerinin tanısal sürece adaptasyonu gerçekçi bir biçimde değerlendirip elde ettiğimiz sonuçların klinik pratikteki geçerliliği artırmaya çalıştık. Levine ve ark.nın çalışmasında tanısal doğruluğun, doğru tanının ilk üç ayırıcı tanı arasında yer alması şeklinde tanımlanmış olması, bildirilen yüksek başarı oranının nedeni olabilir.

Hidde ten Berg ve ark. tarafından yapılan retrospektif çalışmada, yalnızca tek bir kesin tanısı bulunan 30 vaka değerlendirilmiştir. Çalışmada GPT-4 modelinin, doğru tanıyı ilk beş ayırıcı tanı arasında belirtme oranı laboratuvar verileri olmadan ve laboratuvar verileri eklendiğinde değişmeden %87 olarak bildirilmiştir. GPT-3.5 modeli için ise bu oranlar sırasıyla %77 ve %97 olarak saptanmıştır (61). Ancak, çalışmada da levine ve ark. çalışması gibi performans ölçütü olarak yalnızca doğru tanının ön tanılar arasında yer alması kabul edilmiştir (60). Bu nedenle, modelden tek bir kesin tanı istenmesi durumunda modelin performansının ne olacağı belirsizdir. Bizim çalışmamızda ise yalnızca ön tanı ve tanı uyum oranları değerlendirilmemiş, aynı zamanda hekimlerin klinik filtreleme sürecini taklit eden üç aşamalı bir yaklaşımla son kesin tanıya ulaşılması ve bu tanının mutlak doğruluğunun belirlenmesi amaçlanmıştır. Bu metodolojik yaklaşımın, kesin tanı koymadaki güvenilirliği artırdığını düşünüyoruz.

Hoppe ve ark. tarafından 100 rastgele seçilmiş yatan hasta üzerinde yürütülen retrospektif çalışmada, altın standart tanı olarak hastaların taburculuk tanısı kabul edilmiştir. Hastaların öyküleri, tıbbi geçmişleri, laboratuvar testleri ve diğer tanısal tetkikleri ChatGPT'ye girilmiş ve modelin performansı acil servis asistanlarının tanısal doğruluğu ile karşılaştırılmıştır. Tanısal doğruluk, 0: yanlış, 1: kısmen doğru, 2: tamamen doğru arasında puanlanan bir ölçekle değerlendirilmiştir. Çalışmada GPT-

4'ün ortalama skorunun (1,76), acil servis hekimlerinin skorundan (1,59) daha yüksek olduğu saptanmıştır (62). Bizim çalışmamızda altın standart karar olarak, ilgili günün acil tıp uzmanı ile kıdemli hekimin ortak değerlendirmesi esas alınmıştır. Acil servisten taburculuk tanısı veya bir servise yatış tanısı, modelin değerlendirilmesi açısından sonlanım noktası olarak belirlenmiştir. Servis sürecinde ortaya çıkabilecek ek durumlar, komplikasyonlar ya da yeni tetkik sonuçlarının sonraki tanı sürecini ve taburculuk tanısını etkileyebileceği göz önünde bulundurularak, bu bilgiler çalışma kriterlerine dahil edilmemiştir.

Wang ve ark. tarafından, Ten Berg'in 30 vakalık retrospektif veri seti kullanılarak gerçekleştirilen çalışmada, yalnızca semptomlar, tıbbi öykü ve fizik muayene bulguları değerlendirmeye alınmış; laboratuvar verileri kullanılmamıştır. Çalışmada, kesin tanının ilk üç ayırıcı tanı arasında yer alma oranı incelenmiş ve büyük dil modelleri için bu oranın genel olarak %70–80 arasında değiştiği, GPT-4o için ise %78,9 olarak saptanmıştır (61, 63). Wang ve ark., modelin sunduğu ilk ön tanıyı kesin tanı olarak kabul ettiklerinde başarı oranını yaklaşık %60 olarak bildirmiştir. Ancak nihai kesin tanı ön tanı düzeyindeki tanıya bağımlıdır. Bizim çalışmamız, modelin tüm tanısal süreç boyunca gösterdiği tutarlılığı değerlendirmesi bakımından literatürdeki çalışmalardan ayrılmaktadır. Modelin tüm tanısal aşamalardan etkilenmiş nihai performansı, %60–67 arasında bir doğruluk oranı ortaya koymuştur. Bu oran, modelin klinik akış içerisindeki çok aşamalı ilerleme süreci ve bu aşamalardaki değişkenlere rağmen elde edilmiş olup, genel tanısal başarıyı yansıtmaktadır.

Büyük dil modellerinin (LLM) temel amacı, doğal dili işleme ve anlamlandırmadır. Bu modeller metin tabanlı veriler üzerinden güçlü çıkarımlar yapabilseler de görüntü analizi yetenekleri hâlen sınırlıdır. Oysa klinik tanıya ulaşma süreci, yalnızca doğru anamnez ve fizik muayene bulgularını değil, aynı zamanda laboratuvar tetkiklerini ve gerektiğinde görüntüleme sonuçlarını da kapsamaktadır. Bu çok boyutlu parametreler birlikte değerlendirildiğinde, bir hastayı başvuru anından taburculuğa kadar bütüncül şekilde yönetme noktasında LLM'lerin mevcut kapasitesi yetersiz kalabilmektedir.

Çalışmamızda hastaların görüntüleme tetkik raporları yerine doğrudan görüntü verileri (EKG ve akciğer grafileri) kullanılmıştır. Bu tür görsellerin

değerlendirilmesinde büyük dil modellerinin mevcut sınırlılıkları, elde edilen doğruluk oranlarının %60–70 düzeylerinde kalmasına neden olmuş olabilir. Bir hastada tanısal sürecin başından sonuna kadar yapay zekâdan etkin biçimde yararlanabilmek için, LLM'lerin görüntü analizine odaklanan yapay zekâ sistemleriyle entegrasyonunun sağlanması faydalı olacaktır. Aynı zamanda, büyük dil modellerinin yerel hastane bilgi sistemlerine entegre edilmesi, veri girişinde ortaya çıkabilecek kısıtlılıkları ve zaman kayıplarını ortadan kaldırabilir.

Çalışmamızda modellere girilen veriler, hekimler tarafından sağlanan anamnez, fizik muayene ve laboratuvar sonuçlarına dayanmaktadır. Tanı sürecinde ilerlerken model tarafından gerekli görülen ek tetkiklerin de sürece dahil edilmesi, tanısal doğruluğu artırma potansiyeline sahip olup olmayacağı bilinmemekle birlikte, bu uygulamanın etik açıdan dikkatle değerlendirilmesi gerekmektedir.

Elde ettiğimiz bulgular ve mevcut literatür birlikte değerlendirildiğinde, büyük dil modellerinin altın standart tanılarla genellikle iyi düzeyde uyum göstermesi, yapay zekâ destekli ön tanı ve ayırıcı tanı sistemlerinin acil tıp pratiğinde potansiyel kullanım alanları olabileceğine işaret etmektedir. Bu tür sistemler, özellikle yoğun ve zaman baskısı altındaki klinik ortamlarda, deneyimi sınırlı hekimlere karar verme sürecinde ikincil bir görüş sağlayarak destek olabilir. Hastaların doğru bir ön değerlendirmeyle hızlı biçimde ayrıştırılması, uygun tedaviye daha erken başlanmasını ve mevcut kaynakların daha etkin kullanılmasını mümkün kılabilir.

Elbette, yapay zekâ henüz bir hekimin yerini alabilecek düzeyde değildir. Nitekim çalışmamızda modellere verilerinin bir hekim tarafından girilmesi ve fizik muayene bulgularının hekim değerlendirmelerine dayanması da bunun en temel göstergesidir. Ancak uygun şekilde kullanıldığında, ChatGPT gibi büyük dil modelleri klinisyenlerin iş yükünü hafifletme, tanı sürecini hızlandırma ve olası hata payını azaltma potansiyeline sahiptir.

Gelecekte, daha geniş örneklemlere sahip araştırmalarla bu bulguların doğrulanması, ayrıca yapay zekânın klinik sonuçlar üzerindeki etkisinin değerlendirilmesi gerekmektedir. Bundan sonraki adımlar, bu tür modellerin hastane bilgi sistemleriyle entegre edilerek gerçek zamanlı karar destek araçlarına dönüştürüldüğü pilot uygulamaları içermelidir.

Kısıtlılıklar

İlk olarak, çalışma tek bir merkezde yürütülmüş olup nispeten sınırlı bir örneklem büyüklüğüne dayanmaktadır. Ayrıca, çalışmamızda değerlendirdiğimiz parametrelerden biri hekimlerin ön ve ayırıcı tanıları ile LLM çıktıları arasındaki eşleşmeyi ölçmek olsa da hekimler tarafından belirlenen tanıların her zaman mutlak doğruluğu yansıtmayabileceği göz önünde bulundurulmalıdır.

Çalışmamızda hekimlerin belirlediği ön tanılar en olası tanılar olarak kabul edilerek eşleşme oranları değerlendirilmiştir. Ancak bu durum, LLM'lerin sunduğu ve altın standart ile birebir örtüşmeyen ön tanıların klinik açıdan değersiz veya kabul edilemez olduğu anlamına gelmemektedir. Bu konuyla ilgili ayrı bir değerlendirme yapılmamıştır.

Üçüncü olarak, modeller her bir hastanın verileri yalnızca bir kez girilerek değerlendirilmiştir. Aynı hasta verilerinin tekrar girilmesi durumunda farklı sonuçlar elde edilip edilmeyeceği araştırılmamıştır.

6. SONUÇ

Çalışmamızın sonuçlarına göre, büyük dil modelleri altın standart tanılarla iyi düzeyde uyum göstermiş ve ürettikleri kesin tanılar genel olarak %60–70 oranında doğruluk sergilemiştir. Bu bulgular, özelllikle kritik kararların hızla verilmesi gereken klinik ortamlarda, LLM'lerin hekimlere tanısal süreçte yardımcı olma potansiyelini ortaya koymaktadır.

7. KAYNAKLAR

1. Kayambankadzanja RK, Schell CO, Wärnberg MG, Tamras T, Mollazadegan H, Holmberg M, et al. Towards definitions of critical illness and critical care using concept analysis. BMJ open. 2022;12(9):e060972.
2. Maslove DM, Tang B, Shankar-Hari M, Lawler PR, Angus DC, Baillie JK, et al. Redefining critical illness. Nature medicine. 2022;28(6):1141-8.
3. Mboya EA, Ndumwa HP, Amani DE, Nkondora PN, Mlele V, Biyengo H, et al. Critical illness at the emergency department of a Tanzanian national hospital in a three-year period 2019–2021. BMC emergency medicine. 2023;23(1):86.
4. Norman G, Pelaccia T, Wyer P, Sherbino J. Dual process models of clinical reasoning: the central role of knowledge in diagnostic expertise. Journal of Evaluation in Clinical Practice. 2024;30(5):788-96.
5. Adams E, Goyder C, Heneghan C, Brand L, Ajjawi R. Clinical reasoning of junior doctors in emergency medicine: a grounded theory study. Emergency Medicine Journal. 2017;34(2):70-5.
6. Al-Azri NH. How to think like an emergency care provider: a conceptual mental model for decision making in emergency care. International Journal of Emergency Medicine. 2020;13(1):17.
7. Kunitomo K, Harada T, Watari T. Cognitive biases encountered by physicians in the emergency room. BMC Emergency Medicine. 2022;22(1):148.
8. Hansen K. Cognitive bias in emergency medicine. Emergency Medicine Australasia. 2020;32(5).
9. Ng M, Wong E, Sim GG, Heng PJ, Terry G, Yann FY. Dropping the baton: Cognitive biases in emergency physicians. PLoS One. 2025;20(1):e0316361.
10. Hautz WE, Kämmer JE, Hautz SC, Sauter TC, Zwaan L, Exadaktylos AK, et al. Diagnostic error increases mortality and length of hospital stay in patients presenting through the emergency room. Scandinavian journal of trauma, resuscitation and emergency medicine. 2019;27(1):54.
11. Tudela P, Carreres A, Ballester M. Diagnostic errors in emergency departments. Medicina Clínica (English Edition). 2017;149(4):170-5.
12. Giardina TD, Hunte H, Hill MA, Heimlich SL, Singh H, Smith KM. Defining diagnostic error: a scoping review to assess the impact of the national academies' report improving diagnosis in health care. Journal of Patient Safety. 2022;18(8):770-8.
13. Kelen GD, Kaji AH, Schreyer KE, Colucci LB, Keim SM, Schneider SM, et al. The AHRQ report on diagnostic errors in the emergency department: the wrong answer to the wrong question. Annals of Emergency Medicine. 2023;82(3):336-40.
14. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nature medicine. 2023;29(8):1930-40.
15. Torres-Zegarra BC, Rios-Garcia W, Ñaña-Cordova AM, Arteaga-Cisneros KF, Chalco XCB, Ordoñez MAB, et al. Performance of ChatGPT, Bard, Claude, and Bing on the Peruvian national licensing medical examination: a cross-sectional study. Journal of Educational Evaluation for Health Professions. 2023;20.
16. Leiser F, Guse R, Sunyaev A. Large Language Model Architectures in Health Care: Scoping Review of Research Perspectives. Journal of Medical Internet Research. 2025;27:e70315.

17. Nerella S, Bandyopadhyay S, Zhang J, Contreras M, Siegel S, Bumin A, et al. Transformers and large language models in healthcare: A review. Artificial intelligence in medicine. 2024;154:102900.
18. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. Artificial intelligence in healthcare: Elsevier; 2020. p. 295-336.
19. Jassar S, Adams SJ, Zarzeczny A, Burbridge BE, editors. The future of artificial intelligence in medicine: medical-legal considerations for health leaders. Healthcare Management Forum; 2022: SAGE Publications Sage CA: Los Angeles, CA.
20. Pham T. Ethical and legal considerations in healthcare AI: innovation and policy for safe and fair use. Royal Society Open Science. 2025;12(5):241873.
21. Ahun E, Demir A, Yiğit Y, Tulgar YK, Doğan M, Thomas DT, et al. Perceptions and concerns of emergency medicine practitioners about artificial intelligence in emergency triage management during the pandemic: a national survey-based study. Frontiers in Public Health. 2023;11:1285390.
22. Lin MP, Burke RC, Sabbatini AK, Latsko E, Edlow JA, Orav EJ, et al. Potential Diagnostic Error for Emergency Conditions, Mortality, and Healthy Days at Home. JAMA Network Open. 2025;8(6):e2516400-e.
23. Brasen CL, Andersen ES, Madsen JB, Hastrup J, Christensen H, Andersen DP, et al. Machine learning in diagnostic support in medical emergency departments. Scientific Reports. 2024;14(1):17889.
24. Platts-Mills TF, Nagurney JM, Melnick ER. Tolerance of uncertainty and the practice of emergency medicine. Annals of emergency medicine. 2020;75(6):715-20.
25. ALQahtani DA, Rotgans JI, Mamede S, ALAlwan I, Magzoub MEM, Altayeb FM, et al. Does time pressure have a negative effect on diagnostic accuracy? Academic Medicine. 2016;91(5):710-6.
26. Vella KM, Hall AK, van Merrienboer JJ, Hopman WM, Szulewski A. An exploratory investigation of the measurement of cognitive load on shift: application of cognitive load theory in emergency medicine. AEM Education and Training. 2021;5(4):e10634.
27. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 1: origins of bias and theory of debiasing. BMJ quality & safety. 2013;22(Suppl 2):ii58-ii64.
28. Pelaccia T, Sherbino J, Wyer P, Norman G. Diagnostic reasoning and cognitive error in emergency medicine: Implications for teaching and learning. Academic Emergency Medicine. 2025;32(3):320-6.
29. Kovacs G, Croskerry P. Clinical decision making: an emergency medicine perspective. Academic emergency medicine. 1999;6(9):947-52.
30. Kapuralalage TNE, Chan HF, Dulleck U, Hughes JA, Torgler B, Whyte S. Clinical decision-making: Cognitive biases and heuristics in triage decisions in the emergency department. The American Journal of Emergency Medicine. 2025;92:60-7.
31. Reale C, Salwei ME, Militello LG, Weinger MB, Burden A, Sushereba C, et al. Decision-making during high-risk events: a systematic literature review. Journal of Cognitive Engineering and Decision Making. 2023;17(2):188-212.
32. Chan TM, Mercuri M, Turcotte M, Gardiner E, Sherbino J, de Wit K. Making decisions in the era of the clinical decision rule: how emergency physicians use clinical decision rules. Academic Medicine. 2020;95(8):1230-7.

33. Fernandes M, Vieira SM, Leite F, Palos C, Finkelstein S, Sousa JM. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. Artificial intelligence in medicine. 2020;102:101762.
34. Choi A, Choi SY, Chung K, Chung HS, Song T, Choi B, et al. Development of a machine learning-based clinical decision support system to predict clinical deterioration in patients visiting the emergency department. Scientific Reports. 2023;13(1):8561.
35. Borbolla D, Ficheur G, Support SEftIYSoD. Clinical decision support systems and computerized provider order entry: contributions from 2020. Yearbook of Medical Informatics. 2021;30(01):172-5.
36. Kareemi H, Yadav K, Price C, Bobrovitz N, Meehan A, Li H, et al. Artificial intelligence–based clinical decision support in the emergency department: A scoping review. Academic Emergency Medicine. 2025;32(4):386-95.
37. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. Gastrointestinal endoscopy. 2020;92(4):807-12.
38. Hirani R, Noruzi K, Khuram H, Hussaini AS, Aifuwa EI, Ely KE, et al. Artificial intelligence and healthcare: a journey through history, present innovations, and future possibilities. Life. 2024;14(5):557.
39. Kachman MM, Brennan I, Oskvarek JJ, Waseem T, Pines JM. How artificial intelligence could transform emergency care. The American journal of emergency medicine. 2024;81:40-6.
40. Lee S, Lee J, Park J, Park J, Kim D, Lee J, et al. Deep learning-based natural language processing for detecting medical symptoms and histories in emergency patient triage. The American Journal of Emergency Medicine. 2024;77:29-38.
41. Tahayori B, Chini-Foroush N, Akhlaghi H. Advanced natural language processing technique to predict patient disposition based on emergency triage notes. Emergency Medicine Australasia. 2021;33(3):480-4.
42. Porto BM. Improving triage performance in emergency departments using machine learning and natural language processing: a systematic review. BMC Emergency Medicine. 2024;24(1):219.
43. Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. Bioengineering. 2023;10(12):1435.
44. Elhaddad M, Hamam S. AI-driven clinical decision support systems: an ongoing pursuit of potential. Cureus. 2024;16(4).
45. Lee S, Kang WS, Kim DW, Seo SH, Kim J, Jeong ST, et al. An artificial intelligence model for predicting trauma mortality among emergency department patients in South Korea: retrospective cohort study. Journal of medical internet research. 2023;25:e49283.
46. Hsu C-C, Kao Y, Hsu C-C, Chen C-J, Hsu S-L, Liu T-L, et al. Using artificial intelligence to predict adverse outcomes in emergency department patients with hyperglycemic crises in real time. BMC Endocrine Disorders. 2023;23(1):234.
47. Muhammad D, Bendechache M. Unveiling the black box: A systematic review of Explainable Artificial Intelligence in medical image analysis. Computational and structural biotechnology journal. 2024;24:542-60.
48. Rao A, Pang M, Kim J, Kamineni M, Lie W, Prasad AK, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow: development and usability study. Journal of Medical Internet Research. 2023;25:e48659.

49. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Frontiers in artificial intelligence. 2023;6:1169595.
50. Wu J, Ma Y, Wang J, Xiao M. The application of ChatGPT in medicine: a scoping review and bibliometric analysis. Journal of Multidisciplinary Healthcare. 2024:1681-92.
51. Omar M, Nassar S, Hijazi K, Glicksberg BS, Nadkarni GN, Klang E. Generating credible referenced medical research: A comparative study of openAI's GPT-4 and Google's gemini. Computers in Biology and Medicine. 2025;185:109545.
52. Bahir D, Zur O, Attal L, Nujeidat Z, Knaanie A, Pikkel J, et al. Gemini AI vs. ChatGPT: a comprehensive examination alongside ophthalmology residents in medical knowledge. Graefe's Archive for Clinical and Experimental Ophthalmology. 2025;263(2):527-36.
53. Tong L, Zhang C, Liu R, Yang J, Sun Z. Comparative performance analysis of large language models: ChatGPT-3.5, ChatGPT-4 and Google Gemini in glucocorticoid-induced osteoporosis. Journal of Orthopaedic Surgery and Research. 2024;19(1):574.
54. Alain G, Crick J, Snead E, Quatman-Yates CC, Quatman CE. Evaluating User Interactions and Adoption Patterns of Generative AI in Health Care Occupations Using Claude: Cross-Sectional Study. Journal of Medical Internet Research. 2025;27:e73918.
55. Jin H, Guo J, Lin Q, Wu S, Hu W, Li X. Comparative study of Claude 3.5-Sonnet and human physicians in generating discharge summaries for patients with renal insufficiency: assessment of efficiency, accuracy, and quality. Frontiers in digital health. 2024;6:1456911.
56. Kurokawa R, Ohizumi Y, Kanzawa J, Kurokawa M, Sonoda Y, Nakamura Y, et al. Diagnostic performances of Claude 3 Opus and Claude 3.5 Sonnet from patient history and key images in Radiology's "Diagnosis Please" cases. Japanese journal of radiology. 2024;42(12):1399-402.
57. Hosseini MM, Hosseini STM, Qayumi K, Ahmady S, Koohestani HR. The aspects of running artificial intelligence in emergency care; a scoping review. Archives of Academic Emergency Medicine. 2023;11(1):e38.
58. Weiner EB, Dankwa-Mullan I, Nelson WA, Hassanpour S. Ethical challenges and evolving strategies in the integration of artificial intelligence into clinical practice. PLOS digital health. 2025;4(4):e0000810.
59. Petersson L, Vincent K, Svedberg P, Nygren JM, Larsson I. Ethical considerations in implementing AI for mortality prediction in the emergency department: Linking theory and practice. Digital health. 2023;9:20552076231206588.
60. Levine DM, Tuwani R, Kompa B, Varma A, Finlayson SG, Mehrotra A, et al. The diagnostic and triage accuracy of the GPT-3 artificial intelligence model: an observational study. The Lancet Digital Health. 2024;6(8):e555-e61.
61. Ten Berg H, van Bakel B, van De Wouw L, Jie KE, Schipper A, Jansen H, et al. ChatGPT and generating a differential diagnosis early in an emergency department presentation. Annals of emergency medicine. 2024;83(1):83-6.
62. Hoppe JM, Auer MK, Strüven A, Massberg S, Stremmel C. ChatGPT with GPT-4 outperforms emergency department physicians in diagnostic accuracy: retrospective analysis. Journal of medical Internet research. 2024;26:e56110.

63. Wang J, Shue K, Liu L, Hu G. Preliminary evaluation of ChatGPT model iterations in emergency department diagnostics. Scientific reports. 2025;15(1):10426.

8. EKLER

EK 1- VERİ FORMU

HASTA VERİ FORMU

I. HASTA BİLGİLERİ

Adı Soyadı:	Cinsiyet: ☐ Erkek ☐ Kadın	Yaş:	HASTA BARKOD
Telefon:	Boy (cm):	Kilo (kg):	
Kullandığı İlaçlar:	Alerjiler:	Alkol Kullanımı: ☐ Evet ☐ Hayır	
	Sigara Kullanımı: ☐ Evet ☐ Hayır	Kronik Hastalıklar:	

II. BAŞVURU BİLGİLERİ

Başvuru Tarihi ve Saati: ___ / ___ / _____ - ____ : ____

Başvuru Şikayeti: _____

Şikayetin Başlangıç Süresi: _____

Ek Belirtiler: _____

Anamnez: _____

III. VİTAL BULGULAR VE ÖN TANI

Ateş (°C)	Kan Basıncı	Nabız (/dk)	Solunum (/dk)	Oksijen Sat.
___	___ / ___	___	___	___%

Doktorun İlk 5 Ön Tanısı:

1. _____
2. _____
3. _____
4. _____
5. _____

IV. FİZİK MUAYENE

Genel Durum / Sistem Bulguları: _____

V. TETKİKLER VE TANI

İstenen Testler: ☐ Hemogram ☐ Biyokimya ☐ Koagülasyon ☐ Kan Gazı ☐ Diğer: ____

Görüntüleme: ☐ X-ray ☐ USG ☐ BT ☐ MR ☐ EKG ☐ Diğer: ____

Tanılar:

1. _____
2. _____
3. _____

VI. KONSÜLTASYON VE TEDAVİ

Konsültasyon İstenen Bölümler: ☐ Kardiyoloji ☐ Dahiliye ☐ Nöroloji ☐ Genel Cerrahi ☐ Diğer: ____

Son Tanı: _____

Tedavi: ☐ Medikal ☐ Cerrahi ☐ Takip ☐ Yatış ☐ Diğer: ____

Taburculuk: ☐ Eve ☐ Servise ☐ Yoğun Bakıma ☐ Sevk

VII. DOKTOR BİLGİLERİ

Adı Soyadı: _____

Branşı: _____

İmza: _____