T.C.

ALTINBAŞ ÜNİVERSİTESİ

Lisansüstü Eğitim Enstitüsü

Protetik Diş Tedavisi Ana Bilim Dalı

YAPAY ZEKAYA DAYALI BÜYÜK DİL MODELLERİNİN TEMPOROMANDİBULAR EKLEM DÜZENSİZLİKLERİNİ TEŞHİS ETME PERFORMANSLARININ DEĞERLENDİRİLMESİ

Erman DİREKCİ

Doktora Tezi

Danışman

Doç. Dr. Rana TURUNÇ

İstanbul, 2025

YAPAY ZEKAYA DAYALI BÜYÜK DİL MODELLERİNİN TEMPOROMANDİBULAR EKLEM DÜZENSİZLİKLERİNİ TEŞHİS ETME PERFORMANSLARININ DEĞERLENDİRİLMESİ

Erman DİREKCİ

Protetik Diş Tedavisi Ana Bilim Dalı

Doktora Tezi

ALTINBAŞ ÜNİVERSİTESİ

2025

“Erman DİREKCİ” tarafından hazırlanmış ve 18.06.2025 tarihinde sunulmuş “YAPAY ZEKAYA DAYALI BÜYÜK DİL MODELLERİNİN TEMPOROMANDİBULAR EKLEM DÜZENSİZLİKLERİNİ TEŞHİS ETME PERFORMANSLARININ DEĞERLENDİRİLMESİ” başlıklı tez Protetik Diş Tedavisi Ana Bilim Dalında Doktora Tezi olarak **oy birliği** ile kabul edilmiştir.

Doç. Dr. Rana TURUNÇ
Danışman

Tez Savunma Sınavı Jüri Üyeleri:

Doç. Dr. Ayşe ATAY	Klinik Bilimler Bölümü, Altınbaş Üniversitesi	_____
Doç. Dr. Meltem MERT EREN	Klinik Bilimler Bölümü, İstanbul Sağlık ve Teknoloji Üniversitesi	_____
Dr. Öğr. Üyesi Demet Çağıl AYVALIOĞLU ŞAMİLOĞLU	Klinik Bilimler Bölümü, İstanbul Sağlık ve Teknoloji Üniversitesi	_____
Dr. Öğr. Üyesi Beril KOYUNCU	Klinik Bilimler Bölümü, Altınbaş Üniversitesi	_____
Dr. Öğr. Üyesi Elif Ezgi OĞUZ	Klinik Bilimler Bölümü, Altınbaş Üniversitesi	_____

Bu tezin Doktora tezi olarak bütün şartları sağladığını beyan ederim.

Bu tezdeki tüm bilgilerin akademik kurallara ve etik davranışlara uygun olarak edinildiğini ve sunulduğunu beyan ederim. Ayrıca, bu kuralların ve davranışların gerektirdiği şekilde, bu çalışmada, orijinal olmayan tüm materyalleri ve sonuçları tamamen alıntı yaptığımı ve referans gösterdiğimi de beyan ederim.

Erman DİREKCİ

İmzası

İTHAF

Araştırmanın her aşamasında desteğini esirgemeyen değerli danışman hocama, tez sürecinde beni motive eden aileme ve tüm yakınlarıma teşekkür ederim.

ÖNSÖZ

Tez çalışmam boyunca elde ettiğim bilgi ve deneyimlerin, akademik yaşamıma olduğu kadar mesleki ve kişisel gelişimime de büyük katkılar sağladığını düşünüyorum. Karşılaşılan zorluklar ve elde edilen başarılar, bu yolculuğu benim için unutulmaz kılmıştır.

Çalışmanın ortaya konulmasında emeği geçen herkese tekrar teşekkür eder, çalışmamın alanında katkı sağlamasını temenni ederim.

Saygılarımla,

ÖZET

YAPAY ZEKAYA DAYALI BÜYÜK DİL MODELLERİNİN TEMPOROMANDİBULAR EKLEM DÜZENSİZLİKLERİNİ TEŞHİS ETME PERFORMANSLARININ DEĞERLENDİRİLMESİ

DİREKCİ, Erman

Doktora Programı, Protetik Diş Tedavisi Anabilim Dalı, Altınbaş Üniversitesi

Danışman: Doç. Dr. Rana TURUNÇ

Tarih: 06/2025

Sayfa: 103

AMAÇ: Bu çalışmanın amacı, temporomandibular eklem düzensizliklerinin (TMD) tanısında sekiz büyük dil modelinin (ChatGPT-3.5, ChatGPT-4, ChatGPT-4o, Microsoft Copilot, Anthropic Claude 2, Docus AI, Perplexity AI ve Google Gemini) zamana ve dile bağlı performanslarını değerlendirmektir.

GEREÇ VE YÖNTEM: Büyük dil modellerine ilki 2024 Eylül olmak üzere 4 ay arayla Temporomandibular Düzensizlikler için Tanı Kriterleri (TMD/TK) karar ağacı temel alınarak hazırlanan 13 olgu sorusu Türkçe ve İngilizce olarak sorulmuştur. Modellerin yanıtları karar ağacındaki doğru tanıyla karşılaştırılarak değerlendirilmiş; spesifik doğru yanıtlar 1, hatalı, eksik veya çok genel yanıtlar 0 puan olarak kaydedilmiştir. Her model için doğru yanıt oranları hesaplanmış ve farklılıklar Cochran's Q ve McNemar testleri ile istatistiksel olarak analiz edilmiştir.

BULGULAR: Modellerin TMD tanısındaki doğruluk oranları arasında belirgin farklar gözlenmiştir. Her iki değerlendirme döneminde de modeller arası performans farklılıkları istatistiksel olarak anlamlı bulunmuştur ($p<0,05$). ChatGPT-4, ChatGPT-4o ve Anthropic Claude 2, %70–85 düzeyinde doğru tanı oranlarıyla en iyi performansı sergilemiştir. Buna karşın Perplexity AI, Gemini ve Microsoft Copilot'un en düşük doğruluk oranlarına sahip modeller olduğu görülmüştür (<%40). Modellerin büyük çoğunluğu Türkçe ve İngilizce sorularda benzer başarı göstermiştir; özellikle GPT-4 tabanlı modeller ve Claude 2 her iki

dilde de benzer doğruluklara ulaşmıştır. Ancak Gemini modeli Türkçe sorularda oldukça düşük bir performans sergilerken (Ocak ayında %0 doğru), aynı model İngilizce sorularda anlamlı ölçüde daha yüksek doğruluk elde etmiştir.

SONUÇ: Sonuç olarak, büyük dil modellerinin TMD tanısındaki başarıları modelden modele önemli ölçüde değişmektedir. GPT-4 tabanlı gelişmiş modeller ile Claude 2, yüksek doğruluk ve çift dilli tutarlılık göstererek TMD tanısında klinik karar desteği aracı olma potansiyelini ortaya koymuştur. Buna karşılık bazı modellerin (ör. Gemini, Copilot) düşük performansı, bu teknolojilerin henüz tutarlı ve güvenilir olmadığını göstermekte ve daha fazla geliştirme gereksinimine işaret etmektedir. Genel olarak bulgular, büyük dil modellerinin TMD tanısında yardımcı olabileceğini ancak yaygın klinik kullanım öncesinde model seçimi, dil farklılıkları ve sürekli model iyileştirmelerinin dikkate alınması gerektiğini ve konu insan sağlığı olduğu için tamamen büyük dil modellerine tamamen güvenmenin henüz doğru olmayacağını göstermektedir.

Anahtar Kelimeler: TMD/TK, TMD, Büyük Dil Modeli, Yapay Zeka, ChatGPT, Gemini. Copilot.

ABSTRACT

EVALUATION OF THE DIAGNOSTIC PERFORMANCE OF ARTIFICIAL INTELLIGENCE-BASED LARGE LANGUAGE MODELS IN TEMPOROMANDIBULAR JOINT DISORDERS

DİREKCİ, Erman

PhD, Department of Prosthodontics, Altınbaş University

Supervisor: Assoc. Prof. Dr. Rana TURUNÇ

Date: 06/2025

Pages: 103

AIM: The aim of this study is to evaluate the time- and language-dependent performances of eight large language models (ChatGPT-3.5, ChatGPT-4, ChatGPT-4o, Microsoft Copilot, Anthropic Claude 2, Docus AI, Perplexity AI, and Google Gemini) in the diagnosis of temporomandibular disorders (TMD).

MATERIALS AND METHOD: Starting in September 2024, a set of 13 case-based questions, prepared based on the Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) decision tree, were asked to the large language models in both Turkish and English at four-month intervals. The models' responses were evaluated by comparing them to the correct diagnoses in the decision tree; specifically accurate answers were scored as 1, while incorrect, incomplete, or overly general answers were scored as 0. The accuracy rates for each model were calculated, and differences were statistically analyzed using Cochran's Q and McNemar tests.

RESULTS: Significant differences were observed in the accuracy rates of the models in diagnosing TMD. In both evaluation periods, the performance differences between the models were found to be statistically significant ($p<0.05$). ChatGPT-4, ChatGPT-4o, and Anthropic Claude 2 demonstrated the best performance, with correct diagnosis rates ranging from 70% to 85%. In contrast, Perplexity AI, Gemini, and Microsoft Copilot showed the lowest accuracy rates (<40%). The majority of the models performed similarly in both

Turkish and English cases; in particular, GPT-4-based models and Claude 2 achieved comparable accuracy in both languages. However, the Gemini model exhibited notably poor performance in Turkish cases (0% accuracy in January), while achieving significantly higher accuracy in English cases.

CONCLUSION: In conclusion, the diagnostic performance of large language models in identifying TMD varies significantly between models. Advanced GPT-4-based models and Claude 2 demonstrated high accuracy and bilingual consistency, highlighting their potential as clinical decision support tools in TMD diagnosis. In contrast, the low performance of some models (e.g., Gemini, Copilot) indicates that these technologies are not yet consistent or reliable, underscoring the need for further development. Overall, the findings suggest that while large language models can assist in TMD diagnosis, factors such as model selection, language differences, and ongoing model improvements must be considered before widespread clinical implementation. Given the importance of human health, it is also premature to rely entirely on large language models.

Keywords: DC/TMD, TMD, Large Language Model, Artificial Intelligence, ChatGPT, Gemini, Copilot.

İÇİNDEKİLER

Sayfa

TABLO LİSTESİ

Sayfa

ŞEKİL LİSTESİ

Sayfa

KISALTMALAR

BDM	:	Büyük Dil Modeli
DC/TMD	:	Diagnostic Criteria for Temporomandibular Disorders
LLM	:	Large Language Model
TMD	:	Temporomandibular Eklem Düzensizlikleri
TMD/TK	:	Temporomandibular Düzensizlikler için Tanı Kriterleri
TME	:	Temporomandibular Eklem

1. GİRİŞ

Temporomandibular eklem düzensizlikleri, çiğneme kasları, temporomandibuler eklem (TME) ve bağlantılı yapılarla ilgili çeşitli klinik sorunları kapsamaktadır. Bu düzensizlikler, çiğneme kaslarında, TME'de ve ilişkili sert ve yumuşak dokularda ağrıyla kendini gösterebilmektedir. Bunun yanı sıra, alt çenenin hareket aralığında kısıtlanma veya sapma, TME'den gelen sesler ve baş ağrıları ile yüz ağrısı gibi diğer semptomlar da bulunabilmektedir [1].

Temporomandibular eklem düzensizliklerinin teşhisinde, klinik muayene yöntemleri ile radyografik görüntüleme, ultrasonografi ve artrografi gibi ileri görüntüleme teknikleri yaygın olarak kullanılmaktadır. Klinik muayene ve görüntüleme tekniklerinin bir arada kullanımı, temporomandibular eklem düzensizliklerinin kapsamlı bir şekilde değerlendirilmesini sağlar. Çiğneme kaslarının ve temporomandibuler eklemin (TME) klinik muayenesi genellikle kasların detaylı bir biçimde elle muayenesini, TME seslerinin dinlenmesini, alt çenenin hareket aralığının ölçülmesini ve ağrının değerlendirilmesini içermektedir ve bu yaklaşımlar, doğru teşhis konulması ve kişiye özgü etkili tedavi planlarının geliştirilmesi açısından kritik öneme sahiptir [2], [3], [4]. Bu tür değerlendirme genellikle uzman hekimler tarafından gerçekleştirilmektedir [5].

Temporomandibular eklem düzensizliklerinde semptomların değerlendirilmesi düzensizliklerin doğru teşhis edilebilmesi açısından oldukça önemlidir. Semptomlar hangi teşhis testlerinin gerektiği konusunda klinik kararlara yardımcı olmaktadır. Semptomlar, nihai teşhis için tam olarak anlaşılmalıdır [6]. 1934'te ortaya çıkarılan temporomandibular disfonksiyonla ilişkilendirilen semptomlar, çene etrafında yaşanan ağrı, işitme kaybı, baş ağrısı, baş dönmesi, dil, burun ve sinüslerde yanma hissi, kulak çınlaması gibi çeşitli şikayetleri içermektedir. Ayrıca çene hareketlerinin kısıtlanması, konuşma ve yutma güçlükleri, uyku düzensizlikleri ve diş sıkmadan kaynaklı sorunlar da gözlemlenebilmektedir. Tüm bu semptomlar göz önünde bulundurularak doğru tanıyı koymak oldukça önemlidir [7][8].

Günümüzde yapay zeka sağlık alanında teşhis sürecini etkileyen önemli bir teknoloji olarak göze çarpmaktadır. Özellikle medikal görüntüleme ve veri analizi alanında yapay zekanın getirdiği yenilikler hastalıkların daha hızlı ve doğru teşhis edilmesini sağlamaktadır. Yapay

zeka uygulamaları röntgen, MR, bilgisayarlı tomografi gibi radyolojik görüntüler üzerindeki anormallikleri inceleyerek veya semptomların analizini yaparak hastalığın belirlenmesinde yardımcı olmaktadır. Bununla birlikte yapay zeka uygulamaları hekimlerin iş yükünü azaltmaktadır. Sağlık sektöründe yapay zeka uygulamalarının yaygınlaşması ve bu teknolojilerin sürekli geliştirilmesiyle, teşhis süreçlerindeki iyileştirmelerin sağlık hizmetlerinin kalitesini artırması ve tedavi başarısına olumlu katkılar sağlaması amaçlanmaktadır [9], [10], [11], [12], [13].

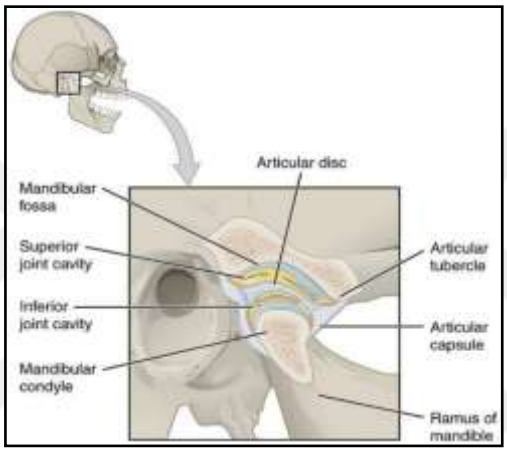
Sağlık alanındaki yapay zekâ uygulamaları, yalnızca görüntüleme temelli sistemlerle sınırlı kalmayıp, dil tabanlı yapılar aracılığıyla klinik karar destek mekanizmalarına entegre edilerek, büyük dil modellerinin bu süreçlerde ön plana çıkmasına olanak tanımaktadır. Bu modeller, internet üzerindeki geniş çaplı verilerden, özellikle tıbbi içeriklerden beslenerek eğitilmekte ve özelleştirilmektedir. Bu durum, onların geleneksel dil işleme araçlarına kıyasla daha güçlü bir dil anlama, örüntü tanıma ve ilişkilendirme kapasitesi geliştirmesine olanak tanımaktadır [14], ancak sağlık alanında etkin kullanılabilmeleri için doğruluk performanslarının değerlendirilmesi şarttır.

Bu çalışmanın amacı, temporomandibular kas ve eklem düzensizlikleriyle ilgili teşhis amaçlı kullanılan standartlaştırılmış anket yöntemi olan 'Temporomandibular Eklem Düzensizlikleri / Tanı Kriterleri'ne (Diagnostic Criteria of Temporomandibular Disorders - DC/TMD) göre günümüzde yaygınlaşan büyük dil modellerinin temporomandibuler eklem düzensizliklerini teşhis etmedeki performanslarını Türkçe ve İngilizce olarak değerlendirmek, birbirleriyle karşılaştırmak ve zaman içerisindeki değişimini gözlemlemektir. Buna bağlı olarak sıfır hipotezi “Yapay zekaya dayalı büyük dil modellerinin temporomandibuler eklem düzensizliklerini teşhis etme performansları arasında anlamlı farklılık yoktur ve büyük dil modellerinin performansı farklı bir dil kullanıldığında veya farklı zamanlarda yapılan değerlendirmelerde değişmez” olarak belirlenmiştir.

2. GENEL BİLGİLER

2.1 TEMPOROMANDİBULAR EKLEM

TME (şekil 2.1), alt çene kondili ile temporal kemiğin glenoid fossası arasında birleşen dinamik bir yapıdır. Bu eklemin içinde, kondil ile fossa arasında hareketi kolaylaştıran ve sürtünmeyi azaltan bir artiküler disk bulunmaktadır. Çene hareketleri, kondil ve disk arasındaki dönme hareketiyle (rotasyon) ve kondil-disk kompleksi ile fossa arasındaki kayma hareketiyle (translasyon) oluşmaktadır [15] .



Şekil 2.1: Temporomandibular Eklem Osteolojisi [16].

Temporomandibuler eklem (TME), ginglymoartrodial bir yapıya sahip olup geniş bir hareket aralığına izin vermektedir. 'Ginglimus' terimi, eklemin bir menteşe gibi çalıştığını ve tipik olarak sadece ileri-geri hareketlerle ilişkilendirildiğini ifade etmektedir. Fakat TME'nin işlevi yalnızca bu hareketle sınırlı değildir. 'Artrodial' terimi ise eklemin yüzeylerinin kayarak birbirine hareket ettiğini tanımlamaktadır. Bu, TME'nin sadece dönme hareketiyle değil aynı zamanda kayma hareketiyle de çalıştığı anlamına gelmektedir [17].

Alt çene hareketleri, yalnızca kemik yapılar, kaslar ve bağ dokularla sınırlı kalmayıp, aynı zamanda dişlerin oklüzal ilişkilerinden de etkilenmektedir. Bu hareketlerin koordinasyonunda, her iki temporomandibular eklem mandibula üzerinde birleşerek birlikte işlev görmektedir [18]. Temporomandibular eklem, mandibula kondili ile temporal kemikte bulundan mandibular fossa arasında yer alan kompleks bir eklemdir. Bu eklemin sağlıklı bir şekilde çalışabilmesi için kondil, disk ve fossanın birbirleriyle uyum içerisinde fonksiyon göstermesi gerekmektedir. Kondil-disk-fossa'nın ideal ilişkisi, kondillerin kendi disklerinin

en ince ve damarsız kısmıyla temasta olduğu ve kondil-disk kompleksinin artiküler eminense karşı en üst ve önde konumlandığı maksillomandibular ilişkidir. [19], [20], [21].

2.2 TEMPOROMANDİBULAR DÜZENSİZLİKLERİN TARİHÇESİ

Tarihsel araştırmalar temporomandibular düzensizliklerin tedavi çalışmalarının mandibular dislokasyonların manuel olarak yerine yerleştirilmesiyle antik Mısırlılar tarafından başladığını öne sürmektedir [22]. Hipokrat ise M.Ö. 5. yüzyılda günümüzde kullanılan yöntemlere çok benzer şekilde mandibular dislokasyon tedavisi yapmıştır. [23].

Çene cerrahları 1800'lü yıllardan beri eklem ve çene ağrılarını gidermeyi amaçlamışlardır ve bu amaçla 20. yüzyılın başlarında eklem içi diski çıkararak eklem düzensizliklerini tedavi etmeyi denemişlerdir. [24]. Bununla birlikte ilerleyen dönemde diş hekimleri, kulak burun boğaz uzmanları ve anatomi uzmanları temporomandibular düzensizliklerin baş, yüz kulak ve çene semptomlarını diş arkı içinde arka bölgede diş kaybına bağlı oluşan atrofi ve dikey boyut kaybıyla açıklamışlardır [25]. Bunun üzerine 1934'te Costen, hastaların protezlerinin yenilenmesinin çene çevresindeki ağrı, baş ağrısı, kulak ağrısı gibi rahatsızlıkları düzeltebileceğini belirtmiştir. Temporomandibular eklem düzensizliklerine bağlı işitme bozukluğu, kulak tıkanıklığı hissi, kulak ağrısı, kulak çınlaması, baş dönmesi, sinüslerde hissedilen ağrı, boğaz ve dilde yanma hissi, baş ağrısı ve trismus gibi semptomlar "Costen Sendromu" olarak adlandırılmıştır.

Shore, 1959 yılında "Temporomandibular Disfonksiyon Sendromu" terimini tanımlarken, Schulte ise 1969'da "Miyoartropati" terimini kullanmıştır. Ardından, 1971 senesinde Ramfjord ve Ash, "fonksiyonel temporomandibular eklem bozuklukları" teriminin kullanımını önermişlerdir [26], [27], [28]. 1982 yılında Amerikan Diş Hekimleri Birliği (American Dental Association) tarafından alınan bir kararla "Temporomandibular Disorders" teriminin kullanımında karar kılınmıştır [29]. 2014 yılında International Association of Dental Research (IADP)'ın aldığı kararda da '' Temporomandibular Disorders'' terimi kullanılmaya devam etmiş [30] ve bu terim Türkçeye ''Temporomandibular Düzensizlikler'' olarak çevirilmiştir.

2.3 TEMPOROMANDİBULAR DÜZENSİZLİKLERİN ETİYOLOJİSİ

Temporomandibular düzensizliklerin etiyolojisi hakkında bir çok çalışma yapılmıştır ancak hala spesifik nedenler belirlenememiştir ve bu konudaki tartışmalar devam etmektedir. Bu düzensizlikler çeşitli faktörlerin bir araya gelmesiyle (Biyomekanik, nöromusküler, biyopsikososyal ve nörobiyolojik faktörler) meydana gelmektedir. Bu faktörler, predispozan (yapısal, metabolik ve/veya psikolojik koşullar), tetikleyici (örneğin, travma veya çiğneme sisteminin tekrarlayan yüklenmesi) ve ağırlaştırıcı (parafonksiyon, hormonal veya psikososyal faktörler) olarak sınıflandırılmaktadır [31], [32], [33], [34].

Predispozan faktörler bireyin belirli bir sorun geliştirme hassasiyetini artıran morfolojik, fizyolojik, psikolojik ve çevresel değişkenlerin bir kombinasyonunu içermektedir. Tetikleyici faktörler travma, stres, hiperfonksiyon ve doğal inhibe edici faktörlerin kombinasyonunu içermektedir. Ağırlaştırıcı faktörler ise zayıf iyileşme potansiyeli, etiyolojik faktörlerin kontrol edilememesi, hastalık sürecinin negatif etkileri ve yanlış uygulanmış tedavilerin olumsuz etkileri gibi durumları kapsamaktadır [35].

Geçmişte, çoğu hekim temporomandibular düzensizliklerin tedavisinde çözümü hastanın oklüzyonunu değiştirmekte aramıştır. Bu yöntem başarısız olursa, hekimler hastayı ciddi psikolojik sorunlar yaşayan bireyler olarak değerlendirmişlerdir. Daha sonraki yıllarda, temporomandibular düzensizliklerin en az beş farklı etiyolojik faktörü olduğu belirlenmiştir. Oklüzyon durumu (derin kapanış, Angle Sınıf II maloklüzyon, çapraz kapanış, eksentrik temaslar),oklüzyondaki akut bir değişiklik, eklemin aşırı yüklenmesi, travma, duygusal stres, derin ağrı girişi ve bruksizm gibi parafonksiyonel aktiviteler bu etiyolojik faktörlerden bazılarıdır ve doğru tedavi için önce temporomandibular düzensizliğin etiyolojisinin belirlenmesi gerekmektedir [36].

2.4 TEMPOROMANDİBULAR DÜZENSİZLİKLERİN EPİDEMİYOLOJİSİ

Toplumda temporomandibular düzensizliklerin epidemiyolijisi, tanı kriterlerindeki değişiklikler nedeniyle tartışmalıdır. Ancak genel kanı, temporomandibular düzensizliklerin semptomlarının toplumda yaygın olduğu yönündedir [37]. Yapılan çalışmalarda toplumun %1 ile %75'i en az 1 kere temporomandibular düzensizliklerin objektif semptomlarını gösterirken %5 ile %33'ü subjektif semptom göstermiştir [38]. Temporomandibular düzensizliklerin semptomları daha çok 20-40 yaş arasında görülürken yaşlılarda ve

çocuklarda daha nadir görülmektedir. Kadınlarda ise erkeklere göre daha sık rastlanmaktadır [39].

Farklı teşhis protokolleri kullanılarak yapılan önceki çalışmalar, temporomandibular düzensizliklerin teşhisinde farklı ve tutarsız sonuçlar ortaya koymuştur. Ancak 1992'de "Research Diagnostic Criteria for Temporomandibular Disorders" (RDC/TMD) – ‘’Temporomandibular Rahatsızlıklar Araştırma Teşhis Kriterleri’’ (TMR/ATK) yayınlanmasıyla standart bir tanı kriteri belirlenmiş ve tutarlı çalışmalar yürütülmüştür [40].

2.5 TEMPOROMANDİBULAR DÜZENSİZLİKLERİN SINIFLANMA TARİHÇESİ

Temporomandibular düzensizlikler ile ilgili onlarca yıldır bir çok sınıflama yapılmıştır. 1987 yılında Amerikan Orofasial Ağrı Akademisi (AAOP) temporomandibular düzensizlikleri, çiğneme kası rahatsızlıkları, temporomandibular eklem rahatsızlıkları, kronik mandibular hipomobilite ve gelişimsel rahatsızlıklar olarak kategorilenmiştir. Bu sınıflama 1993 yılında McNeill ve 1996 yılında Okeson tarafından aşağıdaki şekilde tekrar düzenlenmiştir [41], [42].

Temporomandibular Rahatsızlıklar

a. Çiğneme Rahatsızlıkları
    i. Koruyucu Kas Kasılması
    ii. Lokal Kas Hassasiyeti
    iii. Miyofasiyal Ağrı
    iv. Miyospazm
    v. Miyalji
b. Temporomandibular Eklem Rahatsızlıkları
    i. Kondil Disk Kompleksi Düzensizlikleri
        i. Disk Deplasmanları
        ii. Redüksiyonlu Disk Dislokasyonu
        iii. Redüksiyonsuz Disk Dislokasyonu
    ii. Eklem Düzeylerinin Yapısal Bozuklukları
        i. Şekil Sapmaları
            i. Disk

ii. Kondil
iii. Fossa
ii. Adezyon
i. Diskin Kondile
ii. Diskin Fossaya
iii. TME'nin İltihapsal Rahatsızlıkları
i. Sinovit / Kapsülit
ii. Redrodiskit
iii. Artrit
i. Osteoartrit
ii. Osteartroz
iii. Poliartrit
iv. İlgili Yapıların İltihapsal Rahatsızlıkları
i. Temporal Tendonit
ii. Stilomandibular Ligaman İltihabı

c. Kronik Mandibular Hipomobilite
i. Ankiloz
i. Fibröz
ii. Kemiksel
ii. Kas Kontraktürü
i. Miyostatik
ii. Miyofibrotik
iii. Koronoid Engellemesi

d. Gelişimsel Rahatsızlıklar
i. Doğumasl ve Gelişimsel Kemik Rahatsızlıkları
i. Agenezi
ii. Hipoplazi
iii. Hiperplazi
iv. Neoplazi
i. Doğumsal ve Gelişimsel Kas Rahatsızlıkları
i. Hipotrofi
ii. Hipertrofi

iii. Neoplazi

Okeson ve McNeill'in düzenlediği Temporomandibular Rahatsızlıklar sınıflaması yıllarca kullanılmış olsa da, zamanla daha kapsamlı bir yaklaşıma ihtiyaç duyulmuştur. Yeni bir sınıflamaya duyulan ihtiyaç, hastalıkların daha ayrıntılı alt kategorilere ayrılmasının gerekliliğinden kaynaklanmaktadır; çünkü eski sınıflama sistemi daha çok fiziksel semptomları ele almaktadır, psikososyal faktörleri teşhis sürecine dahil etmemektedir ve spesifik teşhislerin konulmasında yetersiz kalmaktadır [43], [44], [45].

2.5.1 Temporomandibular Düzensizlikler için Teşhis Kriterleri (TMD/TK) (DC/TMD)

Temporomandibular düzensizlikler için hem klinik ortamda hem de araştırmalar için basit, anlaşılır, güvenilir ve geçerli tanı kriterlerine ihtiyaç vardır; bu kriterler, hastalık öyküsü, muayene ve görüntüleme prosedürlerini kapsamalıdır. Ayrıca, hastanın ağrı sürecindeki psikolojik durumu ve psikososyal davranışlarının değerlendirilmesi tanı sürecinin temel bir parçasıdır ve özellikle kronik ağrı varlığında hastanın psikolojik durumu sorgulanmalıdır. Bu nedenle, temporomandibular düzensizliklerin doğru bir şekilde değerlendirilmesi için yapılandırılmış ve bilimsel temellere dayanan tanı sistemlerine ihtiyaç duyulmuştur.

Temporomandibular Düzensizlikler İçin Teşhis Kriterleri (TMD/TK), temporomandibular düzensizliklerin tanı ve sınıflandırılmasında kullanılan güncel, kapsamlı ve bilimsel temelli bir sistemdir. 2014 yılında geliştirilen bu kriterler, önceki sistem olan Araştırma Teşhis Kriterleri (TMD/ATK) ile karşılaşılan sınırlılıkları aşmak ve daha standartlaştırılmış bir tanı yaklaşımı sunmak amacıyla oluşturulmuştur. TMD/TK, TMD/ATK'nın güncellenmiş ve yeniden yapılandırılmış versiyonu olup hem klinik pratikte hem de araştırma ortamlarında kullanılmak üzere tasarlanmıştır. TMD/ATK daha çok araştırmalarda kullanılmaya uygunken, TMD/TK geniş bir uygulama alanı sunarak klinik karar süreçlerinde de güvenilir bir araç haline gelmiştir. Duyarlılık oranı %70'in, özgüllük oranı ise %95'in üzerinde olan bu sistem, temporomandibular bozuklukların doğru şekilde teşhis edilmesini kolaylaştırmakta, tanısal güvenilirliği artırmakta ve yorum farklılıklarını en aza indirmektedir. Ayrıca TMD/TK, yalnızca mekanik faktörlere odaklanan geleneksel yaklaşımların ötesine geçerek, ağrının bilişsel, duygusal ve davranışsal yönlerini içeren biyopsikososyal modeli temel almasıyla, temporomandibular düzensizliklerin etiyolojisine yönelik anlayışta önemli bir paradigma değişikliğine öncülük etmektedir [46] [47].

Yeni çift eksenli Temporomandibular Düzensizlikler için Tanı Kriterleri (TMD/TK), hekimlere hastaları değerlendirirken kullanabilecekleri kanıta dayalı kriterler sunmakta ve tedavi sürecini kolaylaştırmaktadır. Hekimler ve araştırmacılar aynı kriterleri, sınıflandırmayı ve terimleri kullandıklarında, klinikte yapılan teşhis amaçlı sorular ve klinik deneyim daha kolay paylaşılabilir ve erişilebilir hale gelmektedir. Bu durum hekimlere hastalıkları daha rahat tanımlama ve hastalık sürecini daha etkin bir şekilde yönetme imkanı sağlamaktadır [30].

TMD/TK teşhis sistemi, 1. Ekseninde fiziksel değerlendirme yaparken 2. Ekseninde ise hastanın psikososyal durumunu ve ağrıyla ilişkili fonksiyon kaybını ele almaktadır. Araştırmalar, psikososyal ve davranışsal risk faktörlerinin uzun süreli ve kronik ağrılara yol açabileceğini ve tedavi etkinliğini azaltabileceğini göstermektedir. Bu nedenle, temporomandibular düzensizliklerin tedavisinde psikososyal faktörlerin ve psikolojik etkenlerin göz ardı edilmemesi gerektiği vurgulanmaktadır [40], [48], [49], [50].

2.5.2 Temporomandibular Düzensizlikler için Teşhis Kriterleri'ne Göre Sınıflama

TMD/TK belirli temporomandibular düzensizliklerin teşhisinde kullanılabilecek kriterleri içermektedir. Bu sınıflamada sık görülen temporomandibular düzensizlikler bulunmaktadır, ancak, sınıflandırmada yer almayan daha nadir görülen rahatsızlıkların varlığı göz ardı edilmemektedir, bu rahatsızlıklar için henüz güvenilir kaynaklar ve standart yönlendirmeler bulunmadığı için bunlar sınıflandırmada yer almamaktadır [51].

Temporomandibular Düzensizlikler için Teşhis Kriterleri (TMD/TK) Tanı karar ağacı şemasında (şekil 2.2-şekil 2.5) bulunan düzensizlikler aşağıda sıralanmıştır [47].

a. Ağrı ile İlişkili Temporomandibular Düzensizlikler ve Baş Ağrısı
   i. Kas Ağrısı
   ii. Lokal Kas Ağrısı
   iii. Kas-fasya Ağrısı
   iv. Kas-Fasya Ağrısı (yansıyan)
   v. Eklem Ağrısı
   vi. Temporomandibular Düzensizliğe Bağlı Baş Ağrısı

b. Eklem İçi Düzensizlikler
   i. Redüksiyonlu Disk Deplasmanı

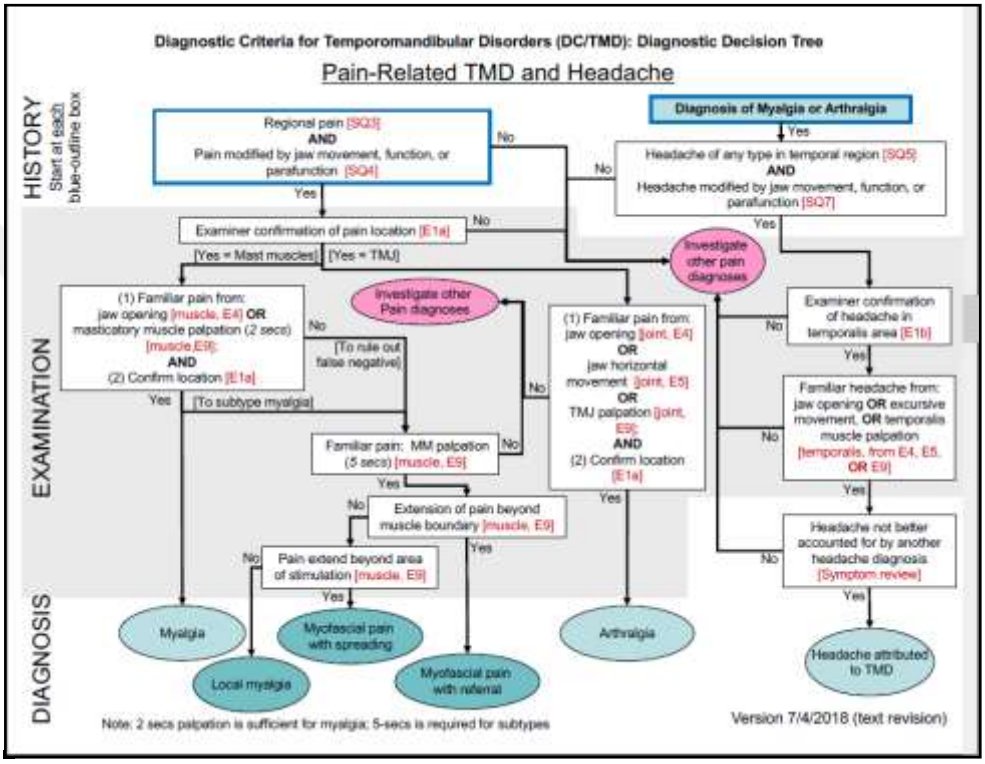
ii. Redüksiyonlu Disk Deplasmanı, Aralıklı Kilitlenme Olan

iii. Redüksiyonsuz Disk Deplasmanı Kısıtlı Ağız Açıklığı Olan

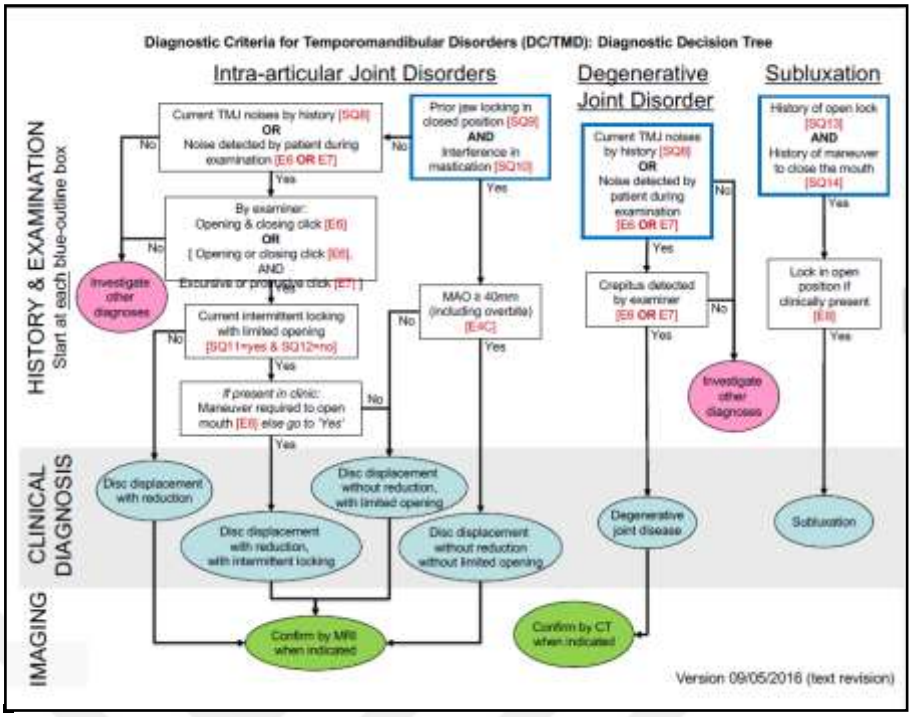
iv. Redüksiyonsuz Disk Deplasmanı Kısıtlı Ağız Açıklığı Olmayan

c. Dejeneratif Eklem Hastalığı

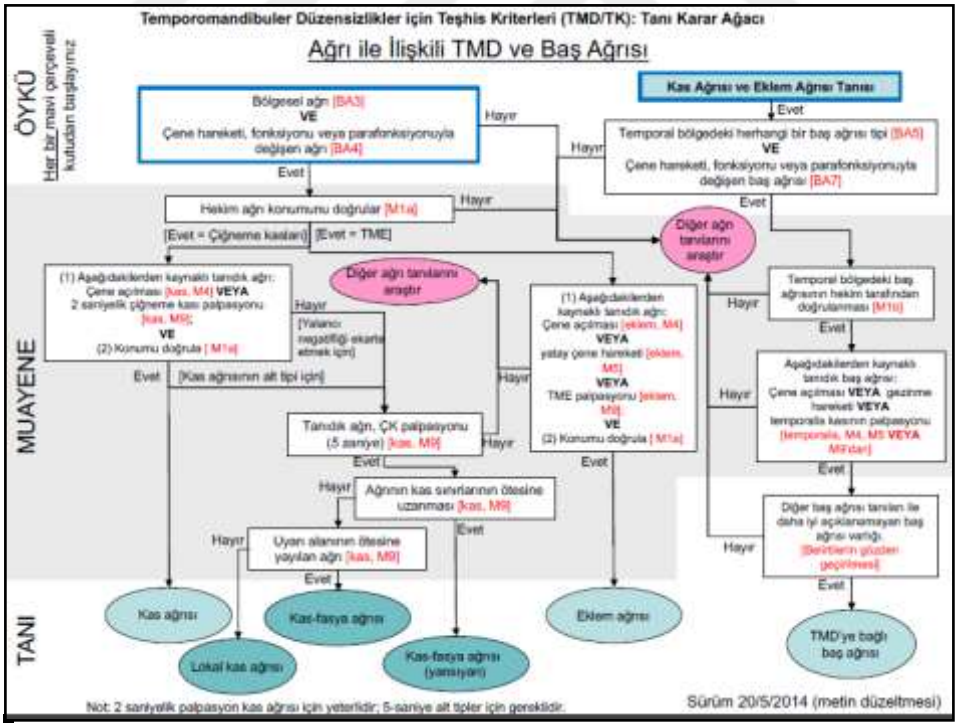
d. Subluksasyon



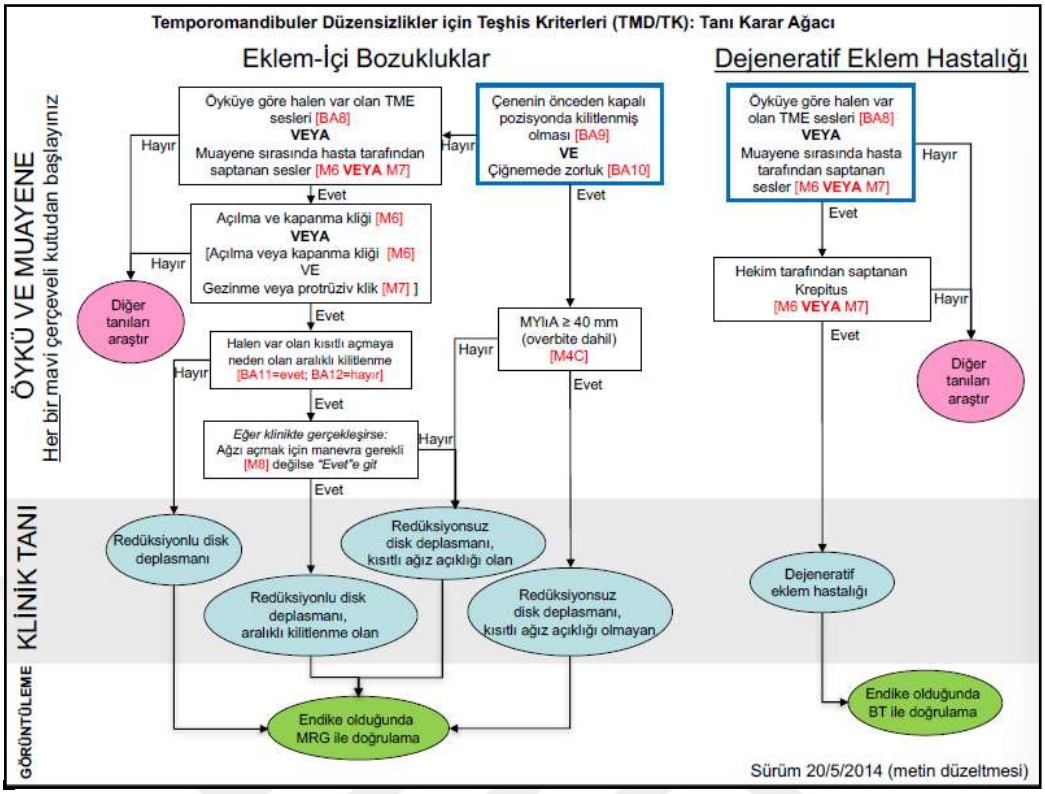
Şekil 2.2: DC/TMD Tanı Karar Ağacı İngilizce [47].



Şekil 2.3: DC/TMD Tanı Karar Ağacı İngilizce [47].



Şekil 2.4: TMD/TK Tanı Karar Ağacı Türkçe [52].



Şekil 2.5: TMD/TK Tanı Karar Ağacı Türkçe [52].

2.6 TEMPOROMANDİBULAR DÜZENSİZLİKLER

2.6.1 Ağrı İle İlişkili Temporomandibular Düzensizlikler ve Baş Ağrısı

2.6.1.1 Kas ağrısı

Temporomandibular kas ağrısı (miyalji), çiğneme kaslarıyla ilişkili çeşitli rahatsızlıkları kapsayan ve temporomandibular eklem disfonksiyonu ile sıklıkla ilişkili olan bir düzensizliktir. Bu durum, genellikle çiğneme kaslarında ağrı ve hassasiyet, çene hareketlerinde kısıtlılık ve kasılmalar gibi semptomlarla kendini göstermektedir. Etiyolojik açıdan değerlendirildiğinde, temporomandibular kas ağrısının nedenleri arasında kas yapısındaki anatomik anormallikler, çiğneme kaslarının aşırı yüklenmesi, travmalar ve stres, anksiyete gibi psikolojik faktörler yer almaktadır [47], [53], [54].

2.6.1.2 Lokal kas ağrısı

Lokal kas ağrısı inflamatuar olmayan ve kas kaynaklı olan bir ağrı türü olup genellikle kasların aşırı kullanımı, stres gibi tetikleyici faktörlere bağlı olarak gelişmektedir. Çiğneme

kaslarında hassasiyet, yorgunluk ve gerginlik hissi gibi belirtilerle kendini göstermektedir. Klinik semptomları ise palpasyonla uyarılan hafif rahatsızlıktan şiddetli ağrıya kadar değişken bir ağrı olarak öne çıkmaktadır. İstirahat durumunda ağrı olmaması fakat fonksiyon sırasında ağrı hissi yaygın olarak görülmektedir [44], [47].

2.6.1.3 Kas-fasya ağrısı

Kas-fasya ağrısı, çiğneme kaslarının ve bu kasları çevreleyen bağ dokusunun (fasya) etkilenmesi sonucu oluşan bir ağrıdır. Genellikle bruksizm gibi davranışlarla kasların aşırı kullanımı ve stres gibi faktörlerle tetiklenebilir. Klinik semptomlar arasında çiğneme kaslarında ve çevresindeki fasyal bölgelerde hissedilen ağrı öne çıkar. Ayrıca, kas dokusunda palpasyonla tespit edilebilen gerginlik, hassasiyet ve miyofasiyal tetik noktalar, mandibular hareketlerde fonksiyonel kısıtlanma ve çiğneme sırasında artan ağrı görülebilir [47], [54], [55].

2.6.1.4 Kas-fasya ağrısı (yansıyan)

Kas-fasya ağrısının önemli bir alt türü olan yansıyan ağrı, ağrının kaynaklandığı bölgeden farklı bir alanda algılanması ile karakterizedir. Örneğin, çiğneme kaslarındaki bir sorun, baş, boyun veya diğer komşu bölgelerde ağrıya neden olabilir. Yansıyan ağrı, klinik değerlendirmelerde dikkate alınması gereken kritik bir unsurdur, çünkü semptomların kaynağını belirlemede zorluklara neden olabilir. Bu tür ağrıların tanı ve yönetimi, multidisipliner bir yaklaşım gerektirerek hem fiziksel hem de potansiyel psikososyal etkenlerin göz önünde bulundurulmasını zorunlu kılmaktadır [43], [47], [56].

2.6.1.5 Eklem ağrısı

Eklem ağrısı (artralji), eklem yüzeyinde hissedilen rahatsızlık veya çeşitli yoğunlukta ağrı ile karakterize olup genelde sağlıklı bir eklemde gözlenmeyen bir enflamasyon veya anomali sonucu ortaya çıkmaktadır. Bu rahatsızlık durumu çiğneme, konuşma ya da ağız açma gibi fonksiyonel aktiviteler sırasında görülebilmektedir [54], [57], [58], [59].

2.6.1.6 Temporomandibular düzensizliğe bağlı baş ağrısı

Temporomandibular düzensizliğe bağlı baş ağrısı çiğneme kaslarının, eklem yapısının ve eklem çevresindeki dokuların dejenerasyonu sonucu gelişen bir ağrı türüdür. Bu tarz baş ağrıları genellikle çiğneme kaslarının aşırı kullanımı, eklem disfonksiyonu, malokluzyon,

bruksizm veya stres gibi faktörlerce tetiklenebilmektedir. Belirgin semptomları başın yan tarafında, şakaklarda veya göz çevresinde duyulan ağrı, yüz ve boyun çevresinde ağrı ve gerginlik hissiyatıdır [44], [47], [58], [60].

2.6.2 Eklem İçi Düzensizlikler

2.6.2.1 Redüksiyonslu disk deplasmanı

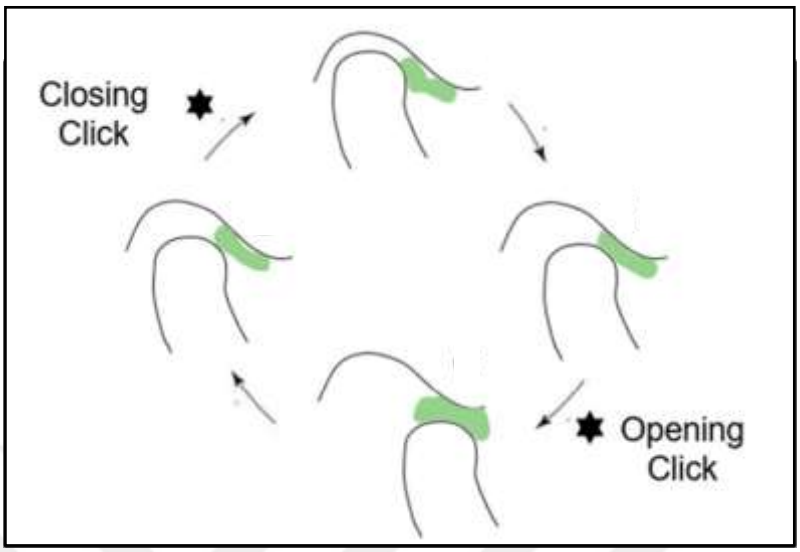
Redüksiyonlu disk deplasmanı (şekil 2.6) eklemin en sık görülen düzensizliğidir. Temporomandibular eklem düzensizliklerinin klinik teşhislerinin %41'i redüksiyonlu disk deplasmanıdır. Ayrıca, herhangi bir semptom göstermeyen bireylerin %33'ünde redüksiyonlu disk deplasmanı görülebilmektedir. Redüksiyonlu disk deplasmanına sahip hastalarda ağız kapalıyken disk kondile göre normal pozisyonundan çıkmakta ve ağız açıldığında normal kondil-disk ilişkisi görülmektedir [30], [61], [62], [63], [64], [65].

Disk herhangi bir yöne (öne, arkaya, yana, içeriye) yön değiştirebilmektedir. Posterior ve lateral yer değiştirmesi ender görülürken anterior deplasmanı en yaygın görülendir [66]. Disk kondil uyumu bozulduğunda çenede herhangi bir hareket kısıtlılığı oluşmayabilir fakat bu hareketler sağlıklı durumda olacağı kadar düzgün ve rahat da gerçekleşmeyebilir [67]. Redüksiyonlu disk deplasmanına sahip bir hastada ağız tam açıldığındaki görüntü sağlıklı bireyle aynı olmaktadır, yani bir defleksiyon (ağız açma esnasında çenenin orta hattan yana kayması ve maksimum ağız açıklığında çenenin tekrar orta hatta dönmemesi) görülmemektedir ancak ağız açılırken deviasyon (ağız açma esnasında çenenin orta hattan yana kayması ve maksimum ağız açıklığında çenenin tekrar orta hatta dönmesi) görülebilmektedir [68] [69].

Normal kondil-disk ilişkisinin bozulması, diskal kollateral ligamanların ve inferior retrodiskal laminanın uzamasıyla ilişkilidir. Bu yapılar uzadığında disk, lateral pterygoid kasın da çekme etkisiyle daha anteriorda konumlanmaktadır. Lateral pterygoid kasın bu çekme etkisi diskin posterior kısmı zamanla inceltmekte ve diski daha anteriora pozisyonlandırmaktadır [36].

Genellikle, redüksiyonlu disk deplasmanı eklem sesleriyle ilişkilendirilmektedir. Klik ve popping sesleri öncelikle redüksiyonlu disk deplasmanını akla getirmektedir. Temporomandibular düzensizlik yaşayan hastaların %26'sında klik sesi tespit edilmiştir. Bu

klik sesi sadece ağız açılırken değil, aynı zamanda ağız kapanırken (reciprocal clicking) de duyulabilmektedir [36], [70].



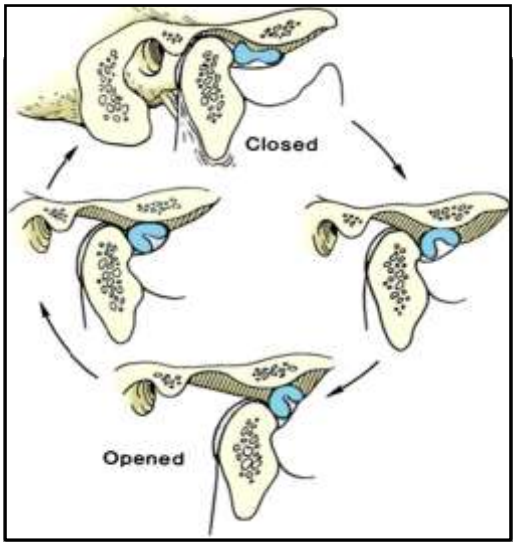
Şekil 2.6: Redüksiyonlu Disk Deplasmanı (Çene Açılıp Kapanırken Kondil-Disk İlişkisi) [70].

2.6.2.2 Redüksiyonsuz disk deplasmanı

Redüksiyonsuz disk deplasmanı (şekil 2.7) özellikle temporomandibular eklem ağrısı ve ağız açıklığında kısıtlılığa yol açabilen bir TME düzensizliğidir. Bu durum kronik veya akut olabilmektedir. Temporomandibular düzensizlikler arasında redüksiyonsuz disk deplasmanının yaygınlığı %2 ile %8 arasında olduğu tahmin edilmektedir [36], [71], [72].

Redüksiyonsuz disk deplasmanı temporomandibular düzensizlikler içinde en zorlayıcı alt gruplardan biridir. Bu durumda, kondil ile disk arasındaki normal ilişki bozulur ve disk, sürekli olarak kondilin önünde olarak konumlanır ve normal pozisyonuna geri dönmez. Redüksiyonsuz disk deplasmanının temel nedenleri genellikle makro ve mikro travmalardır. Ağrı, kısıtlı ağız açıklığı ve defleksiyon en sık görülen belirtiler arasındadır [30].

Redüksiyonsuz disk deplasmanının tedavi seçenekleri, konservatif ve cerrahi yöntemler olmak üzere iki gruba ayrılmaktadır. Konservatif yaklaşımlar arasında manipülasyon, ilaç tedavisi, alışkanlıkların değiştirilmesi, fizik tedavi ve splint tedavisi bulunurken cerrahi müdahaleler arasında artrosentez, artroskopi ve eklem cerrahisi yer almaktadır [73], [74]. Akut vakalarda, redüksiyonsuz disk deplasmanının ilk tedavisi, disk manipülasyonu ile diskin düzeltilmesi veya normal pozisyonuna getirilmesi için yapılan girişimleri içermektedir [75].



Şekil 2.7: Redüksiyonsuz Disk Deplasmanı (Çene Açılıp Kapanırken Disk-Kondil İlişkisi).

2.6.3 Dejeneratif Eklem Hastalığı

Dejeneratif eklem hastalığı, temporomandibular eklemde görülen ve genellikle osteoartrit olarak adlandırılan durumu kapsamaktadır. Osteoartrit, eklem kıkırdağının aşınması, eklem yüzeylerinin bozulması ve eklemde yapısal değişikliklerin meydana gelmesi ile karakterizedir. Semptomlar arasında eklem bölgesinde sürekli veya fonksiyonla artan ağrı, ağız açma ve kapama sırasında hareket kısıtlılığı, klik veya krepitasyon sesleri ve yüz bölgesinde hassasiyet bulunmakdaır. Osteoartritin etiyolojisi çok faktörlü olup, yaşlanmaya bağlı kıkırdak aşınması, aşırı kullanım sonucu ekleme binen yük, çene bölgesine alınan travmalar ve genetik veya metabolik faktörlerin birleşimi ile ilişkilendirilmektedir [20], [47], [76].

2.6.4 Subluksasyon

Subluksasyon, kondilin artiküler eminensin önüne doğru aşırı ve kontrolsüz şekilde hareket etmesiyle, eklem sınırlarının geçici olarak aşılması sonucu kilitlenme hissiyle ortaya çıkan ve genellikle bağ dokusundaki gevşeklik ya da travmalar nedeniyle gelişen bir TME bozukluğudur [77].

2.7 TEMPOROMANDİBULAR EKLEM DÜZENSİZLİKLERİN TANISI

Temporomandibular düzensizlikleri değerlendiren veya tanı koymaya çalışan klinisyen veya araştırmacı, temporomandibular düzensizliklerin farklı yapılarda ve karmaşık bir hastalık grubu olduğunu aklında bulundurmalıdır. Bu rahatsızlıklar için farklı tanı koyma yöntemleri bulunmakla birlikte, TMD/TK'nin doğruluğu yapılan araştırmalarca kanıtlanmıştır [30], [37], [78].

Temporomandibular düzensizliğin tanısını koymada en kritik unsur, detaylı bir klinik muayene ve kapsamlı bir anamnezdir. Muayene yapan doktor, her bir hastada ağrının ve semptomların yerini, ağrının şiddetini ve varsa diğer semptomları belirlemelidir. Hasta ile semptomları konuşurken, doktorun ağrının sürekli, aralıklı, keskin olup olmadığını, uyuşukluk-karıncalanma gibi durumları anlaması gerekmektedir. Bununla birlikte, doktorun semptomların ne zaman ve nasıl başladığını, hastanın herhangi bir travma geçmişi olup olmadığını, daha önce aynı rahatsızlığı yaşayıp yaşamadığını öğrenmesi gerekmektedir. Temporomandibular düzensizliklerin tanısında doğru anamnez almak önemli bir adımdır. Bu noktada, doktorun hastanın hikayesine özen göstermesi ve dikkatli dinleyip doğru soruları sorması gerekmektedir [79].

Klinik muayenede ise hekim hastanın postürünü, yüz ve boyun kaslarının belirginliğini, çene boyutunu ve şeklini, çene simetrisini, dişleri, dişetlerini ve ağız içi yapıları incelemeli, anormallikleri gözlemlemelidir. Orofasiyal yapıların normal ve patolojik görünümünü tanımak için deneyim gerekmektedir. Klinik muayenede ise aktif hareket açıklığı testi (AROM), pasif yardımcı hareket testi, manuel kas testi, palpasyon testi gibi testler yapılmaktadır [79].

2.8 YAPAY ZEKA

Yapay zeka, insanlara çeşitli yollarla yardımcı olabilecek şekilde zekanın ve düşüncenin makinelere aktarılmasına odaklanan bir bilgi işlem alanıdır. İlk kez 1956'da John McCarthy tarafından ortaya atılan yapay zeka terimi, mühendislik, matematik, fizik, teknoloji gibi birçok alanda kullanılmıştır ve kullanım alanı zaman geçtikçe artmaktadır. Yapay zeka fikri, makinelerin zeka kazanabileceğini öngörür yani makinelerin kendi kendilerine öğrenebileceğini, yeni durumlara adapte olabileceğini ve hataları düzeltebileceğini savunur [80].

2.8.1 Yapay Zeka Tarihçesi

İngiliz matematikçi Alan Turing, İkinci Dünya Savaşı'nda "The Bombe" adını verdiği bir kod çözme cihazı geliştirmiştir. Bu cihaz, ilk işleyen elektro-mekanik bilgisayar olarak bilinmektedir. Bombe'un etkileyici performansı, önceden en iyi matematikçilerin bile imkansız gördüğü görevleri başarması, Turing'in bu makinelerin zeka potansiyeli hakkında düşünmesine neden olmuştur. 1950'de Turing, "Bilgi İşlem Makineleri ve Zeka" adlı bir makale yayınlamıştır. Bu makalede, akıllı makinelerin nasıl yaratılabileceği ve özellikle bu makinelerin zekasını test etmenin yollarını açıklamıştır [81]. Bu Turing Testi günümüzde hala yapay bir sistemin zekasını değerlendirmek için bir kriter olarak kabul edilmektedir. Eğer bir insan, başka bir insanla ve bir makineyle etkileşime geçerken makine ile insan arasındaki farkı anlayamıyorsa, o zaman makine "zeki" olarak kabul edilmektedir.

"Yapay zeka" terimi ise tüm bu olaylardan altı yıl sonra 1956'da Marvin Misnky ve John McCarthy tarafından Dartmouth College'da yaklaşık sekiz hafta süren Dartmouth Yaz Araştırma Projesi'nde (DSRPAI) kullanılmıştır. DSRPAI'den yıllar sonra, yapay zeka alanında önemli gelişmeler yaşandı. 1964-1966 yılları arasında Massachusetts Institute of Technology (MIT)'de Joseph Weizenbaum tarafından ELIZA programı geliştirilmiştir. Bu program, insanlarla simüle edilmiş konuşmalar yapabilen bir doğal dil işleme aracıdır ve Turing Testi'ni geçmeye çalışan ilk programlardan biridir. Bu ilham verici gelişmeler, AI araştırmalarına ciddi miktarda fon sağlamış ve bu da birçok yeni projenin ortaya çıkmasına sebep olmuştur. 1970'te Marvin Minsky, Life Magazine'e yaptığı bir röportajda, ortalama bir insanın genel zeka düzeyine sahip bir makinenin üç ila sekiz yıl içinde geliştirilebileceğini söylemiştir. Fakat bu yıllardan sonra yapay zekada beklenen atılım görülmemiş, 1973 yılında Amerika Birleşik Devletleri Meclisi'nde yapay zeka araştırmaları için harcanan yüksek bütçelere eleştiriler başlamıştır. Ardından İngiliz Hükümeti de yapay zeka araştırmalarına destek vermeyi kesmiş ve yapay zeka araştırmaları için sadece belirli üniversitelere (Edinburgh, Essex, Sussex) yetki vermiştir. Bu durum geçtiğimiz on yıla kadar yapay zekanın beklenilen kadar gelişmesini engellemiştir [82].

2.8.2 Yapay Sinir Ağları

Yapay sinir ağları, biyolojik sinir ağlarından ilham alan büyük ölçekli paralel hesaplama sistemleridir. Bu yapay sinirler, doğal sinir hücrelerinin işlevlerini taklit ederek oluşturulmuş

bir hesaplama modelidir. Yapay sinir ağları, bilgi işleme amacıyla kullanılırlar ve genellikle desen tanıma, tahminleme ve veri sıkıştırma gibi alanlarda tercih edilmektedirler [83].

Her bir yapay sinir, çok sayıda basit işlemciden ve birbirleriyle bağlantılı olan çok sayıda bağlantıdan oluşur ve insan beyninin işlevlerini taklit eder. Yapay sinir ağlarının gücü, birçok basit işlemcinin uygun bir şekilde bir araya gelmesinden gelmektedir. Her işlemci tek başına yeterince güçlü değildir; sadece girdilerinin basit bir doğrusal olmayan fonksiyonunu işlemektedir. Ancak bir araya geldiklerinde, belirli bir problem için bir çözüm sağlayabilmektedirler.

2.9 BÜYÜK DİL MODELLERİ (BDM)

Büyük dil modelleri kendi kendini denetleyen veya yarı denetimli öğrenme tekniklerini kullanarak büyük miktarda kategorilendirilmemiş metin üzerinde eğitilen sinir ağı tabanlı dil modelleridir. Bu modeller, insan benzeri metinleri işlemek, anlamak ve üretmek için tasarlanmış gelişmiş yapay zeka sistemleridir. Derin öğrenme teknikleriyle oluşturulan bu modeller genellikle web siteleri, kitaplar, makaleler gibi zengin veri kaynaklarından milyarlarca kelime içeren büyük veri setleriyle eğitilmektedirler [84].

Büyük dil modelleri; çeviri ve sohbet botları yapay zeka asistanları gibi, yapay zeka uygulamalarıyla farklı görevleri tamamlamak üzere eğitilmektedirler. Eğitilmiş modeller metin sınıflandırma, soru yanıtlama, belge özetleme gibi problemleri çözebilmektedir. Bu yapay zeka uygulamaları sağlık, ekonomi, eğitim gibi alanlarda kullanılmaktadır ve birçok firmanın kendi geliştirdiği büyük dil modeli bulunmaktadır [85].

2.9.1 ChatGPT

En yaygın kullanılan büyük dil modellerinden biri olan ChatGPT, OpenAI tarafından 2022 senesinde geliştirilmiştir. Derin öğrenme algoritmalarından faydalanan ChatGPT insan benzeri yanıtlarıyla birçok farklı amaç için kullanılmaktadır. ChatGPT özgün içerikler üretebilmektedir ayrıca sorulara mantıklı, bağlantılı, konu ile alakalı ve akıcı yanıtlar üretebilmektedir. ChatGPT, kendisine tanıtılan kaynaklardan veri toplamaktadır ardından bu veriler arasındaki ilişkileri belirlemekte ve hangi durumlarda hangi metnin kullanılması gerektiğine karar vermektedir ayrıca insanla olan konuşmalarından bilgi edinmekte ve bu bilgileri daha özelleştirilmiş yanıtlar üretmek için kullanmaktadır. Open AI, ChatGPT'yi

yıllar içerisinde güncelleyerek yeni sürümlerini tanıtmaktadır. En son çıkan sürümler mart 2023'te tanıtılan GPT-4 tabanlı versiyon ve mayıs 2024'te tanıtılan Chat GPT4o versiyonudur [86], [87], [88], [89].

2.9.2 Microsoft Copilot

Microsoft Copilot 2023 yılında tanıtılan, yazılım geliştirme süreçlerini daha verimli hale getirmek için tasarlanmış, yapay zeka tabanlı bir yardımcıdır. Bu teknoloji, yazılım geliştirme, satış, hizmet deneyimi, veri analitiği ve sohbet botu gibi çeşitli alanlarda kullanılabilmektedir. Microsoft Copilot, kod yazma, belge oluşturma, hata ayıklama ve içerik üretme gibi birçok işlevi yerine getirebilmektedir. Bu tür araçlar, kullanıcıların taleplerini anlamak ve uygun yanıtlar üretmek için genellikle doğal dil işleme (NLP) ve makine öğrenimi tekniklerini kullanmaktadırlar [90], [91], [92], [93].

2.9.3 Google Gemini

Google Gemini, Google'ın yapay zeka ve dil işleme alanındaki en yeni projelerinden biridir. 2023 yılında tanıtılan bu model, Google'ın yapay zeka stratejisinin bir parçası olarak dil işleme ve yapay zeka uygulamaları için tasarlanmıştır. Google Gemini, bilgileri analiz edebilen, insan dillerini özel bir şekilde anlayabilen, metin oluşturabilen, çeşitli yaratıcı içerikler üretebilen ve soruları bilgilendirici bir şekilde yanıtlayabilen bir makine öğrenimi modelidir. Ayrıca, Gemini, verilen talimatları takip ederek ve açık uçlu soruları dahi detaylı ve açıklayıcı bir şekilde cevaplayarak sonuçlara ulaşma yeteneğine sahiptir [94].

2.9.4 Docus

Merkezi Amerika Birleşik Devletleri'nde olan Docus firması üç girişimci tarafından 2021 yılında kurulmuştur. Bu firma Docus adını verdikleri, sağlık sektöründe yapay zeka destekli teknolojileri kullanarak hastalıkları teşhis etme ve tedavi planları oluşturma amacı güden bir platform oluşturmuştur. Bu platform, tıbbi görüntüleme yazılımları ve makine öğrenimi algoritmalarını entegre ederek teşhis sürecini hem daha hızlı hem de daha hassas hale getirmektedir. Sağlık profesyonelleri, Docus'u kullanarak hastaların sağlık verilerini ayrıntılı bir şekilde analiz edebilmekte, hastalıkları tanımlayabilmekte ve bireyselleştirilmiş tedavi planları geliştirebilmektedir. Böylece, hastaların sağlık hizmetlerine erişimi kolaylaşmakta ve tedavi süreçlerinin etkinliği artırılmaktadır [95], [96].

2.9.5 Claude 2

Claude, Anthropic firması tarafından haziran 2024'te geliştirilmiş bir yapay zeka dil modelidir. Claude, geniş bir bilgi haznesine sahip olup, çeşitli konularda karşılıklı interaktif iletişime geçebilmekte, soruları yanıtlayabilmekte, analizler yapabilmekte, metinler oluşturabilmekte, kod yazabilmekte ve çeşitli görevleri yerine getirebilmektedir [95], [97].

2.9.6 Perplexity

Perplexity, Perplexity AI,Inc. tarafından 2022 yılında San Francisco, Kaliforniya, Amerika Birleşik Devletleri'nde geliştirilmiştir. Kullanıcıların sorularına hızlı ve doğru yanıtlar sunan bir yapay zeka destekli bilgi arama platformudur. Bu sistem, gelişmiş bir arama motoru ile bir AI sohbet asistanının özelliklerini bir araya getirerek bilgi keşfetmeyi ve içerik üretimini daha erişilebilir hale getirmeyi hedeflemektedir. Hem bireyler hem de profesyoneller için bilgi edinme ve içerik oluşturma süreçlerini basitleştiren yenilikçi bir çözüm alternatifidir [98].

2.10 DİŞ HEKİMLİĞİNDE YAPAY ZEKA UYGULAMALARI

Yapay zeka geçmiş verilerden bilgi çıkarma konusunda hekime ve sağlık personellerine yardım etmekte ve hekim adına zaman alıcı görevleri kolaylaştırmaktadır. Diş hekimliğindeki yapay zeka uygulamaları genellikle lezyonları normal yapılardan ayırt etme, tanı koyma, risk faktörlerini değerlendirme, olası sonuçları simüle etme, değerlendirme ve tedavi planı oluşturma konularında kullanılmaktadır. Bu teknoloji, hekimlerin iş yükünü hafifletirken, daha hızlı ve hassas kararlar almalarına olanak tanımaktadır [99].

Diş hekimleri, en etkili tedavi seçeneğini belirlemek için ellerindeki tüm bilgiyi kullanmak zorundadır. Ayrıca, doğru klinik karar verme yeteneklerine sahip olmalarının yanı sıra gelecekteki durumu tahmin etmeleri de beklenmektedir. Fakat bazı durumlarda, kısıtlı bir zaman diliminde doğru klinik kararı vermek için yeterli bilgiye sahip olmayabilirler. Yapay zeka uygulamaları, hekimlere rehberlik ederek daha iyi kararları hızlıca almalarını sağlayabilir ve performanslarını artırabilir [100].

Yapay zekanın diş hekimliğinde kullanım alanları;

Tanı koymak: Hastalıkların teşhisi, hekimin semptomları analizine, test sonuçlarına ve diğer faktörlere dayanır. Bu faktörler klinikte hekimin bilgi birikimine bağımlıdır ancak yüz binlerce vaka ile eğitilmiş yapay zeka en yetkin uzmanın bile klinik deneyimini aşabilmektedir [101].

a. Kariyoloji ve endodonti: Otomatik lezyon segmentasyonu, yapay zekanın kariyoloji ve endodonti teşhislerinde temel bir unsur haline gelmiştir. Yapay zeka, konik ışınlı bilgisayarlı tomografi ile periapikal lezyonları teşhis etmekte, görüntüyü "lezyon", "diş yapısı", "kemik", "restoratif materyal" gibi kısımlara ayırmada klinik uzmanlarla benzer sonuçlar elde etmektedir [102], [103].

b. Temporomandibular düzensizlikler: Teşhisi için yapay zeka uygulamalarından faydalanılabilmektedir [104].

c. Ortodonti: Yapay zeka uygulamaları sefalometrik analiz konusunda oldukça başarılıdır. Yapılan bir çalışmaya göre yapay zeka algoritmaları 20 sefalometrik noktayı yalnızca 2.01 mm'lik bir ortalama hatayla belirlemiştir [83].

d. Periodontoloji: Yapılan bir çalışmada yapay zeka periodontal olarak sağlıklı olmayan premolar ve molarları teşhis etmede %81 ve %76.7 oranında başarılı olmuştur [105].

Tedavi: Yapay zekanın anatomik yapıları ayırt etme ve muhtemel sonuçları simüle etme becerisi, diş tedavisinde, anatomik rehberlik, tedavi planlama ve sonuçların değerlendirilmesinde faydalı bir ek yardımcı haline gelmesini sağlamaktadır [99].

a. Oral ve maksillofasiyal cerrahi: Cerrahi operasyonlarda önemli anatomik noktaların belirlenmesinde ve anatomik noktaların takip edilmesinde yapay zeka uygulamarı hekimlere kolaylık sağlamaktadır [99].

b. Protetik Uygulamalar: CAD/CAM uygulamaları yapay zekaya entegre edilerek hekimlere uygulama kolaylığı sunmaktadır [106].

3. GEREÇ VE YÖNTEM

3.1 SORU ŞABLONLARININ HAZIRLANMASI VE VERİ KÜMESİNİN OLUŞTURULMASI

Bu tez çalışmasında eklem düzensizliklerini teşhis etmek için kullanılan soru şablonları TMD/TK'nın tanı karar ağacı tablosuna göre hazırlanmıştır. TMD/TK tanı karar ağacında belirtilen temporomandibular düzensizlikler ayrıntılı bir şekilde listelenmiş ve her bir düzensizlik için tanı kriterlerine uygun olarak soru şablonu hazırlanmıştır (n=13) [107].

Büyük dil modellerinin performanslarınının dile bağlı olarak değişip değişmediğini belirlemek için bu sorular hem Türkçe hem de İngilizce olarak düzenlenmiştir. Tanı karar ağacı şekil 2.2-2.5'de verilmiştir. Tanı karar ağacına göre düzenlenmiş olan Türkçe soru şablonları şekil 3.1- 3.13 arasında verilmiştir, İngilizce soru şablonları ise şekil 3.14-3.26 arasında verilmiştir.

3.1.1 Tanı Karar Ağacına Göre Düzenlenmiş Türkçe Soru Şablonları

Miyalji (Kas Ağrısı)
Son 30 gün içinde hastanın çenesinde, şakağında, kulağının içinde ya da önünde herhangi bir tarafta ağrı olduysa ve çene hareketleri, fonksiyonel hareketler, parafonksiyonel hareketler son 30 günde meydana gelen ağrıda iyi ya da kötü bir değişiklik oluşturduysa;
Oluşan ağrı çiğneme kaslarındaysa;
Aynı veya benzer bir ağrı çene açılırken ve çiğneme kaslarına yapılan 2 saniyelik palpasyonda görülüyorsa olası temporomandibular düzensizlik nedir?

Şekil 3.1: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

Temporomandibular Düzensizliğe Bağlı Baş Ağrısı
Hastada miyalji veya artralji tanısı varsa;
Hasta son 30 gün içerisinde şakak bölgesini de içine alacak şekilde temporal bölgede herhangi bir baş ağrısı hissettiyse ve çene hareketleri, fonksiyonel hereketler, parafonksiyonel hareketler son 30 günde meydana gelen baş ağrısında iyi ya da kötü bir değişiklik oluşturduysa;
Aynı veya benzer bir ağrı çene açılırken veya lateral ve protrüziv hareketlerde veya temporal kas palpasyonunda hissediliyorsa;
Hastada diğer baş ağrısı teşhisleriyle daha iyi açıklanamayan baş ağrısı varsa olası temporomandibular düzensizlik nedir?

Şekil 3.2: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

Eklem Ağrısı (Artralji)
Son 30 gün içinde hastanın çenesinde, şakağında, kulağının içinde ya da önünde herhangi bir tarafta ağrı olduysa ve çene hareketleri, fonksiyonel hareketler, parafonksiyonel hareketler son 30 günde meydana gelen ağrıda iyi ya da kötü bir değişiklik oluşturduysa;
Oluşan ağrı temporomandibular eklemdeyse;
Aynı ağrıya benzer ağrı çene açılırken veya lateral ve protrüziv hareketlerde veya TME palpasyonunda hissediliyorsa olası temporomandibular düzensizlik nedir?

Şekil 3.3: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

Kas – fasya ağrısı (yansıyan) (Myofascial Pain with Referral)
Son 30 gün içinde hastanın çenesinde, şakağında, kulağının içinde ya da önünde herhangi bir tarafta ağrı olduysa ve çene hareketleri, fonksiyonel hareketler, parafonksiyonel hareketler son 30 günde meydana gelen ağrıda iyi ya da kötü bir değişiklik oluşturduysa;
Oluşan ağrı çiğneme kaslarındaysa;
Aynı veya benzer bir ağrı çene açılırken veya çiğneme kaslarına yapılan 2 saniyelik palpasyonda görülüyorsa
Palpasyonda oluşan ağrı kas sınırlarının ötesine uzanıyorsa olası temporomandibular düzensizlik nedir?

Şekil 3.4: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

Kas Fasya Ağrısı (Myofascial Pain with Spreading)
Son 30 gün içinde hastanın çenesinde, şakağında, kulağının içinde ya da önünde herhangi bir tarafta ağrı olduysa ve çene hareketleri, fonksiyonel hareketler, parafonksiyonel hareketler son 30 günde meydana gelen ağrıda iyi ya da kötü bir değişiklik oluşturduysa;
Oluşan ağrı çiğneme kaslarındaysa;
Aynı veya benzer bir ağrı çene açılırken veya çiğneme kaslarına yapılan 2 saniyelik palpasyonda görülüyorsa;
Palpasyonda oluşan ağrı kas sınırlarının ötesine uzanmıyorsa;
Palpasyonda oluşan ağrı uyarı alanının ötesine yayılıyorsa olası temporomandibular düzensizlik nedir?

Şekil 3.5: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

Lokal Kas Ağrısı (Local Myalgia)
Son 30 gün içinde hastanın çenesinde, şakağında, kulağının içinde ya da önünde herhangi bir tarafta ağrı olduysa ve çene hareketleri, fonksiyonel hereketler, parafonksiyonel hareketler son 30 günde meydana gelen ağrıda iyi ya da kötü bir değişiklik oluşturduysa;
Oluşan ağrı çiğneme kaslarındaysa;
Aynı veya benzer bir ağrı çene açılırken veya çiğneme kaslarına yapılan 2 saniyelik palpasyonda görülüyorsa;
Palpasyonda oluşan ağrı kas sınırlarının ötesine uzanmıyorsa;
Palpasyonda oluşan ağrı uyarı alanının ötesine yayılmıyorsa olası temporomandibular düzensizlik nedir?

Şekil 3.6: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

Redüksiyonlu Disk Deplasmanı
Hastanın çenesi daha önce kapalı pozisyonda kilitlenmediyse
Hastada son 30 gün içinde hastanın çene hareketleri sırasında ya da çenesini kullanırken herhangi bir eklem sesi oluştuysa veya muayene sırasında temporomandibular eklemden (TME) çene açma-kapama, lateral ya da protrüziv hareketler sırasında ses geliyorsa;
Muayenede klik sesi çene hem açılırken hem de kapanırken geliyorsa ya da sadece açılırken veya kapanırken ve lateral, protrüziv hareketler sırasında geliyorsa;
Hastada aralıklı kilitlenme ve kısıtlı ağız açıklığı yoksa olası temporomandibular düzensizlik nedir?

Şekil 3.7: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

Redüksiyonlu Disk Deplasmanı Aralıklı Kilitlenme Olan

Hastanın çenesi daha önce kapalı pozisyonda kilitlenmediyse
Hastada son 30 gün içinde hastanın çene hareketleri sırasında ya da çenesini kullanırken herhangi bir eklem sesi oluştuysa veya muayene sırasında temporomandibular eklemden (TME) çene açma-kapama, lateral ya da protrüziv hareketler sırasında ses geliyorsa;
Muayenede klik sesi çene hem açılırken hem de kapanırken geliyorsa veya sadece açılırken ya da kapanırken ve lateral, protrüziv hareketler sırasında geliyorsa;
Hastada halen var olan kısıtlı ağız açıklığına neden olan aralıklı kilitlenme varsa;
Eklem kilitlenmesi gerçekleşirse ağzı açmak için manevra gerekli değilse olası temporomandibular düzensizlik nedir?

Şekil 3.8: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

Redüksiyonsuz Disk Deplasmanı Kısıtlı Ağız Açıklığı Olan

Hastanın çenesi daha önce kapalı pozisyonda kilitlenmediyse
Hastada son 30 gün içinde hastanın çene hareketleri sırasında ya da çenesini kullanırken herhangi bir eklem sesi oluştuysa veya muayene sırasında temporomandibular eklemden (TME) çene açma-kapama, lateral ya da protrüziv hareketler sırasında ses geliyorsa;
Muayenede klik sesi çene hem açılırken hem de kapanırken geliyorsa veya sadece açılırken ya da kapanırken ve lateral, protrüziv hareketler sırasında geliyorsa;
Hastada halen var olan kısıtlı ağız açıklığına neden olan aralıklı kilitlenme varsa;
Eklem kilitlenmesi gerçekleşirse ağzı açmak için manevra gerekliyse olası temporomandibular düzensizlik nedir?

Şekil 3.9: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

Redüksiyonsuz Disk Deplasmanı Kısıtlı Ağız Açıklığı Olmayan

Hasta bir an bile olsa çenesini tamamen açamadığı kilitlenme veya takılma yaşadıysa ve bu kilitlenme-takılma çeneyi açmayı ve yemek yemeyi kısıtladıysa
Overbite dahil maksimum yardımlı ağız açıklığı 40mm'den fazlaysa olası temporomandibular düzensizlik nedir?

Şekil 3.10: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

Redüksiyonsuz Disk Deplasmanı Kısıtlı Ağız Açıklığı Olan

Hasta bir an bile olsa çenesini tamamen açamadığı kilitlenme veya takılma yaşadıysa ve bu kilitlenme-takılma çeneyi açmayı ve yemek yemeyi kısıtladıysa Overbite dahil maksimum yardımlı ağız açıklığı 40mm'den azsa olası temporomandibular düzensizlik nedir?

Şekil 3.11: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

Dejeneratif Eklem Hastalığı

Hastanın son 30 gün içinde çenesini hareket ettiğinde veya kullandığında bir eklem sesi hissi varsa veya hasta muayene sırasında açma, kapama, lateral veya protrüziv hareketlerde bir ses hissettiyse;
Hekim tarafından muayene sırasında açma-kapama veya lateral, protrüziv hareketlerde krepitus saptandıysa olası temporamandibular rahatsızlık nedir?

Şekil 3.12: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

Sublüksasyon

Hasta son 30 gün içinde geniş bir şekilde ağzını açtığında, bu geniş açma pozisyonunda bir an bile olsa kapatamadığı bir şekilde kilitlenme-takılma olduysa ve geniş açılma pozisyonunda kilitlenme-takılma olduğunda çenesini kapatmak için dinlendirme, hareket ettirme, bastırma veya manevra yapmak zorunda kaldıysa;
Kilitlenme-takılma olayı klinik olarak mevcutsa olası temporomandibular düzensizlik nedir?

Şekil 3.13: Tanı Karar Ağacına Göre Düzenlenmiş Olan Türkçe Soru Şablonu Örneği.

3.1.2 Tanı Karar Ağacına Göre Düzenlenmiş İngilizce Soru Şablonları

Myalgia
If the patient has experienced pain in the jaw, temple, inside or in front of the ear on either side in the last 30 days, and if jaw movements, functional movements, or parafunctional movements have caused an improvement or worsening in the pain over the last 30 days;
If the pain is located in the masticatory muscles;
If the familiar pain is felt when opening the jaw or during 2 second palpation of the masticatory muscles, what is the possible temporomandibular disorder?

Şekil 3.14: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

Headache Attributed to TMD
If the patient has been diagnosed with myalgia or arthralgia;
In the last 30 days if the patient has experienced a headache in the temporal region, including the temple area, and if jaw movements, functional movements, or parafunctional movements have caused an improvement or worsening in the headache during the last 30 days;
If the familiar headache is felt when opening the jaw, during lateral or protrusive movements, or upon palpation of the temporal muscle;
If the patient has a headache that cannot be better explained by other headache diagnoses, what is the possible temporomandibular disorder?

Şekil 3.15: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

Arthralgia
If the patient has experienced pain in the jaw, temple, inside or in front of the ear on either side in the last 30 days, and if jaw movements, functional movements, or parafunctional movements have caused an improvement or worsening in the pain over the last 30 days;
If the pain is located in the temporomandibular joint;
If a familiar pain is felt when opening the jaw, during lateral and protrusive movements, or upon palpation of the TMJ, what is the possible temporomandibular disorder?

Şekil 3.16: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

Myofascial Pain with Referral
If the patient has experienced pain in the jaw, temple, inside or in front of the ear on either side in the last 30 days, and if jaw movements, functional movements, or parafunctional movements have caused an improvement or worsening in the pain over the last 30 days;
If the pain is located in the masticatory muscles;
If the familiar pain is felt when opening the jaw or during 2 second palpation of the masticatory muscles
If the pain during palpation extends beyond the muscle boundaries, what is the possible temporomandibular disorder?

Şekil 3.17: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

Myofascial Pain with Spreading
If the patient has experienced pain in the jaw, temple, inside or in front of the ear on either side in the last 30 days, and if jaw movements, functional movements, or parafunctional movements have caused an improvement or worsening in the pain over the last 30 days;
If the pain is located in the masticatory muscles;
If the familiar pain is felt when opening the jaw or during 2 second palpation of the masticatory muscles
If the pain during palpation does not extends beyond the muscle boundaries
If the pain during palpation extends beyond the area of stimulation, what is the possible temporomandibular disorder?

Şekil 3.18: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

Local Myalgia
If the patient has experienced pain in the jaw, temple, inside or in front of the ear on either side in the last 30 days, and if jaw movements, functional movements, or parafunctional movements have caused an improvement or worsening in the pain over the last 30 days;
If the pain is located in the masticatory muscles;
If the familiar pain is felt when opening the jaw or during 2 second palpation of the masticatory muscles
If the pain during palpation does not extends beyond the muscle boundaries
If the pain during palpation does not extends beyond the area of stimulation, what is the possible temporomandibular disorder?

Şekil 3.19: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

Disc Displacement with Reduction

If the patient has never experienced jaw locking in closed position before;
If there has been any joint sound during jaw movements or while using the jaw within the last 30 days, or if during the examination, a sound is heard from the temporomandibular joint (TMJ) during jaw opening-closing, lateral, or protrusive movements;
If a clicking sound is heard during both jaw opening and closing, or only during opening or closing, and during lateral or protrusive movements;
If the patient does not have intermittent locking or limited mouth opening, what is the possible temporomandibular disorder?

Şekil 3.20: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

Disc Displacement with Reduction, with Intermittent Locking
If the patient has never experienced jaw locking in closed position before;
If there has been any joint sound during jaw movements or while using the jaw within the last 30 days, or if during the examination, a sound is heard from the temporomandibular joint (TMJ) during jaw opening-closing, lateral, or protrusive movements;
If a clicking sound is heard during both jaw opening and closing, or only during opening or closing, and during lateral or protrusive movements;
If the patient has current intermittent locking with limited opening;
If joint locking occurs but no maneuver is required to open the mouth, what is the possible temporomandibular disorder?

Şekil 3.21: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

Disc Displacement without Reduction with Limited Opening
If the patient has never experienced jaw locking in closed position before;
If there has been any joint sound during jaw movements or while using the jaw within the last 30 days, or if during the examination, a sound is heard from the temporomandibular joint (TMJ) during jaw opening-closing, lateral, or protrusive movements;
If a clicking sound is heard during both jaw opening and closing, or only during opening or closing and during lateral or protrusive movements;
If the patient has current intermittent locking with limited opening;
If joint locking occurs and maneuver is required to open the mouth, what is the possible temporomandibular disorder?

Şekil 3.22: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

Disc Displacement without Reduction without Limited Opening
If the patient ever had jaw lock or catch, even for a moment, so that it would not open all the way and that lock or catch severe enough to limit patient's jaw opening and interference in mastification;
If the maximum assisted mouth opening, including overbite, is greater than 40mm, what is the possible temporomandibular disorder?

Şekil 3.23: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

Disc Displacement without Reduction with Limited Opening
If the patient ever had jaw lock or catch, even for a moment, so that it would not open all the way and that lock or catch severe enough to limit patient's jaw opening and interference in mastification;
If the maximum assisted mouth opening, including overbite, is less than 40mm, what is the possible temporomandibular disorder?

Şekil 3.24: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

Degeneretive Joint Disease
If the patient has felt a joint sound when moving or using their jaw within the last 30 days, or if the patient has detected a sound during opening, closing, lateral, or protrusive movements during the examination;
If crepitus is detected by the physician during opening-closing or lateral and protrusive movements during the examination, what is the possible temporomandibular disorder?

Şekil 3.25: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

Subluxation
If the patient experienced locking or catching in the wide open position, even for a moment, within the last 30 days, where they were unable to close their mouth, and when this locking or catching occurred in the wide open position, the patient had to rest, move, press, or perform a maneuver to close their jaw;
If the lock in open position clinically present, what is the possible temporomandibular disorder?

Şekil 3.26: Tanı Karar Ağacına Göre Düzenlenmiş Olan İngilizce Soru Şablonu Örneği.

3.2 BÜYÜK DİL MODELLERİNE SORULARIN SORULMASI

Bu tez çalışmasında sık kullanılmakta olan 8 adet büyük dil modeline (Microsoft Copilot, Open AI Chat GPT-3.5, Open AI Chat GPT-4o, Open AI Chat GPT 4, Google Gemini, Anthropic Claude 2, Docus AI, Perplexity AI) TMD/TK tanı karar ağacına göre hazırlanan soru şablonları ilk olarak 8-10 Eylül 2024 tarihleri arasında sorulmuş ve her modelin verdiği yanıt titizlikle değerlendirilmiştir.

Modellerin teşhis performansının ve dile bağımlılığının tespiti için büyük dil modellerinin Türkçe ve İngilizce sorulara verdikleri yanıtlar bu incelemeler sonucunda TMD/TK tanı karar ağacı ile kıyaslanıp başarılı ve başarısız teşhisler belirlenmiş ve bu sonuçlar tablo haline getirilmiştir.

4 ay sonra aynı sorular tekrar büyük dil modellerine yöneltilerek modellerin zaman içerisindeki öğrenme performansları ve tutarlılıkları değerlendirilmiştir. Bu sayede dil modellerinin sadece teşhis başarısını ölçmek dışında semptomlar arasında ilişki kurabilme, bunları anlamlandırma, analiz yetenekleri sınanmıştır.

3.3 İSTATİSTİKSEL ANALİZ

Çalışmada elde edilen bulgular değerlendirilirken, istatistiksel analizler için IBM SPSS Statistics 22 programı kullanılmıştır. Verilerin karşılaştırılmasında Cochran’s Q test ve Mc Nemar testleri kullanılmıştır. Anlamlılık $p<0,05$ düzeyinde değerlendirilmiştir.

4. BULGULAR

Çalışmada yer alan tüm büyük dil modellerinin 2024 Eylül ayına ait doğru/yanlış yanıt verileri Tablo 4.1'de verilmiştir. 2025 Ocak ayına ait doğru/yanlış yanıt verileri ise Tablo 4.2'de verilmiştir.

Tablo 4.1: Çalışmada Kullanılan Büyük Dil Modellerinin 2024 Eylül Ayında Yapılan Değerlendirmedeki Doğru/Yanlış Verileri.

	Microsoft CopRot		Open AI Chat GPT-3.5		Open AI Chat GPT-4o		Open AI Chat GPT 4		Google Gemini		Docus AI		Anthropic Claude 2		Perplexity AI		
	T	İ	T	İ	T	İ	T	İ	T	İ	T	İ	T	İ	T	İ	
S1,Q1	0	0	0	0	0	1	0	1	0	0	0	0	1	1	0	0	
S2Q2	0	1	1	0	1	1	1	1	0	0	0	0	1	1	0	0	
S3,Q3	0	1	1	1	1	1	1	1	0	0	0	0	1	1	0	0	
S4,Q4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Ağrı ile İlişkili TMD ve Baş Ağrısı
S5,Q5	1	1	0	0	0	0	0	0	1	1	1	1	0	1	1	1	
S6,Q6	0	0	1	1	1	1	1	1	0	0	0	1	1	1	0	0	
S7,Q7	0	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0	
S8,Q8	1	0	0	1	0	1	0	1	0	1	0	0	1	1	0	1	
S9,Q9	1	0	1	1	1	1	1	0	0	0	0	1	0	0	0	0	Eklem İçi Düzensizlikler
S10,Q10	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	
S11,Q11	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	
S12,Q12	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	Dejeneratif Eklem Hastalığı
S13,Q13	0	0	0	0	0	1	0	1	0	0	0	0	1	1	0	1	Subluksasyon
TOPLAM	5	8	8	8	9	11	8	10	3	4	6	7	10	11	2	5	
GENEL TOPLAM	13		16		20		18		7		13		21		7		

Tablo 4.2: Çalışmada Kullanılan Büyük Dil Modellerinin 2025 Ocak Ayında Yapılan Değerlendirmedeki Doğru/Yanlış Verileri.

	Microsoft Copilot		Open AI Chat GPT-3.5		Open AI Chat GPT-4o		Open AI Chat GPT 4		Google Gemini		Docus AI		Anthropic Claude 2		Perplexity AI		
	T	İ	T	İ	T	İ	T	İ	T	İ	T	İ	T	İ	T	İ	
S1,Q1	0	0	0	0	1	1	0	1	0	0	0	0	1	1	0	0	
S2Q2	0	0	1	1	1	1	1	1	0	0	1	1	1	1	0	1	
S3,Q3	0	0	0	0	1	1	1	1	0	0	0	1	1	1	0	0	
S4,Q4	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	Ağrı ile İlişkili TMD ve Baş Ağrısı
S5,Q5	0	0	1	1	1	1	0	1	0	1	1	0	0	0	1	1	
S6,Q6	0	1	1	1	1	1	1	1	0	0	1	1	0	0	0	0	
S7,Q7	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	
S8,Q8	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	
S9,Q9	0	0	1	1	1	0	1	1	0	1	1	1	0	0	0	1	Eklem İçi Düzensizlikler
S10,Q10	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	
S11,Q11	0	0	1	1	1	1	1	1	0	1	1	1	1	1	0	1	
S12,Q12	0	0	1	1	0	1	1	1	0	0	1	1	1	1	1	0	Dejeneratif Eklem Hastalığı
S13,Q13	0	0	1	1	1	1	1	1	0	0	0	1	1	1	1	0	Subluksasyon
TOPLAM	1	1	9	10	12	11	11	12	0	7	9	11	9	9	6	7	
GENEL TOPLAM	2		19		23		23		7		20		18		13		

Tabloda yer alan puanlamalar, modellerin verdiği yanıtların doğruluğuna göre yapılmış; doğru cevaplar (teşhisler) ‘1’, yanlış, birden fazla teşhis içeren ya da sorunun alt başlığına inmeden belirlenen genel teşhisler ise ‘0’ olarak değerlendirilmiştir.

Dil modelleri arasında 2024 Eylül ayındaki Türkçe sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı farklılık bulunmuştur (p:0,004; p<0,05) (Tablo 4.3). Anthropic Claude 2'nin doğru cevap verme oranı (%76.9), Gemini (%23.1) ve Perplexity AI (%15.4) gruplarından anlamlı şekilde yüksek bulunmuştur (p<0,05). Ayrıca şekil 4.1 ve 4.2'de bu farklılığa dair örnek sunulmuştur. ChatGPT-4o'ın doğru cevap verme oranı (%69.2), Perplexity AI (%15.4) grubundan anlamlı şekilde yüksek bulunmuştur. (p<0,05). Diğer dil modelleri arasında Türkçe sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı bir farklılık bulunamamıştır (Tablo 4.4).

Dil modelleri arasında Eylül ayındaki İngilizce sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı farklılık bulunmuştur (p:0,013; p<0,05) (Tablo 4.3). Anthropic Claude 2'nin doğru cevap verme oranı (%84.6), Gemini (%30.8) ve Perplexity AI (%38.5) gruplarından anlamlı şekilde yüksektir (p<0,05). ChatGPT-4o'ın doğru cevap verme oranı (%84.6), Gemini (%30.8) ve Perplexity AI (%38.5) gruplarından anlamlı şekilde yüksek bulunmuştur (p<0,05). Diğer dil modelleri arasında İngilizce sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı bir farklılık bulunamamıştır (Tablo 4.4).

Microsoft Copilot grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %38.5, İngilizce sorulara doğru cevap verme oranı %61.5 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-3.5 grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %61.5, İngilizce sorulara doğru cevap verme oranı %61.5 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4o grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %69.2, İngilizce sorulara doğru cevap verme oranı %84.6 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

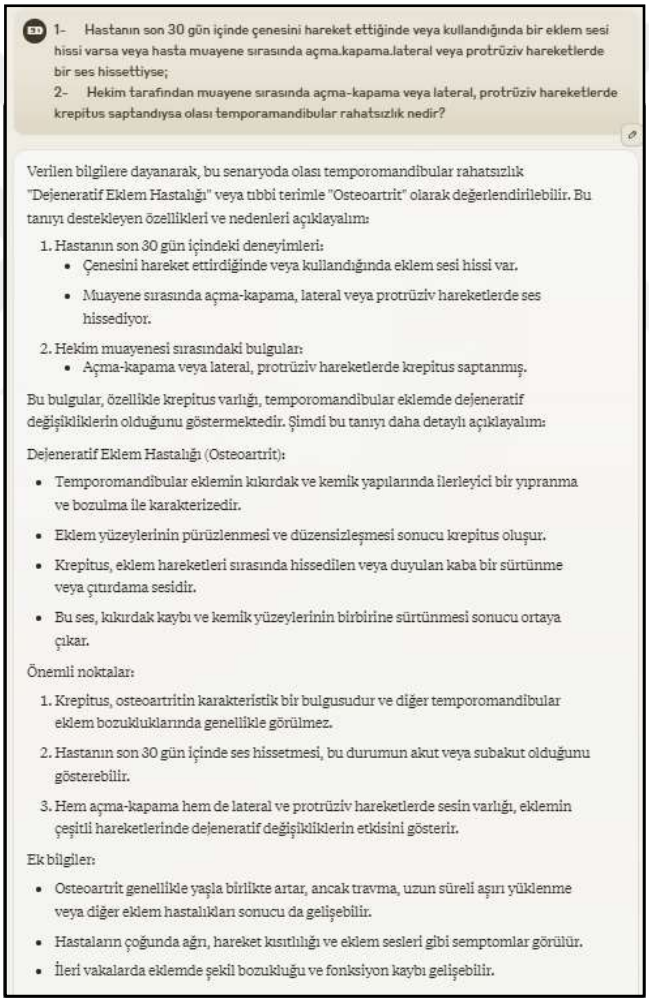
ChatGPT-4 grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %61.5, İngilizce sorulara doğru cevap verme oranı %76.9 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Gemini grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %23.1, İngilizce sorulara doğru cevap verme oranı %30.8 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

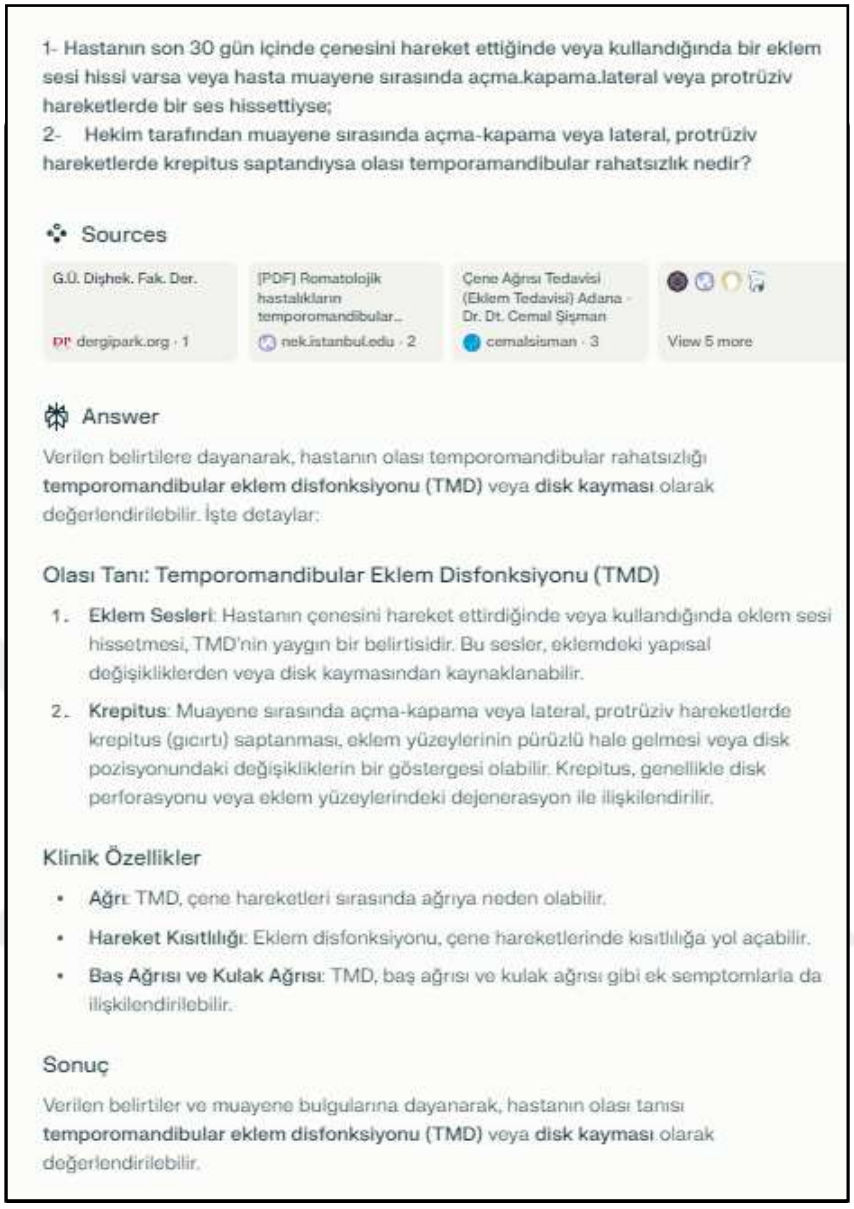
Docus AI grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %46.2, İngilizce sorulara doğru cevap verme oranı %53.8 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Anthropic Claude 2 *grubunda Eylül ayında;* Türkçe sorulara doğru cevap verme oranı %76.9 (şekil 4.1), İngilizce sorulara doğru cevap verme oranı %84.6 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Perplexity AI grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %15.4 (şekil 4.2), İngilizce sorulara doğru cevap verme oranı %38.5 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).



Şekil 4.1: Anthropic Claude 2'ye Ait 2024 Eylül Ayı Türkçe Doğru Cevap Örneği.



Şekil 4.2: Perplexity AI'a Ait 2024 Eylül Ayı Türkçe Yanlış Cevap Örneği.

Tablo 4.3: Büyük Dil Modellerinin 2024 Eylül Ayı Türkçe ve İngilizce Değerlendirilmelerindeki Performanslarının (Doğru Cevap Verme Oranı) Analizi.

	Türkçe (n=13)	**İngilizce (n=13)**	
Dil modelleri (Eylül)	**n (%)**	**n (%)**	**[1]p**
Microsoft Copilot	5 (%38,5)	8 (%61,5)	0,453
ChatGPT-3.5	8 (%61,5)	8 (%61,5)	1,000
ChatGPT-4o	9 (%69,2)	11 (%84,6)	0,625
ChatGPT-4	8 (%61,5)	10 (%76,9)	0,625
Gemini	3 (%23,1)	4 (%30,8)	1,000
Docus AI	6 (%46,2)	7 (%53,8)	1,000
Anthropic Claude 2	10 (%76,9)	11 (%84,6)	1,000
Perplexity AI	2 (%15,4)	5 (%38,5)	0,250
[2]p	0,004*	0,013*	

[1]Mc Nemar Test [2]Cochran's Q Test *p<0,05

Tablo 4.4: 2024 Eylül Ayında Sorulan Tüm Sorulara Verilen Yanıtlar Üzerinden Yapılan Post Hoc Değerlendirme.

		Türkçe	**İngilizce**
Dil modelleri (Eylül)		**p**	**p**
Microsoft Copilot	ChatGPT-3.5	0,453	1.000
	ChatGPT-4o	0,289	0,453
	ChatGPT-4	0,453	0,687
	Gemini	0,625	0,219
	Docus AI	1,000	1,000
	Anthropic Claude 2	0,180	0,375
	Perplexity AI	0,250	0,453
ChatGPT-3.5	ChatGPT-4o	1,000	0,250
	ChatGPT-4	1,000	0,625
	Gemini	0,125	0,219
	Docus AI	0,687	1.000
	Anthropic Claude 2	0,625	0,375
	Perplexity AI	0,070	0,453

Tablo 4.4: 2024 Eylül Ayında Sorulan Tüm Sorulara Verilen Yanıtlar Üzerinden Yapılan Post Hoc Değerlendirme (Devam).

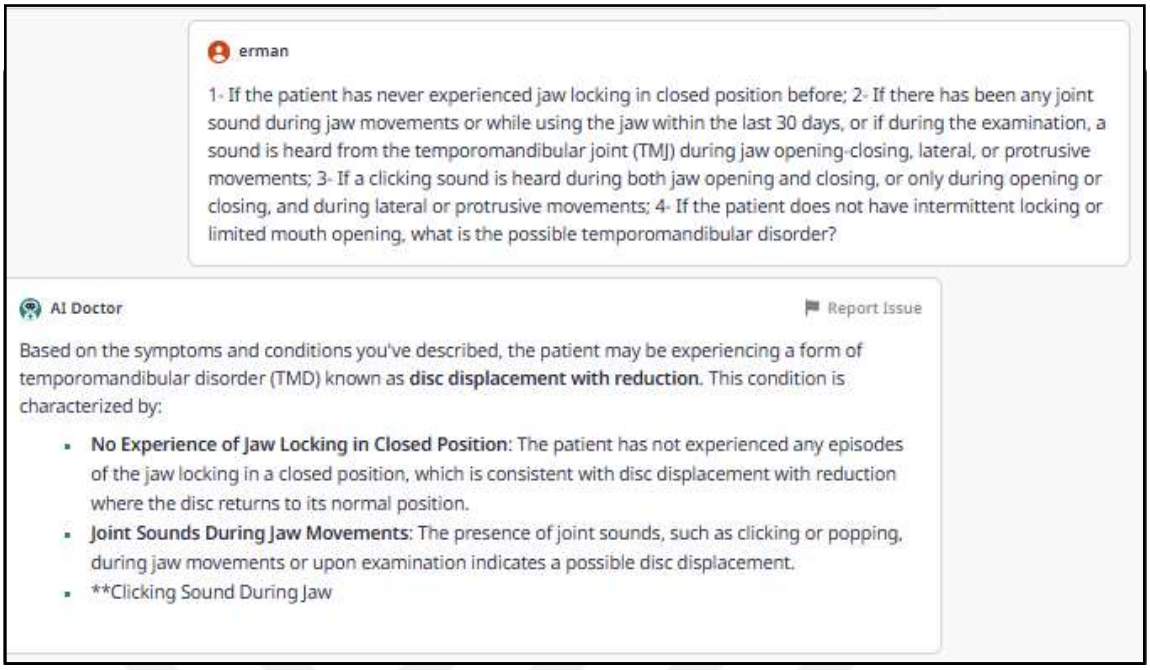
ChatGPT-4o	ChatGPT-4	1,000	1,000
	Gemini	0,070	**0,016***
	Docus AI	0,375	0,219
	Anthropic Claude 2	1,000	1,000
	Perplexity AI	**0,039***	**0,031***
ChatGPT-4	Gemini	0,125	0,070
	Docus AI	0,687	0,453
	Anthropic Claude 2	0,625	1,000
	Perplexity AI	0,070	0,125
Gemini	Docus AI	0,250	0,375
	Anthropic Claude 2	**0,039***	**0,016***
	Perplexity AI	1,000	1,000
Docus AI	Anthropic Claude 2	0,289	0,219
	Perplexity AI	0,125	0,687
Anthropic Claude 2	Perplexity AI	**0,021***	**0,031***

Mc Nemar Test *p<0,05

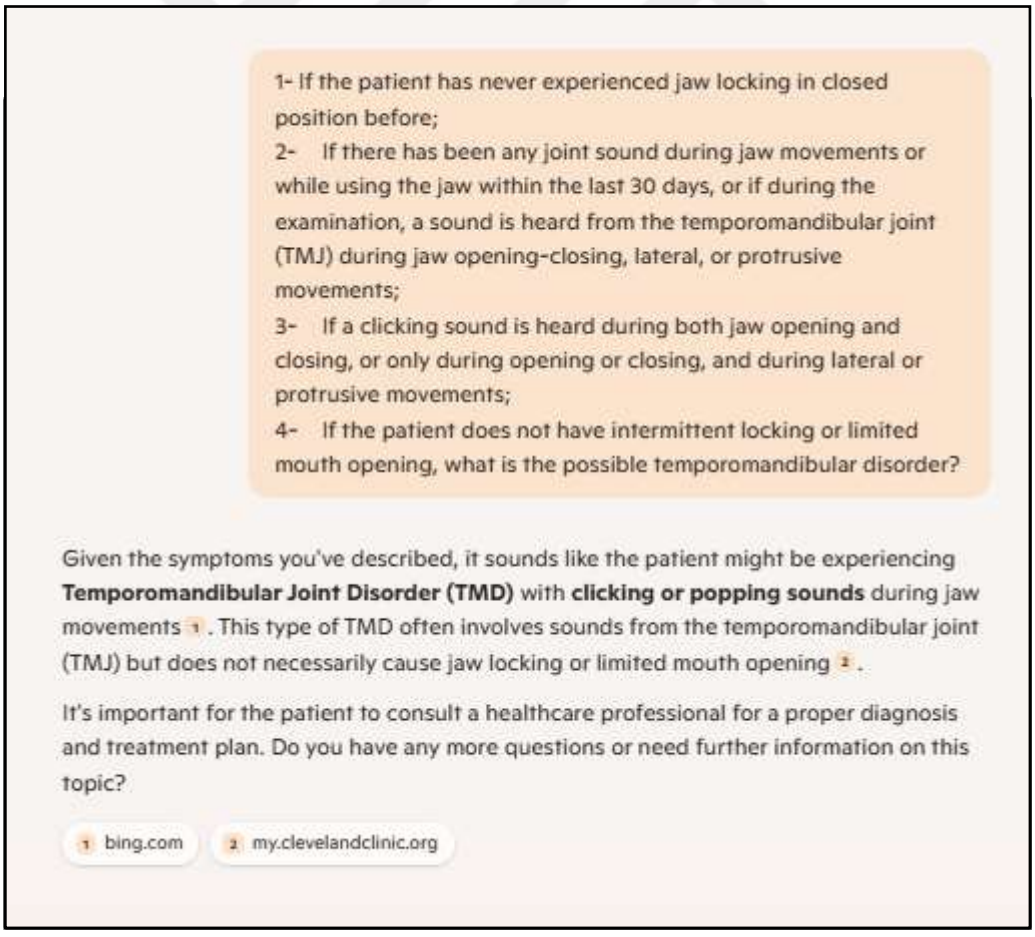
Dil modelleri arasında 2025 Ocak ayındaki Türkçe sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı farklılık bulunmuştur (p:0,001; p<0,05) (Tablo 4.5). Microsoft Copilot (%7.7) ve Gemini'nin (%0) doğru cevap verme oranları, ChatGPT-3.5 (%69.2), ChatGPT-4o (%92.3), ChatGPT-4 (%76.9), Docus AI (%69.2) ve Anthropic Claude 2 (%69.2) gruplarından anlamlı şekilde düşük bulunmuştur (p<0,05). Perplexity AI'nın doğru cevap verme oranı (%46.2), ChatGPT-4o'dan (%92.3) anlamlı şekilde düşük bulunmuştur. (p<0,05). Diğer dil modelleri arasında Türkçe sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı bir farklılık bulunamamıştır (Tablo 4.6).

Dil modelleri arasında Ocak ayındaki İngilizce sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı farklılık bulunmuştur (p:0,001; p<0,05) (Tablo 4.5). Microsoft Copilot'un (%7.7) doğru cevap verme oranları, ChatGPT-3.5 (%76.9), ChatGPT-4o (%84.6), ChatGPT-4 (%92.3), Docus AI (%84.6) ve Anthropic Claude 2 (%69.2) gruplarından anlamlı şekilde düşük bulunmuştur (p<0,05). Ayrıca şekil 4.3 ve 4.4'de bu farklılığa dair örnek sunulmuştur. Diğer dil modelleri arasında İngilizce sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı bir farklılık bulunamamıştır (Tablo 4.6).

Microsoft Copilot grubunda Ocak ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %7.7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).
ChatGPT-3.5 grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %69.2, İngilizce sorulara doğru cevap verme oranı %76.9 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).
ChatGPT-4o grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %92.3, İngilizce sorulara doğru cevap verme oranı %84.6 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).
ChatGPT-4 grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %76.9, İngilizce sorulara doğru cevap verme oranı %92.3 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).
Gemini grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %0, İngilizce sorulara doğru cevap verme oranı %53.8 olup, görülen bu artış istatistiksel olarak anlamlı bulunmuştur (p:0,016; p<0,05).
Docus AI grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %69.2, İngilizce sorulara doğru cevap verme oranı %84.6 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).
Anthropic Claude 2 grubunda Ocak ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %69.2 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).
Perplexity AI grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %46.2, İngilizce sorulara doğru cevap verme oranı %53.8 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).



Şekil 4.3: Docus AI'a Ait 2025 Ocak Ayı Doğru Cevap Örneği.



Şekil 4.4: Microsot Copilot'a Ait 2025 Ocak Ayı Yanlış Cevap Örneği.

Tablo 4.5: Büyük Dil Modellerinin 2025 Ocak Ayı Türkçe ve İngilizce Değerlendirilmelerindeki Performanslarının (Doğru Cevap Verme Oranı) Analizi.

	Türkçe (n=13)	**İngilizce (n=13)**	
Dil modelleri (Ocak)	**n (%)**	**n (%)**	**[1]p**
Microsoft Copilot	1 (%7.7)	1 (%7.7)	1,000
ChatGPT-3.5	9 (%69.2)	10 (%76.9)	1,000
ChatGPT-4o	12 (%92.3)	11 (%84.6)	1,000
ChatGPT-4	10 (%76.9)	12 (%92.3)	0,500
Gemini	0 (%0)	7 (%53.8)	0,016*
Docus AI	9 (%69.2)	11 (%84.6)	0,625
Anthropic Claude 2	9 (%69.2)	9 (%69.2)	1,000
Perplexity AI	6 (%46.2)	7 (%53.8)	1,000
[2]p	0,001*	0,001*	

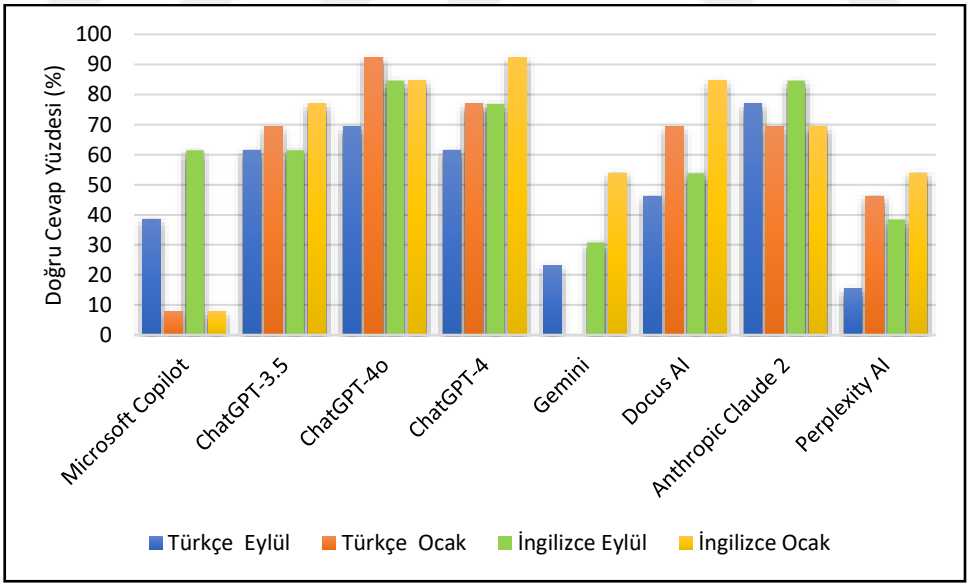
[1]Mc Nemar Test [2]Cochran's Q Test *p<0,05

Tablo 4.6: 2025 Ocak Ayında Sorulan Tüm Sorulara Verilen Yanıtlar Üzerinden Yapılan Post Hoc Değerlendirme.

		Türkçe	**İngilizce**
Dil modelleri (Ocak)		**p**	**p**
Microsoft Copilot	ChatGPT-3.5	**0,008***	**0,004***
	ChatGPT-4o	**0,001***	**0,002***
	ChatGPT-4	**0,004***	**0,001***
	Gemini	1,000	0,070
	Docus AI	**0,008***	**0,002***
	Anthropic Claude 2	**0,008***	**0,021***
	Perplexity AI	0,063	0,070
ChatGPT-3.5	ChatGPT-4o	0,250	1,000
	ChatGPT-4	1,000	0,500
	Gemini	**0,004***	0,375
	Docus AI	1,000	1,000
	Anthropic Claude 2	1,000	1,000
	Perplexity AI	0,375	0,250

Tablo 4.6: 2025 Ocak Ayında Sorulan Tüm Sorulara Verilen Yanıtlar Üzerinden Yapılan Post Hoc Değerlendirme (Devam).

ChatGPT-4o	ChatGPT-4	0,500	1,000
	Gemini	**0,001***	0,289
	Docus AI	0,250	1,000
	Anthropic Claude 2	0,250	0,500
	Perplexity AI	**0,031***	0,219
ChatGPT-4	Gemini	**0,002***	0,125
	Docus AI	1,000	1,000
	Anthropic Claude 2	1,000	0,250
	Perplexity AI	0,219	0,063
Gemini	Docus AI	**0,004***	0,219
	Anthropic Claude 2	**0,004***	0,727
	Perplexity AI	**0,031***	1,000
Docus AI	Anthropic Claude 2	1,000	0,625
	Perplexity AI	0,375	0,219
Anthropic Claude 2	Perplexity AI	0,375	0,687
Mc Nemar Test	*p<0,05		



Şekil 4.5: Tüm Soruları Kapsayan Doğru Cevap Oranları.

Microsoft Copilot grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %38.5, Ocak ayında %7.7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05) (Tablo 4.7 ve Şekil 4.5).

ChatGPT-3.5 grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %61.5, Ocak ayında %69.2 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4o grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %69.2, Ocak ayında %92.3 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4 grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %61.5, Ocak ayında %76.9 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Gemini grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %23.1, Ocak ayında %0 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Docus AI grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %46.2, Ocak ayında %69.2 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Anthropic Claude 2 grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %76.9, Ocak ayında %69.2 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Perplexity AI grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %15.4, Ocak ayında %46.2 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Microsoft Copilot grubunda İngilizce sorularda; Eylül ayında sorulara doğru cevap verme oranı %61.5, Ocak ayında %7.7 olup, görülen bu düşüş istatistiksel olarak anlamlı bulunmuştur (p:0,039; p<0,05).

ChatGPT-3.5 grubunda İngilizce sorularda; Eylül ayında sorulara doğru cevap verme oranı %61.5, Ocak ayında %76.9 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4o grubunda İngilizce sorularda; Eylül ayında ve Ocak ayında sorulara doğru cevap verme oranı %84.6 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4 grubunda İngilizce sorularda; Eylül ayında sorulara doğru cevap verme oranı %76,9, Ocak ayında %92,3 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Gemini grubunda İngilizce sorularda; Eylül ayında sorulara doğru cevap verme oranı %30,8, Ocak ayında %53.8 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Docus AI grubunda İngilizce sorularda; Eylül ayında sorulara doğu cevap verme oranı %53,8, Ocak ayında %84,6 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Anthropic Claude 2 grubunda İngilizce sorularda; Eylül ayında sorulara doğu cevap verme oranı %84,6, Ocak ayında %69,2 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Perplexity AI grubunda İngilizce sorularda; Eylül ayında sorulara doğu cevap verme oranı %38,5, Ocak ayında %53.8 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Tablo 4.7: Büyük Dil Modellerinin Türkçe ve İngilizce Değerlendirmelerindeki 2025 Ocak ve 2024 Eylül Ayındaki Performanslarının (Doğru Cevap Verme Oranı) Analizi.

	Türkçe (n=13)			**İngilizce (n=13)**		
	Eylül	**Ocak**		**Eylül**	**Ocak**	
Dil modelleri	**n (%)**	**n (%)**	**p**	**n (%)**	**n (%)**	**p**
Microsoft Copilot	5 (%38,5)	1 (%7,7)	0,125	8 (%61,5)	1 (%7,7)	0,039*
ChatGPT-3.5	8 (%61,5)	9 (%69,2)	1,000	8 (%61,5)	10 (%76,9)	0,625
ChatGPT-4o	9 (%69,2)	12 (%92,3)	0,375	11 (%84,6)	11 (%84,6)	1,000
ChatGPT-4	8 (%61,5)	10 (%76,9)	0,500	10 (%76,9)	12 (%92,3)	0,500
Gemini	3 (%23,1)	0 (%0)	0,250	4 (%30,8)	7 (%53,8)	0,375
Docus AI	6 (%46,2)	9 (%69,2)	0,375	7 (%53,8)	11 (%84,6)	0,219
Anthropic Claude 2	10 (%76,9)	9 (%69,2)	1,000	11 (%84,6)	9 (%69,2)	0,500
Perplexity AI	2 (%15,4)	6 (%46,2)	0,125	5 (%38,5)	7 (%53,8)	0,625
Mc Nemar Test	*p<0,05					

Ağrı ile ilgili TMD ve baş ağrısı sorularında dil modelleri arasında Eylül ayındaki Türkçe sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p:0,284; p>0,05) (Tablo 4.8).

Ağrı ile ilgili TMD ve baş ağrısı sorularında dil modelleri arasında Eylül ayındaki İngilizce sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p:0,121; p>0,05) (Tablo 4.8).

Microsoft Copilot grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %33.3, İngilizce sorulara doğru cevap verme oranı %66,7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-3.5 grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %66,7, İngilizce sorulara doğru cevap verme oranı %50 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4o grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %66,7, İngilizce sorulara doğru cevap verme oranı %83,3 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4 grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %66,7, İngilizce sorulara doğru cevap verme oranı %83,3 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Gemini grubunda Eylül ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %33,3 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Docus AI grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %33,3, İngilizce sorulara doğru cevap verme oranı %50 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Anthropic Claude 2 grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %83.3, İngilizce sorulara doğru cevap verme oranı %100 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Perplexity AI grubunda Eylül ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %33.3 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Tablo 4.8: Ağrı ile İlgili TMD ve Baş Ağrısı Sorularında Büyük Dil Modellerinin 2024 Eylül Ayı Türkçe ve İngilizce Değerlendirmelerindeki Performanslarının (Doğru Cevap Verme Oranı) Analizi.

Ağrı ile ilgili TMD ve baş ağrısı soruları (Eylül)	**Türkçe (n=6)**	**İngilizce (n=6)**	
	n (%)	**n (%)**	**[1]p**
Microsoft Copilot	2 (%33,3)	4 (%66,7)	0,500
ChatGPT-3.5	4 (%66,7)	3 (%50)	1,000
ChatGPT-4o	4 (%66,7)	5 (%83,3)	1,000
ChatGPT-4	4 (%66,7)	5 (%83,3)	1,000
Gemini	2 (%33,3)	2 (%33,3)	1,000
Docus AI	2 (%33,3)	3 (%50)	1,000
Anthropic Claude 2	5 (%83,3)	6 (%100)	1,000
Perplexity AI	2 (%33,3)	2 (%33,3)	1,000
[2]p	0,284	0,121	

[1]Mc Nemar Test [2]Cochran's Q Test *p<0,05

Eklem içi düzensizlikler sorularında dil modelleri arasında Eylül ayındaki Türkçe sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p:0,193; p>0,05) (Tablo 4.9).

Eklem içi düzensizlikler sorularında dil modelleri arasında Eylül ayındaki İngilizce sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p:0,402; p>0,05) (Tablo 4.9).

Microsoft Copilot grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %40, İngilizce sorulara doğru cevap verme oranı %60 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-3.5 grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %60, İngilizce sorulara doğru cevap verme oranı %80 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4o grubunda Eylül ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %80 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4 grubunda Eylül ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %60 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Gemini grubunda Eylül ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %20 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).
Docus AI grubunda Eylül ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %60 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).
Anthropic Claude 2 grubunda Eylül ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %60 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).
Perplexity AI grubunda Eylül ayında; Türkçe sorulara doğru cevap verme oranı %0 iken İngilizce sorulara doğru cevap verme oranı %40 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Tablo 4.9: Eklem İçi Düzensizlikler Sorularında Büyük Dil Modellerinin 2024 Eylül Ayı Türkçe ve İngilizce Değerlendirmelerindeki Performanslarının (Doğru Cevap Verme Oranı) Analizi.

Eklem içi düzensizlikler soruları (Eylül)	**Türkçe (n=5)**	**İngilizce (n=5)**	
	n (%)	**n (%)**	**[1]p**
Microsoft Copilot	2 (%40)	3 (%60)	1.000
ChatGPT-3.5	3 (%60)	4 (%80)	1.000
ChatGPT-4o	4 (%80)	4 (%80)	1.000
ChatGPT-4	3 (%60)	3 (%60)	1.000
Gemini	1 (%20)	1 (%20)	1.000
Docus AI	3 (%60)	3 (%60)	1.000
Anthropic Claude 2	3 (%60)	3 (%60)	1.000
Perplexity AI	0 (%0)	2 (%40)	0,500
[2]p	0,193	0,402	

[1]Mc Nemar Test [2]Cochran's Q Test *p<0,05

Ağrı ile ilgili TMD ve baş ağrısı sorularında dil modelleri Ocak ayındaki Türkçe sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı farklılık bulunmaktadır (p:0,010; p<0,05) (Tablo 4.10). ChatGPT-4o'ın doğru cevap verme oranı (%100), Microsoft Copilot (%16,7) ve Gemini (%0) gruplarından anlamlı şekilde yüksektir (p<0,05). Diğer dil modelleri arasında Türkçe sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı bir farklılık bulunamamıştır (Tablo 4.11).
Ağrı ile ilgili TMD ve baş ağrısı sorularında dil modelleri 2025 Ocak ayındaki İngilizce sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı farklılık bulunmuştur (p:0,035; p<0,05) (Tablo 4.10). Microsoft Copilot'un doğru cevap verme oranı

(%16,7), ChatGPT-4o (%100) ve ChatGPT-4 (%100) gruplarından anlamlı şekilde düşüktür ($p<0,05$). Diğer dil modelleri arasında İngilizce sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı bir farklılık bulunamamıştır (Tablo 4.11).

Microsoft Copilot grubunda Ocak ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %16,7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır ($p>0,05$).

ChatGPT-3.5 grubunda Ocak ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %66,7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır ($p>0,05$).

ChatGPT-4o grubunda Ocak ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %100 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır ($p>0,05$).

ChatGPT-4 grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %66,7, İngilizce sorulara doğru cevap verme oranı %100 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır ($p>0,05$).

Gemini grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %0, İngilizce sorulara doğru cevap verme oranı %33,3 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır ($p>0,05$).

Docus AI grubunda Ocak ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %66.7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır ($p>0,05$).

Anthropic Claude 2 grubunda Ocak ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %66,7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır ($p>0,05$).

Perplexity AI grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %33,3, İngilizce sorulara doğru cevap verme oranı %50 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır ($p>0,05$).

Tablo 4.10: Ağrı ile İlgili TMD ve Baş Ağrısı Sorularında Büyük Dil Modellerinin 2025 Ocak Ayı Türkçe ve İngilizce Değerlendirmelerindeki Performanslarının (Doğru Cevap Verme Oranı) Analizi.

Ağrı ile ilgili TMD ve baş ağrısı soruları (Ocak)	**Türkçe (n=6)**	**İngilizce (n=6)**	
	n (%)	**n (%)**	**[1]p**
Microsoft Copilot	1 (%16,7)	1 (%16,7)	1.000
ChatGPT-3.5	4 (%66,7)	4 (%66,7)	1.000
ChatGPT-4o	6 (%100)	6 (%100)	1.000
ChatGPT-4	4 (%66,7)	6 (%100)	0,500
Gemini	0 (%0)	2 (%33,3)	0,500
Docus AI	4 (%66,7)	4 (%66,7)	1.000
Anthropic Claude 2	4 (%66,7)	4 (%66,7)	1.000
Perplexity AI	2 (%33,3)	3 (%50)	1.000
[2]p	0,010*	0,035*	

[1]Mc Nemar Test [2]Cochran's Q Test *p<0,05

Tablo 4.11: Ağrı İle İlgili TMD ve Baş Ağrısı Sorularında 2025 Ocak Ayı İçin Post Hoc Değerlendirme Sonuçları.

Ağrı ile ilgili TMD ve baş ağrısı soruları (Ocak)		**Türkçe**	**İngilizce**
		p	**p**
Microsoft Copilot	ChatGPT-3.5	0,250	0,250
	ChatGPT-4o	**0,043***	**0,043***
	ChatGPT-4	0,250	**0,043***
	Gemini	1.000	1.000
	Docus AI	0,250	0,250
	Anthropic Claude 2	0,250	0,375
	Perplexity AI	1.000	0,625
ChatGPT-3.5	ChatGPT-4o	0,500	0,500
	ChatGPT-4	1.000	0,500
	Gemini	0,125	0,500
	Docus AI	1.000	1.000
	Anthropic Claude 2	1.000	1.000
	Perplexity AI	0,500	1.000

Tablo 4.11: Ağrı İle İlgili TMD ve Baş Ağrısı Sorularında 2025 Ocak Ayı İçin Post Hoc Değerlendirme Sonuçları (Devam)

ChatGPT-4o	ChatGPT-4	0,500	1.000
	Gemini	**0,012***	0,125
	Docus AI	0,500	0,500
	Anthropic Claude 2	0,500	0,500
	Perplexity AI	0,125	0,250
ChatGPT-4	Gemini	0,125	0,125
	Docus AI	1.000	0,500
	Anthropic Claude 2	1.000	0,500
	Perplexity AI	0,625	0,250
Gemini	Docus AI	0,125	0,625
	Anthropic Claude 2	0,125	0,625
	Perplexity AI	0,500	1.000
Docus AI	Anthropic Claude 2	1.000	1,000
	Perplexity AI	0,500	1,000
Anthropic Claude 2	Perplexity AI	0,625	1.000

Mc Nemar Test *p<0,05

Eklem içi düzensizlikler sorularında dil modelleri Ocak ayındaki Türkçe sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı farklılık bulunmaktadır (p:0,007; p<0,05) (Tablo 4.12). Microsoft Copilot ve Gemini'nin doğru cevap verme oranları (%0), ChatGPT-4o (%80), ChatGPT-4 (%80) ve Docus AI (%80) gruplarından anlamlı şekilde düşük bulunmuştur (p<0,05). Diğer dil modelleri arasında Türkçe sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı bir farklılık bulunamamıştır (Tablo 4.13).

Eklem içi düzensizlikler sorularında dil modelleri Ocak ayındaki İngilizce sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı farklılık bulunmuştur (p:0,004; p<0,05) (Tablo 4.12). Microsoft Copilot'un doğru cevap verme oranı (%0), ChatGPT-3.5 (%80), ChatGPT-4 (%80), Gemini (%100), Docus AI (%100) ve Perplexity AI (%80) gruplarından anlamlı şekilde düşük bulunmuştur (p<0,05). Diğer dil modelleri arasında İngilizce sorulara doğru cevap verme oranları açısından istatistiksel olarak anlamlı bir farklılık bulunamamıştır (Tablo 4.13).

Microsoft Copilot grubunda Ocak ayında; Türkçe ve İngilizce sorulara doğru cevap vermemiştir.

ChatGPT-3.5 grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %60, İngilizce sorulara doğru cevap verme oranı %80 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4o grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %80, İngilizce sorulara doğru cevap verme oranı %60 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4 grubunda Ocak ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %80 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Gemini grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %0, İngilizce sorulara doğru cevap verme oranı %100 olup bu farklılık istatistiksel olarak anlamlı bulunmuştur (p:0,043; p<0,05).

Docus AI grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %80, İngilizce sorulara doğru cevap verme oranı %100 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Anthropic Claude 2 grubunda Ocak ayında; Türkçe ve İngilizce sorulara doğru cevap verme oranı %60 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Perplexity AI grubunda Ocak ayında; Türkçe sorulara doğru cevap verme oranı %40, İngilizce sorulara doğru cevap verme oranı %80 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Tablo 4.12: Eklem İçi Düzensizlikler Sorularında Büyük Dil Modellerinin 2025 Ocak Ayı Türkçe ve İngilizce Değerlendirmelerindeki Performanslarının (Doğru Cevap Verme Oranı) Analizi.

Eklem içi düzensizlikler soruları (Ocak)	**Türkçe (n=5)**	**İngilizce (n=5)**	
	n (%)	**n (%)**	**[1]p**
Microsoft Copilot	0 (%0)	0 (%0)	1.000
ChatGPT-3.5	3 (%60)	4 (%80)	1.000
ChatGPT-4o	4 (%80)	3 (%60)	1.000
ChatGPT-4	4 (%80)	4 (%80)	1.000
Gemini	0 (%0)	5 (%100)	0,043*
Docus AI	4 (%80)	5 (%100)	1.000
Anthropic Claude 2	3 (%60)	3 (%60)	1.000
Perplexity AI	2 (%40)	4 (%80)	0,500
[2]p	0,007*	0,004*	

[1]Mc Nemar Test [2]Cochran's Q Test *p<0,05

Tablo 4.13: Eklem İçi Düzensizlikler Sorularında 2025 Ocak Ayı İçin Post Hoc Değerlendirme Sonuçları.

Eklem içi düzensizlikler soruları (Ocak)		**Türkçe**	**İngilizce**
		p	**p**
Microsoft Copilot	ChatGPT-3.5	0,250	**0,043***
	ChatGPT-4o	**0,043***	0,250
	ChatGPT-4	**0,043***	**0,043***
	Gemini	-	**0,004***
	Docus AI	**0,043***	**0,004***
	Anthropic Claude 2	0,250	0,250
	Perplexity AI	0,500	**0,043***
ChatGPT-3.5	ChatGPT-4o	1.000	1.000
	ChatGPT-4	1.000	1.000
	Gemini	0,250	1.000
	Docus AI	1.000	1.000
	Anthropic Claude 2	1.000	1.000
	Perplexity AI	1.000	1.000
ChatGPT-4o	ChatGPT-4	1.000	1.000
	Gemini	**0,043***	0,500
	Docus AI	1,000	0,500
	Anthropic Claude 2	1.000	1,000
	Perplexity AI	0,500	1.000
ChatGPT-4	Gemini	**0,043***	1.000
	Docus AI	1.000	1.000
	Anthropic Claude 2	1.000	1.000
	Perplexity AI	0,500	1.000
Gemini	Docus AI	**0,043***	1.000
	Anthropic Claude 2	0,250	0,500
	Perplexity AI	0,500	1.000
Docus AI	Anthropic Claude 2	1.000	0,500
	Perplexity AI	0,500	1.000
Anthropic Claude 2	Perplexity AI	1.000	1.000

Mc Nemar Test *p<0,05

Ağrı ile ilgili TMD ve baş ağrısı sorularında (Şekil 4.6 ve Tablo 4.14);

Microsoft Copilot grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %33.3, Ocak ayında %16.7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-3.5 grubunda Türkçe sorularda; Eylül ve Ocak ayında sorulara doğru cevap verme oranı %66,7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4o grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %66,7, Ocak ayında %100 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4 grubunda Türkçe sorularda; Eylül ve Ocak ayında sorulara doğru cevap verme oranı %66,7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Gemini grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %33,3, Ocak ayında %0 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Docus AI grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %33.3, Ocak ayında %66,7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Anthropic Claude 2 grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %83,3, Ocak ayında %66,7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Perplexity AI grubunda Türkçe sorularda; Eylül ve Ocak ayında sorulara doğru cevap verme oranı %33,3 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Microsoft Copilot grubunda İngilizce sorularda; Eylül ayında sorulara doğru cevap verme oranı %66,7, Ocak ayında %16,7 olup, görülen bu düşüş istatistiksel olarak anlamlı bulunmuştur (p:0,039; p<0,05).

ChatGPT-3.5 grubunda İngilizce sorularda; Eylül ayında sorulara doğru cevap verme oranı %50, Ocak ayında %66,7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4o grubunda İngilizce sorularda; Eylül ayında sorulara doğru cevap verme oranı %83.3, Ocak ayında %100 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

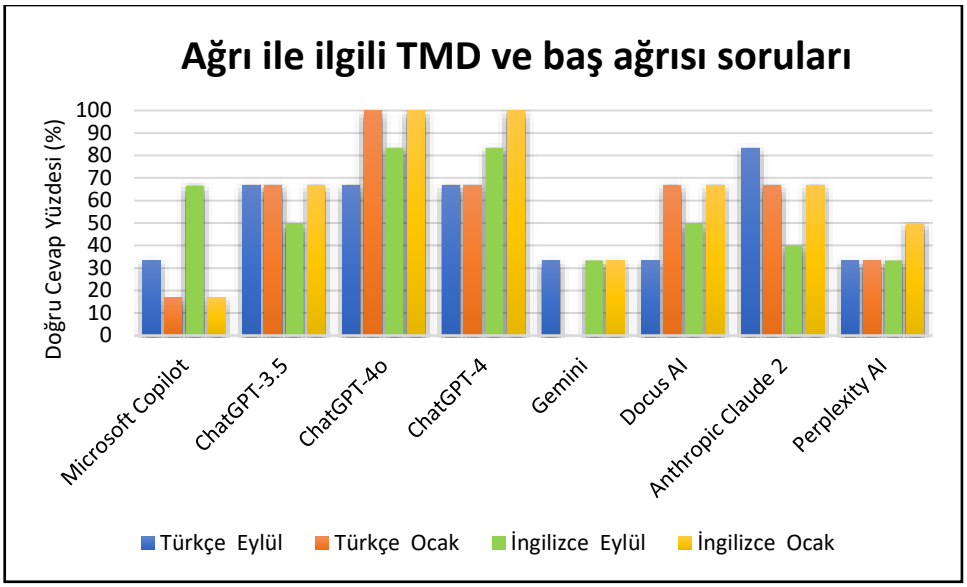
ChatGPT-4 grubunda İngilizce sorularda; Eylül ayında sorulara doğru cevap verme oranı %83.3, Ocak ayında %100 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Gemini grubunda İngilizce sorularda; Eylül ve Ocak ayında sorulara doğru cevap verme oranı %33.3 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Docus AI grubunda İngilizce sorularda; Eylül ayında sorulara doğu cevap verme oranı %50, Ocak ayında %66.7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Anthropic Claude 2 grubunda İngilizce sorularda; Eylül ayında sorulara doğu cevap verme oranı %40, Ocak ayında %66.7 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Perplexity AI grubunda İngilizce sorularda; Eylül ayında sorulara doğu cevap verme oranı %33.3, Ocak ayında %50 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

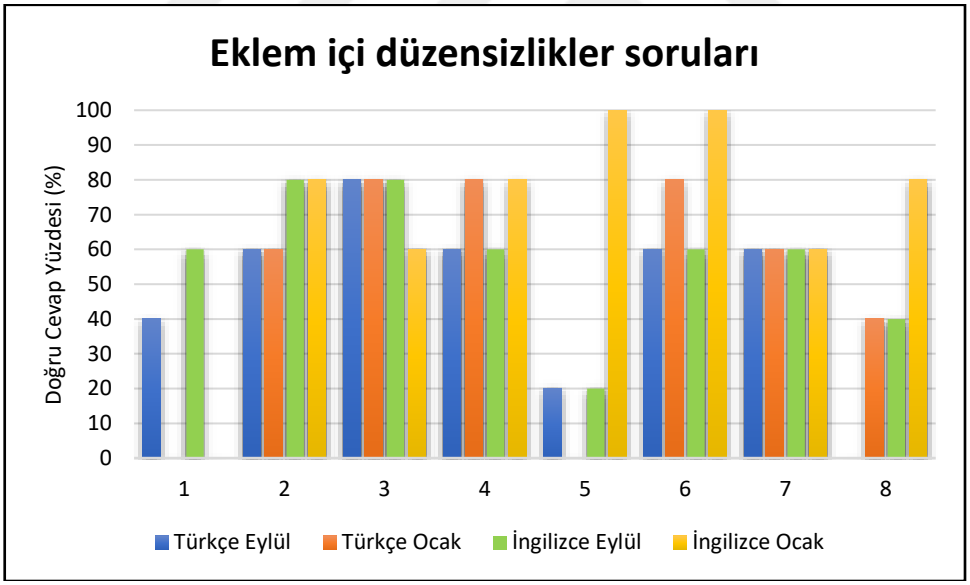


Şekil 4.6: Ağrı ile İlgili TMD ve Baş Ağrısı Sorularında Doğru Cevap Oranları.

Tablo 4.14: Ağrı ile İlgili TMD ve Baş Ağrısı Sorularında Büyük Dil Modellerinin Türkçe ve İngilizce Değerlendirmelerindeki 2025 Ocak ve 2024 Eylül Ayındaki Performanslarının (Doğru Cevap Verme Oranı) Analizi.

	Türkçe (n=6)			İngilizce (n=6)		
	Eylül	Ocak		Eylül	Ocak	
Dil modelleri	**n (%)**	**n (%)**	**p**	**n (%)**	**n (%)**	**p**
Microsoft Copilot	2 (%33,3)	1 (%16,7)	1.000	4 (%66,7)	1 (%16,7)	0,375
ChatGPT-3.5	4 (%66,7)	4 (%66,7)	1.000	3 (%50)	4 (%66,7)	1.000
ChatGPT-4o	4 (%66,7)	6 (%100)	0,500	5 (%83,3)	6 (%100)	1.000
ChatGPT-4	4 (%66,7)	4 (%66,7)	1.000	5 (%83,3)	6 (%100)	1.000
Gemini	2 (%33,3)	0 (%0)	0,500	2 (%33,3)	2 (%33,3)	1.000
Docus AI	2 (%33,3)	4 (%66,7)	0,500	3 (%50)	4 (%66,7)	1.000
Anthropic Claude 2	5 (%83,3)	4 (%66,7)	1.000	2 (%40)	4 (%66,7)	0,500
Perplexity AI	2 (%33,3)	2 (%33,3)	1.000	2 (%33,3)	3 (%50)	1.000

Mc Nemar Test



Şekil 4.7: Eklem İçi Düzensizlikler Sorularında Doğru Cevap Oranları.

Eklem içi düzensizlikler sorularında (Şekil 4.7 ve Tablo 4.15);

Microsoft Copilot grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %40, Ocak ayında %0 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-3.5 grubunda Türkçe sorularda; Eylül ve Ocak ayında sorulara doğru cevap verme oranı %60 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4o grubunda Türkçe sorularda; Eylül ve Ocak ayında sorulara doğru cevap verme oranı %80 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4 grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %60, Ocak ayında %80 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Gemini grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %20, Ocak ayında %0 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Docus AI grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %60, Ocak ayında %80 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Anthropic Claude 2 grubunda Türkçe sorularda; Eylül ve Ocak ayında sorulara doğru cevap verme oranı %60 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Perplexity AI grubunda Türkçe sorularda; Eylül ayında sorulara doğru cevap verme oranı %0, Ocak ayında %40 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Microsoft Copilot grubunda İngilizce sorularda; Eylül ayında sorulara doğru cevap verme oranı %60, Ocak ayında %0 olup, görülen bu düşüş istatistiksel olarak anlamlı bulunmamıştır (p>0,05).

ChatGPT-3.5 grubunda İngilizce sorularda; Eylül ve Ocak ayında sorulara doğru cevap verme oranı %80 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4o grubunda İngilizce sorularda; Eylül ayında sorulara doğru cevap verme oranı %80, Ocak ayında %60 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

ChatGPT-4 grubunda İngilizce sorularda; Eylül ayında sorulara doğru cevap verme oranı %60, Ocak ayında %8 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Gemini grubunda İngilizce sorularda; Eylül ayında sorulara doğru cevap verme oranı %20, Ocak ayında %100 olup, görülen bu artış istatistiksel olarak anlamlı bulunmamıştır (p>0,05).
Docus AI grubunda İngilizce sorularda; Eylül ayında sorulara doğu cevap verme oranı %60, Ocak ayında %100 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).
Anthropic Claude 2 grubunda İngilizce sorularda; Eylül ve Ocak ayında sorulara doğru cevap verme oranı %60 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).
Perplexity AI grubunda İngilizce sorularda; Eylül ayında sorulara doğu cevap verme oranı %40, Ocak ayında %80 olup, aralarında istatistiksel olarak anlamlı bir farklılık bulunamamıştır (p>0,05).

Tablo 4.15: Eklem İçi Düzensizlikler Sorularında Büyük Dil Modellerinin Türkçe ve İngilizce Değerlendirmelerindeki 2025 Ocak ve 2024 Eylül Ayındaki Performanslarının (Doğru Cevap Verme Oranı) Analizi.

	Türkçe (n=5)			**İngilizce (n=5)**		
	Eylül	**Ocak**		**Eylül**	**Ocak**	
Dil modelleri	**n (%)**	**n (%)**	**p**	**n (%)**	**n (%)**	**p**
Microsoft Copilot	2 (%40)	0 (%0)	0,500	3 (%60)	0 (%0)	0,250
ChatGPT-3.5	3 (%60)	3 (%60)	1.000	4 (%80)	4 (%80)	1.000
ChatGPT-4o	4 (%80)	4 (%80)	1.000	4 (%80)	3 (%60)	1.000
ChatGPT-4	3 (%60)	4 (%80)	1.000	3 (%60)	4 (%80)	1.000
Gemini	1 (%20)	0 (%0)	1.000	1 (%20)	5 (%100)	0,125
Docus AI	3 (%60)	4 (%80)	1.000	3 (%60)	5 (%100)	0,500
Anthropic Claude 2	3 (%60)	3 (%60)	1.000	3 (%60)	3 (%60)	1.000
Perplexity AI	0 (%0)	2 (%40)	0,500	2 (%40)	4 (%80)	0,500

Mc Nemar Test

5. TARTIŞMA

Temporomandibular eklem düzensizlikleri, çene eklemini, çiğneme kaslarını ve bu bölgedeki diğer anatomik yapıları etkileyen, çeşitli klinik belirtilerle ortaya çıkan bir sağlık sorunudur. Buna bağlı olarak kişiler konuşma, beslenme ve gün içerisinde yaptıkları basit çene hareketlerinde zorluk yaşayabilir. Ayrıca hastalar sıklıkla baş ve boyun bölgesinde uzun süre devam eden ağrılardan şikâyet ederler. Bu durum, bireylerin günlük yaşam aktivitelerini olumsuz yönde etkileyebilmektedir [37], [108], [109].

Temporomandibular düzensizliklerin semptomları, klinik bulguları ve toplum içindeki görülme sıklığını araştıran çalışmalarda birbirinden oldukça farklı sonuçlar elde edilmiştir. Bu farklılıkların ortaya çıkmasında, araştırmacıların hastalığı tanımlama şekillerindeki farklılıklar ve çalışma yöntemlerindeki çeşitlilik önemli rol oynamaktadır [110].

Temporomandibular rahatsızlıkların sınıflandırılmasında yaşanan güçlükleri gidermek ve bilimsel çalışmalara standart teşhis ölçütleri sunmak amacıyla Samuel F. Dworkin ve Linda Le Resche liderliğinde oluşan bir çalışma grubu tarafından "Temporomandibular Rahatsızlıklar için Araştıma Teşhis Kriterleri" (TMR/ATK) geliştirilmiştir. Dworkin ve çalışma arkadaşları tarafından oluşturulan bu kriterler, uzun süre pek çok araştırmada yaygın biçimde kullanılmıştır. Ancak TMR/ATK sisteminin zaman içinde ortaya çıkan eksiklikleri sebebiyle 2014 yılında Richard Ohrbach ve Eric Schiffman öncülüğünde uluslararası bir araştırmacı ekibi tarafından "Temporomandibular Düzensizlikler için Tanı Kriterleri" (TMD/TK) hazırlanarak yayımlanmıştır [40], [47], [111].

Bu çalışma Temporomandibuler Düzensizlikler için Tanı Kriterlerine (TMD/TK) göre günümüzde yaygın kullanıma sahip büyük dil modellerinin, temporomandibuler eklem düzensizliklerinin teşhisindeki performanslarını Türkçe ve İngilizce dillerinde değerlendirmek, modellerin performanslarını birbirleriyle karşılaştırmak ve bu performansın zaman içerisindeki değişimini incelemeyi hedeflemiştir. Elde edilen bulgular, büyük dil modellerinin temporomandibular eklem düzensizlikleri tanısında dikkate değer bir doğruluk sergileyebildiğini ancak model tipine, zamana ve dile bağlı olarak performanslarında varyasyonlar olabileceğini göstermektedir. Dolayısıyla “Yapay zekaya dayalı büyük dil modellerinin temporomandibuler eklem düzensizliklerini teşhis etme performansları arasında anlamlı farklılık yoktur ve büyük dil modellerinin performansı farklı bir dil

kullanıldığında veya farklı zamanlarda yapılan değerlendirmelerde değişmez" sıfır hipotezi reddedilmiştir. Her ne kadar birçok karşılaştırmada modeller arasında benzer başarı oranları görülmüş olsa da, tespit edilen istatistiksel anlamlılıklar, tüm modellerin eş düzeyde performans sergilemediğini göstermektedir. Bu durum, büyük dil modellerinin temporomandibular eklem düzensizliklerinin tanısındaki yeteneğinin özdeş olmadığını ve performanslarının zaman veya dile bağlı olarak değişebileceğini göstermektedir.

Literatürdeki önceki çalışmalar, büyük dil modellerinin teşhis süreçlerindeki kullanımını genellikle sınırlı sayıda modelle incelemiştir. Örneğin, Kanjee ve arkadaşları çalışmalarında sadece ChatGPT-4 modelinin New England Journal of Medicine'de yayımlanan 70 adet klinikopatolojik konferans (CPC) vakasında hekimlere tanısal destek olarak etkisini değerlendirmiştir [112]. Benzer şekilde, Ríos-Hoyo'nun Massachusetts General Hospital'a ait ve New England Journal of Medicine'da yayımlanmış 75 ardışık karmaşık klinik vakayı temel alan çalışmasında, büyük dil modellerinin tanısal destek aracı olarak kullanım potansiyeli değerlendirilmiştir, Karahan'ın yürüttüğü çalışmada ise büyük dil modellerinin ağız ve çene cerrahisine yönelik sorulara yanıt verebilme kapasiteleri incelenmiştir. Her iki araştırmada da OpenAI'ın ChatGPT-3.5 ve ChatGPT-4 modelleri arasında sınırlı bir karşılaştırma gerçekleştirilmiştir [113], [114]. BDM yelpazesinin bu denli dar tutulması, elde edilen bulguların genellenebilirliğini kısıtlayarak farklı model mimarileri arasındaki performans farklarının anlaşılmasını zorlaştırmaktadır.

Bu çalışmada ise sekiz farklı büyük dil modeli (Microsoft Copilot, ChatGPT-3.5, ChatGPT-4, ChatGPT-4o, Google Gemini, Anthropic Claude 2, Docus AI, Perpexity AI) sistematik olarak değerlendirilmiştir.

Böylesine geniş bir model yelpazesinin incelenmesi, modeller arası performans farklılıklarının çok daha kapsamlı biçimde analiz edilmesine olanak tanımıştır. Farklı altyapı ve eğitim verisine sahip bu modellerin aynı tanı görevlerinde nasıl sonuçlar verdiği karşılaştırılarak, belli bir alanda en başarılı ya da tutarlı modelin hangisi olduğu konusunda daha sağlam çıkarımlar yapılması amaçlanmıştır.

Mevcut literatürdeki çalışmaların büyük çoğunluğu yalnızca tek bir dilde, özellikle de İngilizce odaklı olarak gerçekleştirilmiştir. Örneğin, Gupta'nın çalışmasında büyük dil modelleri belirtilere dayalı yaygın hastalıkların teşhisinde tanısal destek aracı olarak değerlendirilirken; Moglia ve arkadaşlarının araştırmasında ise robotik cerrahiye yönelik

çevrim içi testlerde kullanılmıştır [115], [116]. Her iki çalışmada da BDM'ler yalnızca İngilizce vakalar ve sorular üzerinden test edilmiş, çok dilli performans analizi gerçekleştirilmemiştir. Mauro Giuffrè ve arkadaşlarının yaptığı bir derlemede de büyük dil modellerinin tıbbi sorulara yanıt verirkenki doğruluğu yalnızca İngilizce içerik üzerinden değerlendirilmiştir [117]. Ancak modellerin farklı dillerdeki performansını anlamak kritik önemdedir; çünkü yapılan araştırmalarda ChatGPT'nin farklı dillerdeki başarısının önemli ölçüde değiştiği, özellikle İngilizce dilinde en yüksek doğruluğa ulaştığı bildirilmiştir [118]. Tek dille sınırlı değerlendirmeler, BDM'lerin çok dilli ortamlardaki etkinliğine dair eksik bir tablo sunmaktadır.

Bu çalışmada dil bariyerinin etkilerini incelemek için çift dilli bir değerlendirme yaklaşımı benimsenmiştir. Tanı senaryoları Türkçe ve İngilizce olmak üzere iki farklı dilde modellere sunulmuş ve modellerin her iki dilde nasıl performans gösterdiği analiz edilmiştir. İki dilde test edilmesi sayesinde, büyük dil modellerinin dil bariyerini aşıp aşamadığı, yani benzer doğruluk ve tutarlılığı farklı dillerde koruyup koruyamadığı gözlemlenmiştir. Sonuçlar, modellerin çift dilli performans tutarlılığını ortaya koyarak, pratikte dil engelinin aşılmasında bu yapay zekâ araçlarının potansiyelini değerlendirmeye olanak tanımıştır.

Önceki araştırmaların çoğu, modelleri tek seferlik bir değerlendirmeye tabi tutmuştur; dolayısıyla bir BDM'nin zaman içerisindeki performans tutarlılığı veya olası öğrenme/ güncellenme etkileri yeterince incelenmemiştir. Oysa büyük dil modelleri, sürümleri tamamlandıktan sonra bile güncellemeler veya kullanım şekline bağlı değişikliklerle farklı zamanlarda farklı performans sergileyebilmektedir. Nitekim, kısa aralıklarla yayınlanan model güncellemelerinin bile davranış değişikliklerine yol açabileceği gösterilmiştir. Örneğin ChatGPT-4 modelinin Mart 2023 ve Haziran 2023 sürümlerini karşılaştıran bir çalışmada, aynı modelin farklı versiyonları arasında matematik problemleri çözme ve talimatlara uyma becerileri gibi konularda belirgin performans farkları saptanmıştır [119]. Mevcut çalışmada da bu çalışmayla uyumlu sonuçlar bulunmuştur. Bu bulgu, bir BDM servisinin "aynı" kalmasına rağmen kısa süre içerisinde önemli davranış değişimleri sergileyebileceğini ve dolayısıyla sürekli izlenmesi gerektiğini ortaya koymaktadır.

Bu araştırmada modellerin zamanla değişen performansını değerlendirebilmek için iki ayrı zaman diliminde ölçüm yapılmıştır. İki farklı veri seti oluşturularak, TMD/TK'nın tanı karar ağacından oluşturulan soru şablonları Eylül 2024 ve Ocak 2025'te modellere sorulmuştur.

Bu iki aşamalı tasarım sayesinde, BDM'lerin yaklaşık dört aylık bir sürede sergilediği performans değişiklikleri analiz edilebilmiştir. Olası model güncellemelerinin veya algoritmik iyileştirmelerin tanısal doğruluk ve tutarlılık üzerindeki etkileri, birinci ve ikinci değerlendirme sonuçları karşılaştırılarak incelenmiştir. Sonuç olarak, her bir modelin zaman içindeki tutarlılığı (performansını koruyup koruyamadığı) veya gelişimi (daha iyi hale gelip gelmediği) hakkında veriler elde edilmiştir. Bu dinamik değerlendirme, BDM'lerin statik değil, zamanla değişebilen araçlar olduğunu vurgulayarak, klinik kullanımda düzenli aralıklarla yeniden değerlendirilmesi gerekliliğine dikkat çekmektedir.

Bu çalışmayla, TMD/TK'da belirtilen temporomandibular düzensizlikler ayrıntılı bir şekilde listelenmiş ve her bir düzensizlik için tanı kriterlerine uygun olarak soru kalıpları hazırlanmıştır. Bu soru hazırlama süreci, TMD/TK'nın tanı karar ağacından faydalanarak düzensizliklerin yapısal ve fonksiyonel özelliklerini dikkate alarak gerçekleştirilmiştir. Daha sonra, bu sorular yapay zeka destekli büyük dil modellerine yöneltilmiştir ve temporomandibular eklem düzensizliklerini teşhis performansı değerlendirilmiştir.

Medikal rahatsızlıkların teşhisi için yapay zeka uygulamaları ve büyük dil modellerinden faydalanmak hem klinik uygulamalar için hem de bilimsel araştırmalar için önemli derecede kolaylık sağlamaktadır. Yapay zeka destekli BDM'nin fazla miktardaki veriyi kısa sürede analiz etme becerisi, teşhis süresini oldukça kısaltmaktadır. Tüm bunların yanında yapay zeka destekli bu modeller güncel verileri analiz ederek başarılı teşhis koyma yeteneklerini zaman içerisinde geliştirebilmektedir, yapay zekanın sürekli kendini yenileyebilme becerisi teşhislerde daha başarılı ve güvenilir olmasını sağlamaktadır. Günümüzde BDM, medikal teşhis sürecinde önemli bir rol oynadığından bu kıyaslama çalışması hem modellerin doğruluğunu sınamak hem de aralarındaki farkı ortaya koymak açısından önemlidir.

Büyük dil modelleri metin tabanlı girdilerden insan benzeri yanıtlar oluşturabilen makine öğrenimi teknolojileridir. Araştırmalar, bu modellerin karmaşık vakaları analiz edebilme, klinik karar süreçlerini insan gibi yürütebilme, tanı belirleme, hasta öyküsünü değerlendirme gibi yeteneklere sahip olduğunu göstermektedir. Geniş kullanım alanları sayesinde, büyük dil modelleri sağlık sektörünün farklı alanlarına giderek daha fazla entegre edilmektedir [112], [120], [121], [122].

Büyük dil modelleri, insan dilini kavrayıp yorumlayabilmek için geniş ölçekli metin veri kümeleriyle eğitilmiş, yapay sinir ağları temelli gelişmiş yapay zeka sistemleri olarak da

tanımlanmaktadır. Tıp alanında farklı görevlerde kullanılan bu modeller; radyoloji raporlarının daha anlaşılır hale getirilmesi, internet forumlarında hastaların tıbbi sorularına yanıt verilmesi, tıbbi özetler oluşturulması ve hekimlerin tanı sürecini desteklemesi gibi uygulamalarda yer bulmaktadır [123], [124], [125], [126]. BDM'ler, çeşitli araştırmalarda tıbbi lisans sınavı sorularını belirli bir doğruluk oranıyla yanıtlayabilme yeteneği sergilese de, bu başarının gerçek klinik ortamlarda karar destek mekanizması olarak ne derece etkin olabileceği halen kesinleşmemiştir. Bunun temel sebebi, gerçek vakaların standart test sorularına kıyasla çok daha fazla ayrıntı ve karmaşıklık içermesidir. Geniş kullanım alanları göz önüne alındığında, BDM'lerin tıp öğrencileri ve hekimler için tanı koyma sürecinde yardımcı bir araç olarak önerildiği ve bazı modellerin performansına dair sınırlı bilgiye rağmen bu amaçla kullanılmasının mümkün olduğu düşünülmektedir [123], [124], [125], [126], [127], [128]. Tıbbi bilgi konusunda kapsamlı bir semantik anlayışa sahip olmaları ve klinik akıl yürütme yetenekleri, onların tıbbi soruları yanıtlayabilme, radyoloji raporlarını sadeleştirme ve tıbbi sınavlarda yüksek başarı gösterme gibi avantajlarını ortaya koymaktadır [114], [123], [124], [125], [126], [129].

Kanjee ve ekibi tarafından gerçekleştirilen bir araştırmada ChatGPT-4 modelinin 70 adet klinikopatolojik vakadan elde edilen teşhislerle uyumu incelenmiştir. Bu araştırmanın sonuçlarına göre, ChatGPT-4'ün ilk öngördüğü teşhis, nihai teşhisle %39 oranında eşleşirken, ayırıcı tanı listesinde doğru teşhisin bulunma oranı %64 olarak tespit edilmiştir [112].

Ríos-Hoyo ve ekibinin Massachusetts General Hospital'den seçtiği 75 ardışık klinik vaka üzerinde yaptığı çalışmada, OpenAI'nin ChatGPT-4 modelinin karmaşık vakaları doğru teşhis etme konusunda ChatGPT-3.5'e kıyasla daha başarılı olduğu, ancak yanlış teşhislerin yaygın olduğu ve bu modellerin mevcut haliyle en fazla bir karar destek aracı olarak kullanılabileceği belirtilmiştir [114].

Gupta ve ekibi tarafından yürütülen bir araştırmada, ChatGPT-4, ChatGPT-4o, Gemini, o1 Preview ve ChatGPT-3.5 gibi beş ileri düzey büyük dil modelinin teşhis yetenekleri ve doğruluk seviyeleri, adil bir karşılaştırma yapılabilmesi için yapılandırılmış bir çerçeve kullanılarak sistematik bir şekilde incelenmiştir. Çalışmanın bulguları, ChatGPT-4'ün geniş veri setleriyle eğitilmesinin, karmaşık tıbbi vakaları doğru analiz etme konusunda üstün performans sağladığını ortaya koymaktadır. Gemini, yüksek riskli durumlarda duyarlılığı ön

planda tutarak yanlış pozitif sonuçları en aza indirirken, ChatGPT-4o ve o1 Preview, güvenilir teşhis doğruluğu ile gerçek zamanlı kullanım arasında dengeli bir performans sunmaktadır. ChatGPT-3.5 ise daha temel bir model olmasına rağmen, genel teşhis süreçlerinde ve sınırlı kaynaklara sahip sağlık ortamlarında etkili bir yardımcı araç olarak öne çıkmaktadır [115].

Moglia ve arkadaşları tarafından gerçekleştirilen çalışmada, 2023 yılında yayımlanan beş farklı ChatGPT sürümü ile bir ChatGPT-4 modeli, robotik cerrahi testine tabi tutulmuştur. Değerlendirme kapsamında, her bir sürüm için yedi deneme yapılmış olup, geçme notu %79,5 olarak belirlenmiştir. Çalışmanın sonuçlarına göre, ChatGPT'nin beş farklı sürümü, sırasıyla %64,6, %65,6, %75,0, %78,9 ve %72,7 oranında başarı elde ederken, ChatGPT-4 modeli %91,5 puana ulaşarak tüm denemelerde en yüksek performansı sergilemiştir. ChatGPT-4'ün sonuçları, ChatGPT ile kıyaslandığında istatistiksel olarak anlamlı bir üstünlük göstermektedir. Bu araştırmada elde edilen bulgular, ChatGPT'nin robotik cerrahi testlerindeki performansının ilk sürümde %64,6 seviyesinden dördüncü sürümde %78,9'a kadar yükseldiğini, ancak beşinci sürümde %72,7'ye düştüğünü ortaya koymaktadır. Özellikle, ChatGPT'nin ilk iki ve beşinci sürümlerinde gerçekleştirilen yedi denemenin hiçbirinde robotik cerrahi testini geçemediği, ancak üçüncü ve dördüncü sürümlerde bu başarıya ulaştığı tespit edilmiştir. Bununla birlikte, modelin tutarlı bir başarı sergileyememesi, zaman içinde sürüm güncellemeleriyle verdiği yanıtların değişkenlik gösterdiğini ortaya koymaktadır [116]. Mevcut çalışmanın bulguları da büyük dil modellerinin ardışık sürüm güncellemeleriyle birlikte yanıt içeriklerinde ve doğruluk düzeylerinde anlamlı değişiklikler meydana geldiğini ortaya koymuştur.

ChatGPT-4'ün, sıfır atış (zero-shot) ve az atış öğrenme (few-shot learning) stratejilerini kullanarak gerçekleştirilen değerlendirmelerde, ABD Tıp Lisanslama Sınavı (USMLE) ve Çince ile Japonca sorular içeren diğer ulusal yeterlilik sınavlarını önemli bir farkla geçtiği tespit edilmiştir. Ayrıca, ChatGPT-4'ün Japonya Ulusal Tıp Uygulayıcıları Yeterlilik Sınavı, Kore Geleneksel Tıp Doktorları Ulusal Lisanslama Sınavı, Birleşik Krallık Nöroloji Uzmanlık Sertifika Sınavı, Amerikan Nöroşirürji Kurulu sınavı, Amerikan Anesteziyoloji Kurulu sınavı, Kore Genel Cerrahi Kurul Sınavları, Avustralya Acil Tıp Koleji sınavı ve Dermatoloji Uzmanlık Sertifika Sınavı gibi sınavlarda ChatGPT ve Bard'a kıyasla daha

üstün performans sergilediği görülmüştür [130], [131], [132], [133], [134], [135], [136], [137], [138].

Giuffrè ve ekibi tarafından yapılan değerlendirme, büyük dil modellerinin sindirim hastalıkları alanındaki performansını incelemiş ve bu modellerin klinik kullanım potansiyeline sahip olmasına rağmen, mevcut doğruluk ve güvenilirlik seviyelerinin tıbbi uygulamalar için yeterli olmadığını ortaya koymuştur. Çalışmanın bulguları, büyük dil modellerinin sindirim hastalıklarına yönelik tanı ve tedavi süreçlerinde güvenilir bir karar destek aracı olarak kullanılabilmesi için daha fazla iyileştirme ve doğrulama gerektirdiğini vurgulamaktadır [117].

Huo ve ekibi, ChatGPT-4, Copilot, Google Bard ve Perplexity AI gibi yapay zeka destekli sohbet botlarının gastroözofageal reflü hastalığının cerrahi yönetimi konusundaki önerilerini değerlendirmiştir. Çalışma sonuçlarına göre, Google Bard, hem hekimler hem de hastalar açısından en doğru ve kapsamlı önerileri sunarken, ChatGPT-4 benzer şekilde yüksek doğruluk oranına sahip olmuştur. Copilot ve Perplexity AI ise diğer modellere kıyasla daha düşük doğruluk seviyeleri sergilemiştir [139].

Lim ve arkadaşları, ChatGPT-3.5, Claude, Gemini ve Copilot gibi büyük dil modeli tabanlı sohbet botlarının, abdominoplasti öncesinde verilen cerrahi tavsiyeler açısından performansını değerlendirmiştir. Çalışmanın sonuçlarına göre, Claude, en güvenilir model olarak öne çıkarken, ChatGPT-3.5 ona yakın bir performans sergilemiştir. Buna karşılık, Copilot ve Gemini, diğer modellere kıyasla daha düşük güvenilirlik göstermiştir [140].

Duran ve çalışma arkadaşları tarafından yürütülen araştırmada, aralarında ChatGPT-4, Gemini ve Copilot gibi modellerin bulunduğu çeşitli büyük dil modellerinin estetik cerrahi prosedürlerine ilişkin sağladıkları tıbbi bilgilerin okunabilirlik, netlik ve doğruluk açısından değerlendirilmesi yapılmıştır. Elde edilen sonuçlar, ChatGPT-4'ün özellikle hasta odaklı bilgilendirici içerikler oluşturma açısından diğer modellere kıyasla daha başarılı olduğunu ortaya koymuştur [141].

Ponzo ve ekibinin yaptığı çalışmada, farklı klinik özelliklere sahip iki obezite vakasında yapay zeka tabanlı sohbet botlarının beslenme önerisi sunma becerileri incelenmiştir. Çalışmada, obez ancak başka hastalığı olmayan 35 yaşındaki bir erkek (Vaka 1) ile obezite yanında tip 2 diyabet, sarkopeni ve kronik böbrek hastalığı bulunan 65 yaşındaki bir kadın

(Vaka 2) için üç gün boyunca 10 farklı sohbet botundan beslenme tavsiyeleri istenmiştir. Yanıtlar, üç kayıtlı diyetisyen tarafından doğruluk, yeterlilik ve tekrarlanabilirlik açısından değerlendirilmiştir. Vaka 1'de ChatGPT 3.5, %67,2 ile doğruluk açısından en başarılı model olurken, Copilot %21,1 ile en düşük performansı göstermiştir. Yeterlilik yönünden ChatGPT 3.5 ve ChatGPT 4.0 en yüksek puanı (%87,3) elde etmiş; Gemini (%55,6) ve Copilot (%42,9) ise en düşük yeterliliği sergilemiştir. Tekrarlanabilirlik bakımından Chatsonic (%86,1) en yüksek, ChatGPT 4.0 (%50) ve ChatGPT 3.5 (%52,8) en düşük puanları almıştır. Vaka 2'de ise genel doğruluk oranları düşmüş ve hiçbir sohbet botu %50 doğruluk düzeyine ulaşamamıştır. Bu vakada yeterlilik açısından en iyi performans ChatGPT 4.0 ve Claude (%77,8) tarafından gösterilirken, Copilot %23,3 ile en zayıf sonucu vermiştir. Tekrarlanabilirlik açısından ChatGPT 4.0 ve Pi AI en düşük puanları kaydetmiştir [142].

Karahan ve ekibi tarafından gerçekleştirilen araştırmada, ChatGPT-4o ve Claude 3.5 gibi yapay zekâ tabanlı sohbet botlarının mamografi görüntülerini değerlendirme ve yorumlama performansları, özellikle BI-RADS sınıflandırması ve meme parenkim tiplerinin belirlenmesi açısından karşılaştırmalı olarak analiz edilmiştir. Ocak ile Temmuz 2024 tarihleri arasında toplanan toplam 53 mamografi görüntüsüne ait analizlerin sonucunda, her iki sohbet botunun da düşük düzeyde doğruluk ve güvenilirlik sergilediği ortaya konulmuştur. Çalışmada elde edilen bulgulara göre, BI-RADS sınıflamasında ChatGPT-4o'nun doğruluk oranı %18,87 ile %26,42 arasında değişen bir aralıkta gerçekleşirken, Claude 3.5 modeli için sabit olarak %18,7 olarak ölçülmüştür. BI-RADS kategorileri iyi huylu ve kötü huylu olarak yeniden sınıflandırıldığında ise ChatGPT-4o'nun doğruluk seviyesi %55 ila %57,5 arasında, Claude 3.5 içinse %47,5 olarak belirlenmiştir. Meme parenkim tipini doğru değerlendirme açısından ChatGPT-4o modeli %22,64 ila %30,19 arasında bir başarı gösterirken, Claude 3.5 modeli %26,42 doğruluk seviyesine ulaşmıştır. Sonuç olarak, bu çalışma, mevcut yapay zekâ tabanlı sohbet botlarının mamografi görüntülerinin değerlendirilmesinde yeterli düzeyde performans göstermediğini, bu nedenle daha kapsamlı veri setleri ve geliştirilmiş eğitim yöntemleri kullanılarak performanslarının artırılmasına ihtiyaç duyulduğunu vurgulamaktadır [113].

Günay ve arkadaşlarının yaptığı çalışmada ChatGPT-4, GPT-4o ve Gemini Advanced gibi büyük dil modellerinin elektrokardiyografi (EKG) sonuçlarını yorumlama becerileri analiz edilerek, tıbbi uygulamalardaki potansiyelleri değerlendirilmiştir. Araştırmada, ChatGPT-4o

modelinin genel EKG değerlendirmelerinde diğer iki modele kıyasla daha yüksek başarı gösterdiği saptanmıştır. Ayrıca, günlük klinik pratiğe yönelik EKG değerlendirmelerinde de ChatGPT-4o, Gemini Advanced modelinden daha iyi performans sergilemiştir. Modellerin verdiği yanıtların tutarlılığı açısından bakıldığında ise ChatGPT-4 ve Gemini Advanced modellerinin düşük seviyede bir uyum gösterdiği görülürken, ChatGPT-4o'nun ise orta düzeyde tutarlı yanıtlar sunduğu tespit edilmiştir. Elde edilen bu bulgular, büyük dil modellerinin EKG yorumlama sürecinde doğruluk ve tutarlılık açısından halen sınırlı olduklarını ve günümüzde klinik uzmanların yerini alacak düzeye henüz ulaşamadıklarını göstermektedir. Bu nedenle söz konusu modellerin klinik uygulamalarda rutin olarak kullanılabilmesi için daha kapsamlı veri setleriyle eğitilerek geliştirilmesi gerektiği sonucuna varılmıştır [143].

Hoppe ve arkadaşlarının yürüttüğü araştırmada, ChatGPT'nin GPT-3.5 ve GPT-4 sürümlerinin acil servise başvuran yetişkin hastaların tanısında gösterdiği performans, birinci basamak hasta değerlendirmesini yürüten asistan doktorların tanısal doğruluk oranlarıyla karşılaştırılmıştır. Araştırma kapsamında, iç hastalıkları kaynaklı sağlık sorunları nedeniyle acil servise başvuran yetişkin hastaların klinik bilgileri incelenmiş, yapay zeka modelleri ve asistan hekimlerin koyduğu tanılar, hastaların hastaneden taburcu edilirken konulan son tanılar ile karşılaştırılarak analiz edilmiştir. Araştırmaya dahil edilen hastaların çoğunluğunun ileri yaş grubunda olduğu ve kardiyovasküler sistem, endokrin sistem, gastrointestinal sistem ve enfeksiyon kaynaklı hastalıklar nedeniyle başvurdukları saptanmıştır. Yapılan değerlendirme sonucunda ChatGPT-4 modelinin, hem ChatGPT-3.5 sürümüne hem de acil servis asistan doktorlarına kıyasla tanısal doğruluk açısından daha başarılı olduğu tespit edilmiştir [144].

Mevcut çalışmada genel olarak dil modellerinin tanı başarısında Eylül 2024'ten Ocak 2025'e belirgin bir tutarlılık gözlenmiştir. Çoğu modelin performansı zaman içinde istatistiksel olarak anlamlı bir değişim göstermemiştir. Örneğin ChatGPT-3.5'in doğru yanıt oranı Eylül ayında Türkçe için %61,5 iken Ocak ayında %69,2 olarak hafifçe artsa da bu fark istatistiksel olarak anlamlı bulunmamıştır ($p>0,05$). Benzer şekilde, ChatGPT-4 ve Claude 2 modelleri de her iki dilde küçük oranlı iyileşmeler veya düşüşler sergilemekle birlikte tutarlı bir performans düzeyini korumuştur. Bu durum, gelişmiş dil modellerinin kısa vadede kararlı bir performans sunduğuna işaret etmektedir.

Bununla birlikte, Microsoft Copilot modelinde belirgin bir performans düşüşü saptanmıştır. Eylül 2024'te İngilizce senaryolarda %61,5 doğru yanıt oranına sahip olan Copilot, Ocak 2025'te aynı senaryolarda yalnızca %7,7 oranında doğru yanıt verebilmiştir (p=0,039). Türkçe senaryolarda da Copilot'un doğruluk oranı %38,5'ten %7,7'ye gerilemiş ancak örneklem boyutu nedeniyle istatistiksel anlamlılığa ulaşmamıştır. Bu anlamlı düşüş, Copilot'un model güncellemeleri veya yanıt biçimindeki değişiklikler nedeniyle önceki başarısını sürdüremediğini göstermektedir. Diğer bir deyişle, Microsoft Copilot'un performansındaki istikrarsızlık, dil modelinin zaman içinde tutarlı bir şekilde geliştirilmesinin önemini vurgulamaktadır.

Dikkate değer bir diğer değişim Perplexity AI modelinde görülmüştür. Perplexity AI, Eylül 2024'te Türkçe senaryolarda %15,4'lük doğru yanıt oranı ile en düşük performansı sergilerken Ocak 2025'te bu oranı %46,2'ye yükseltmiştir. Bu ~%30'luk artış istatistiksel anlamlılık eşiğini aşmasa da (p=0,125), modelin zaman içinde güncellenerek belirgin bir iyileşme gösterdiğine işaret etmektedir. Benzer şekilde, ChatGPT-4 (güncel sürüm) modelinin de Türkçe senaryolarda %61,5'ten %76,9'a, İngilizce'de %76,9'dan %92,3'e yükselen bir performansı vardır. Bu artışlar istatistiksel olarak anlamlı bulunmasa bile (%15-16'lık artışlar, p>0,05), gözlemlenen iyileşme eğilimi gelişmiş modellerin sürekli eğitim ve güncellemelere paralel olarak tanı başarısını artırabildiğini göstermektedir. Nitekim OpenAI, ChatGPT-4 modelinin selefi ChatGPT-3.5'e kıyasla "daha güvenilir, yaratıcı ve nüanslı talimatları daha iyi anlama" yeteneğine sahip olduğunu ve İngilizce dışındaki dillerde de geliştirilmiş performans sunduğunu bildirmiştir [145]. Literatürde de GPT-4 bazlı modellerin zamanla ve güncellemelerle birlikte tıbbi sorulara verdiği yanıtların doğruluk oranında artışlar görülebildiği belirtilmektedir [145]. Sonuç olarak, sekiz büyük dil modelinden yalnızca biri zaman içinde performansında istatistiksel olarak anlamlı bir değişim göstermiş (Copilot'ta düşüş), diğer modeller ise benzer başarı düzeylerini korumuştur. Bu bulgu, güncel büyük dil modellerinin kısa vadede göreceli bir kararlılıkla çalıştığını ancak bazı modellerin beklenmedik gerilemeler yaşayabileceğini ortaya koymaktadır. Kullanılan büyük dil modellerinin zaman içerisinde güncellenebilir olması nedeniyle, Eylül 2024 ve Ocak 2025 değerlendirmeleri sırasında bulut tabanlı modellerin (örneğin ChatGPT, Claude, Bing Copilot, Perplexity ve Gemini gibi) farklı sürümleri kullanılmış olabilir. Örneğin, OpenAI'nin ChatGPT-4 modelinin periyodik olarak güncellendiği bilinmektedir; dolayısıyla Ocak ayında test edilen sürümün Eylül

ayındakinden daha gelişmiş olması mümkündür. Benzer durum, Perplexity AI veya Aralık 2024'te ilk kez sunulan Gemini modeli için de geçerli olabilir. Bu nedenle bu çalışmadaki zaman temelli performans farklılıkları yalnızca modellerin kendi öğrenmelerinden değil, aynı zamanda güncellemelerden kaynaklanmış olabilir. Modellerin altyapısındaki güncellemeler veya bilgi tabanlarındaki değişiklikler performansı etkileyebileceği için, bu tür karşılaştırmalı zaman serisi değerlendirmelerinin devam etmesi önem taşımaktadır.

Dil modellerinin Türkçe ve İngilizce senaryolardaki performansları genellikle benzer bulunmuştur, bu da gelişmiş modellerin dil bariyerini büyük ölçüde aşabildiğine işaret etmektedir. Eylül 2024 değerlendirmesinde hiçbir modelin Türkçe ve İngilizce performansı arasında istatistiksel olarak anlamlı bir fark saptanmamıştır (McNemar $p>0,05$). Örneğin ChatGPT-3.5 modeli her iki dilde de %61,5 doğru yanıt oranı ile aynı başarıyı göstermiştir. Benzer şekilde Anthropic Claude 2 modeli de Eylül ayında Türkçe %76,9 ve İngilizce %84,6 oranlarıyla oldukça yüksek ve birbirine yakın doğruluk sergilemiştir ($p=1,00$). Bu sonuçlar, birçok büyük dil modelinin eğitiminde çok dilli verilerin kullanılması sayesinde her iki dilde de tutarlı bir performans yakalayabildiğini göstermektedir.

Ocak 2025 değerlendirmesinde de genel tablo benzerdir: sekiz modelden yalnızca biri iki dil arasında belirgin bir performans farkı göstermiştir. Google'ın Gemini modeli, İngilizce senaryolarda %53,8 doğru yanıt oranına ulaşırken Türkçe senaryolarda hiçbir doğru tanı koyamamıştır (%0) ve bu fark istatistiksel olarak anlamlıdır ($p=0,016$). Gemini dışındaki modeller için Türkçe ve İngilizce doğru yanıt oranları arasında anlamlı bir farklılık saptanmamıştır ($p>0,05$). Örneğin, ChatGPT-4 (Ocak sürümü) modeli Türkçe sorularda %76,9, İngilizce sorularda %92,3 oranında doğru yanıt vermiş ve aradaki fark anlamlı bulunmamıştır ($p=0,500$). Aynı şekilde Claude 2 modeli Ocak ayında her iki dilde de %69,2 oranında doğru tanı koyarak tam bir dil tutarlılığı göstermiştir. Bu bulgular, özellikle en güçlü modellerin dil bariyerini büyük ölçüde aştığını, soruları Türkçe'de de İngilizce'de olduğu kadar doğru yanıtlayabildiğini ortaya koymaktadır.

Literatürde de GPT-4 tabanlı modeller gibi ileri modellerin çok dilli tıbbi metinlerde başarılı olduğuna dair bulgular mevcuttur. Örneğin, ChatGPT-4'ün tıbbi notları İngilizce, İspanyolca ve İtalyanca dillerinde analiz ederken tümünde yüksek doğrulukla soruları yanıtlayabildiği gösterilmiştir [146]. Benzer şekilde, Japonca tıp lisans sınavında ChatGPT-4'ün yüksek başarı gösterip ChatGPT-3.5'i geride bıraktığı ve böylece İngilizce olmayan bir dilde de

güvenilir klinik akıl yürütebildiği rapor edilmiştir [145]. Mevcut çalışmada da GPT-4 tabanlı modellerin (ChatGPT-4 ve 4o) Türkçe tanı senaryolarında en az İngilizce kadar başarılı olması, bu çok dillilik potansiyelini doğrular niteliktedir. ChatGPT-4o modeli, Ocak ayında Türkçe sorularda %92,3 ile bu çalışmadaki en yüksek doğru tanı oranına ulaşmıştır ve aynı model İngilizce'de %84,6 oran yakalamıştır, her iki dilde de yüksek bir performans söz konusudur. Bu durum, gelişmiş dil modellerinin eğitim verilerindeki dil çeşitliliği sayesinde tanı kriterlerini dil fark etmeksizin uygulayabildiğini göstermektedir.

Öte yandan, daha kısıtlı dil verisiyle eğitilmiş modellerde dil bariyeri etkisi belirgin olmuştur. Google Gemini modelinin Türkçe sorulardaki başarısızlığı, bu modelin henüz Türkçe dilinde yeterli eğitimi almadığını veya Türkçe klinik terimleri anlamlandırmada sorun yaşadığını düşündürmektedir. Benzer şekilde, Eylül ayında Perplexity AI ve Microsoft Copilot gibi modeller İngilizce'de Türkçe'den daha yüksek doğruluk oranlarına sahiptir, ancak aradaki farklar istatistiksel olarak anlamlı olmasa da bu eğilim İngilizce dilinde eğitim verilerinin daha zengin olmasından kaynaklanıyor olabilir. Nitekim bazı çalışmalar, dil modellerinin sorular İngilizce sunulduğunda orijinal dile kıyasla daha doğru yanıtlar verebildiğini, çeviri kalitesinin model performansını etkilediğini ortaya koymaktadır [147].

Bulgularımız, bazı dil modellerinin temporomandibular eklem düzensizlikleri tanı süreçlerinde klinik karar desteği olarak kullanılabilecek kadar yüksek doğruluk sergilediğini, bazılarının ise mevcut halleriyle sınırlı bir değere sahip olduğunu göstermektedir. Özellikle OpenAI tarafından geliştirilen GPT-4 tabanlı modeller (ChatGPT-4 ve ChatGPT-4o) ile Anthropic'in Claude 2 modeli, her iki dilde de yüksek tanı doğruluk oranlarına ulaşmıştır. Ocak ayı verilerine göre ChatGPT-4 ve ChatGPT-4o modelleri tanı senaryolarında %85–92 gibi dikkat çekici doğruluk oranları sunmuştur. Bu düzey, birçok hastada doğru tanıya varılabileceğini ve model cevaplarının klinisyene gerçekçi bir ikinci görüş sağlayabileceğini göstermektedir. Nitekim literatürde ChatGPT-4'ün tıbbi sınavlarda %80'i aşan başarı oranları yakalayabildiği ve zorlu klinik vakalarda dahi güçlü bir akıl yürütme sergilediği rapor edilmektedir [145]. Özellikle diş hekimliği ve çene eklemi bozuklukları konusunda yayımlanan bazı çalışmalar, büyük dil modellerinin hasta eğitimi ve klinik bilgiye erişim konularında faydalı olabileceğini göstermiştir. Örneğin, yakın zamanda yapılan bir çalışmada ChatGPT-4'e temporomandibular eklem düzensizlikleri ile ilgili sorular sorulmuş ve yanıtların güvenilirlik ve kullanılabilirlik açısından orta-yüksek kalitede olduğu

bildirilmiştir [104]. Bu çalışmada da ChatGPT-4'ün senaryo bazlı tanı doğruluğunun %75–85 aralığında olması, bu modelin temporomandibular düzensizlikler konusunda önemli ölçüde doğru ve faydalı cevaplar üretebildiğini göstermektedir ve bu durum söz konusu literatür bulgusuyla tutarlıdır.

Model bazında literatür karşılaştırması yapıldığında, Anthropic Claude 2 ve diğer alternatif büyük modellerle ilgili yayınların daha sınırlı olduğu görülmektedir. ChatGPT/GPT-4 kadar olmasa da, Claude 2'nin de genel dil ve mantık yeteneklerinin yüksek olduğu bilinmektedir [148]. Literatürde doğrudan Claude 2'nin klinik performansını ele alan kapsamlı çalışma bulunamamıştır ancak bizim bulgumuz, Claude 2'nin temporomandibular eklem düzensizlikleri tanısında ChatGPT-4'e yakın bir başarı sergilediği yönündedir (özellikle Eylül ayında %84,6 ile en yüksek doğruluklardan birine ulaşmıştır). Bu, Claude 2'nin de klinik destek için değerlendirilebilecek bir model olduğu yönündeki düşünceyi güçlendirmektedir. Öte yandan, Microsoft Copilot ve Perplexity AI gibi modellerin başarısızlığı literatürde de öngörülebilir bir sonuçtur [139]. Mevcut çalışmada da GPT-4 türevi modellerin ve Claude 2'nin temporomandibular eklem düzensizlikleri tanısında benzer bir başarı ortaya koyması, bu modellerin klinik pratikte destekleyici bir araç olarak potansiyelini vurgulamaktadır. Özellikle tanı sürecinde hekimlerin zorlandığı veya birden fazla olasılığın değerlendirilmesi gereken durumlarda, ChatGPT-4 gibi modeller ek bir danışman olarak kullanılabilir. Claude 2 modeli de %69–85 aralığında doğru tanı oranlarıyla güvenilirliğini göstermiştir. Bu modeller, tutarlı performansları sayesinde hekimin gözardı etmiş olabileceği bir tanıyı önerme veya mevcut tanıyı doğrulama noktasında faydalı olabilir fakat dil modellerinin hiçbiri mutlak güvenilirlik seviyesinde değildir. En iyi modeller dahi %100 başarıya ulaşamamıştır, Dolayısıyla, bu araçlar yardımcı karar destek sistemleri olarak düşünülmeli, nihai kararı verecek olanın her zaman klinisyen olduğu akılda tutulmalıdır. Klinik kullanımda, model çıktılarının doğruluğunu ve önerilerini hekimlerin değerlendirmesi gerekmektedir. Özellikle temporomandibular düzensizlikler gibi karmaşık ve birden fazla alt tipi olan bozukluklarda, modelin verdiği cevabın doğru alt tipi mi işaret ettiği dikkatle kontrol edilmelidir. Ayrıca dil modellerinin olası "halüsinasyon" sorunu yani gerçekte olmayan bilgiler üretme riski klinik bağlamda önemli bir endişe kaynağıdır [104]. Bu nedenle, model önerileri kesinlikle doğrulanmadan tek başına tedavi veya müdahale kararı için kullanılmamalıdır.

Modeller arasında klinik kullanım potansiyeli bakımından belirgin farklılıklar mevcuttur. ChatGPT-4/4o ve Claude 2, klinik destek için en güçlü adaylar olarak öne çıkarken, ChatGPT-3.5 ve Docus AI gibi modeller nispeten daha düşük doğruluk oranı sunmaktadır. Özellikle ChatGPT-3.5, GPT-4 tabanlı modeller kadar başarılı olmasa da dikkat çekmektedir. Bu model, ücretsiz veya daha yaygın erişilebilir olduğundan, hekim olmayan sağlık personeli veya öğrenciler tarafından bilgi almak amacıyla değerlendirilebilir. Docus AI modeli de benzer bir başarı düzeyine sahiptir ve sağlık alanına özel geliştirildiği için özellikle temporomandibular düzensizlikler tanı kriterlerine aşinalığı sayesinde makul bir performans sergilemiştir (Türkçe %69,2, İngilizce %84,6). Öte yandan, Microsoft Copilot, Gemini ve Perplexity AI gibi modellerin mevcut halleriyle doğrudan klinik kullanıma uygun olmadığı anlaşılmaktadır. Microsoft Copilot'un özellikle Ocak 2025 değerlendirmesinde neredeyse işlevsiz kalacak düzeyde (%7,7 doğruluk) performans sergilemesi, bu platformun temporomandibular düzensizlik tanısı gibi özel bir alanda yeterli bilgi tabanına sahip olmadığını veya tutarlı sonuç üretemediğini göstermiştir. Benzer şekilde, Gemini modeli Türkçe dilinde hiç doğru yanıt verememiş, İngilizce'de ise %53,8 ile düşük bir başarı yakalayabilmiştir. Bu çalışmadaki düşük performanslı modeller, yanlış yönlendirebileceklerinden dolayı TMD ile ilgili klinik karar desteğinde kullanılmaması daha uygundur.

Sonuç olarak, bazı ileri büyük dil modelleri TMD tanısında yardımcı araçlar olarak umut vaat etmektedir. Bu modeller, doğru kullanıldığında tanı sürecini hızlandırıp destekleyebilir ve klinisyenler için ikinci bir fikir sağlayabilir. Ancak mevcut halleriyle hiçbir modelin bağımsız bir tanı aracı olarak güvenilirliği kanıtlanmış değildir. Bu nedenle, modellerin çıktıları rehberlik amacıyla alınmalı, klinik bağlam ve hastanın tüm bulguları göz önüne alınarak son karar klinisyene bırakılmalıdır.

Elde ettiğimiz bulgular, literatürde büyük dil modellerinin tıbbi alandaki performansına ilişkin mevcut bilgilerle genel hatlarıyla uyumludur [114], [116], [139], [141], [144]. İleri seviye daha güncel büyük dil modellerinin önceki nesil modellere göre belirgin şekilde üstün performans gösterdiği bu çalışmada da doğrulanmıştır. Modellerin çok dilli performansına dair literatür bulguları da bu çalışmayla uyumludur, özellikle GPT-4 tabanlı modellerin farklı dillerdeki tıbbi metinleri anlama ve soruları cevaplama becerisi, diğer çalışmalar tarafından da rapor edilmiştir [134], [135]. Örneğin, yakın tarihli bir model değerlendirme çalışmasında

ChatGPT-4'ün İngilizce, İtalyanca ve İspanyolca olarak sunulan klinik notları sorunsuz analiz edip, hekimler tarafından değerlendirilen sorulara %79–88 oranında doğru yanıtlar ürettiği bulunmuştur [146]. Bu bulgu, ChatGPT-4'ün çok dilli ortamlarda dahi etkili çalışabildiğini ve dil bariyerinin getirdiği engelleri büyük ölçüde aşabildiğini vurgulamaktadır. Mevcut çalışmada da GPT-4 tabanlı iki modelin (ChatGPT-4 ve 4o) Türkçe ve İngilizce başarılarının yakın olması literatürdeki bu eğilimi destekler niteliktedir. Öte yandan, Google Gemini gibi bir modelin Türkçe dilinde düşük performans sergilemesinin sebebi. Türkçe gibi İngilizce'ye nazaran sınırlı kullanım alanı olan dillerin eğitim verisi açısından da sınırlı kalması olabilir.

Bu çalışmanın bulguları göz önüne alındığında bu çalışma literatürdeki büyük dil modellerinin tıbbi uygulamalarına dair mevcut bilgilerle büyük ölçüde uyumlu olduğu görülmüştür [116], [127], [130], [131], [132], [134], [138], [139], [140], [141], [144]. Bu uyum, mevcut çalışmanın bulgularının güvenilirliğini pekiştirdiği gibi, literatürde dile getirilen kaygı ve umutların temporomandibular eklem düzensizlikleri için de geçerli olduğunu göstermektedir [113], [114], [117], [143], [149]. Elbette ki her çalışma gibi bazı farklılıklar ve yeni gözlemler de mevcuttur; örneğin Copilot'un performansındaki dramatik düşüş veya Gemini'nin Türkçe dilinde TMD teşhisindeki başarısızlığı gibi noktalar literatürde pek vurgulanmamış özgün katkılar olarak değerlendirilebilir. Bu nedenle ileride yapılacak çalışmalar, literatürle karşılaştırmayı zenginleştirmek adına farklı model ve dil kombinasyonlarını da inceleyerek büyük dil modellerinin sağlık alanındaki rolünü daha kapsamlı biçimde ortaya koyacaktır.

Bu çalışmanın birtakım limitasyonları bulunmaktadır. Bunlardan biri kullanılan soru şablonlarıdır. Bu şablonlar TMD/TK'ya dayanılarak hazırlanmış olmasına rağmen, temporomandibular eklem düzensizliklerinin tüm klinik çeşitliliğini yeterli düzeyde temsil etmek için sınırlı kalabilir. Ayrıca büyük dil modellerine yöneltilen 13 adet olgu temelli soru şablonu, örneklem sayısının sınırlı kalmasına neden olmuştur; bu durum, modeller arasında gözlemlenen belirgin performans farklılıklarına rağmen istatistiksel olarak anlamlı sonuçlara ulaşılamamasına yol açmıştır. Örneğin ChatGPT-4o'nun Türkçe performansının %69,2'den %92,3'e yükselmesi gibi pratik açıdan belirgin performans artışları, sınırlı örneklem büyüklüğü nedeniyle istatistiksel anlamlılık göstermemiştir. Benzer şekilde, modellerin dil performanslarında İngilizce lehine gözlenen küçük farklar da istatistiksel olarak anlamlı

bulunmamıştır. Bu durum, Perplexity AI gibi modellerde zamana bağlı gerçek performans iyileşmelerinin, sınırlı vaka sayısı sebebiyle tespit edilemediğini göstermektedir. Gelecekte daha geniş örneklemle yapılacak çalışmaların bu tür performans farklılıklarını istatistiksel olarak doğrulayabileceği düşünülmektedir.

Çalışmada kullanılan değerlendirme kriterleri de bir diğer limitasyon olarak sayılabilir. Örneğin, bir model tarafından birden fazla olası tanı belirtildiğinde, yanlış yanıt olarak değerlendirilmesi eksiklik olarak düşünülebilir çünkü gerçek klinik koşullarda modellerin birden fazla olası tanıyı sunması, teşhis sürecinde kısmi bir fayda sağlayabilmektedir ancak değerlendirme kriterlerinin yoruma kapalı olarak doğrudan TMD/TK tanı karar ağacına göre uygun teşhis koyacak biçimde düzenlenmesi ile objektif ve net veriler elde etmek amaçlanmıştır. Ayrıca model sürümlerinin değişkenliği bu çalışmanın sınırlılıklarından biri olarak değerlendirilebilir. Bu değerlendirmeler, 2024 Eylül ve 2025 Ocak aylarına ait veriler doğrultusunda gerçekleştirilmiştir; ancak büyük dil modellerinin öğrenme kapasiteleri ve zamanla gerçekleşen sistem güncellemeleri nedeniyle, elde edilen sonuçlar ilerleyen dönemlerde farklılık gösterebilir.

Bu çalışmada büyük dil modellerinin performansı, TMD/TK'ya göre yazılı hasta senaryoları üzerinden tanı koyma yetenekleriyle değerlendirilmiştir. Ancak gerçek klinik koşullarda tanı süreci, sözlü iletişim, fiziksel muayene bulguları ve radyolojik görüntüleme gibi çoklu veri türleri kullanılarak yürütülür. Araştırmamızdaki tasarım ise temporomandibular düzensizlikler için tanıyı yalnızca sözel vaka bilgilerinden oluşturmayı kapsamaktadır. Bu yöntem pratik bir yaklaşım olsa da, manyetik rezonans (MR) görüntüleri veya çene eklemi ses kayıtları gibi önemli klinik unsurların değerlendirme dışında kalmasına neden olmuştur. Bu açıdan değerlendirildiğinde, model performansları sınırlı ve tek yönlü bir veri seti üzerinden ölçülmüştür. Ayrıca, mevcut çalışmadaki modellerden yalnızca nihai tanıyı vermeleri istenmiştir; önerilen tedavi yaklaşımları ya da ilave tetkik önerileri değerlendirme kapsamında değildir. Bu durum, belirli bir konuda model performansının net bir şekilde ortaya konulmasını sağlarken, klinik uygulamanın tüm boyutlarını kapsamamaktadır. Örneğin, bir modelin tanı koymadaki başarısı ile tedavi önerilerindeki başarısı aynı olmayabilir ancak mevcut çalışmada sadece tanı doğruluğu değerlendirilmiştir. Gelecekte gerçekleştirilecek çalışmalarda, büyük dil modellerinin tanı koyma performanslarının yanı

sıra yönetim ve tedavi planlaması konusunda da değerlendirilmesi, modellerin klinik pratikteki gerçek potansiyelini daha bütüncül bir şekilde ortaya koyacaktır.

Bu sınırlılıklar ışığında, mevcut çalışmanın sonuçlarının dikkatli ve bağlamı içinde yorumlanması gerekmektedir. Bulgularımız, sekiz büyük dil modelinin belirli koşullar altındaki performansını ortaya koyarken, bu koşulların dışında farklı sonuçlar alınabileceğini unutmamak gerekir. Gelecekte daha kapsamlı, farklı dil ve senaryo çeşitlerini içeren, daha büyük örneklemli çalışmalar yapılarak burada sunulan bulguların doğrulanması ve genişletilmesi uygun olacaktır. Bu sayede, büyük dil modellerinin sağlık alanındaki rolü ve güvenilirliği konusunda daha kesin ve genellenebilir yargılara varmak mümkün olacaktır.

6. SONUÇ

Bu çalışmada, yapay zekâ tabanlı büyük dil modellerinin temporomandibuler eklem düzensizlikleri tanısındaki performansları; Türkçe ve İngilizce olarak, iki farklı zamanda (Eylül 2024 ve Ocak 2025) değerlendirilmiştir. Bu çalışmanın limitasyonları dahilinde erişilen sonuçlar ve öneriler şunlardır:

a. Modeller arasında istatistiksel olarak anlamlı farklar saptanmış, dil ve zaman gibi faktörlerin bazı büyük dil modellerinin performansını değiştirdiği görülmüştür.

b. GPT-4 tabanlı modeller (ChatGPT-4 ve ChatGPT-4o) ile Anthropic Claude 2, her iki değerlendirme döneminde ve her iki dilde en yüksek tanı doğruluk oranlarını göstermiştir.

c. Microsoft Copilot, Gemini ve Perplexity AI gibi bazı modeller, birçok senaryoda düşük doğruluk oranları göstermiş ve gelişmiş modellerden istatistiksel olarak anlamlı biçimde geride kalmıştır. Özellikle Gemini modeli, Ocak 2025 döneminde Türkçe sorularda %0 doğruluk göstermiştir.

d. Genel olarak modellerin performansları zamana bağlı olarak değişmemiştir. Ancak Microsoft Copilot modelinde, Ocak 2025 döneminde İngilizce değerlendirmelerde anlamlı bir performans düşüşü saptanmıştır.

e. Gelişmiş modellerin çoğu, Türkçe ve İngilizce sorulara benzer doğrulukla yanıt vermiştir. Dil farklılığına bağlı anlamlı performans değişikliği yalnızca Gemini modelinde gözlenmiştir. Bu durum, istisnalar haricinde modellerin çok dilli eğitim yapılarının başarısını ortaya koymaktadır.

f. ChatGPT-4 tabanlı modeller ve Claude 2, TMD tanısında yüksek doğrulukları ile klinik karar desteği aracı olma potansiyeline sahiptir. Ancak tüm modellerin aynı başarıyı göstermemesi, yapay zekâ sistemlerinin klinik ortamlarda dikkatle seçilmesi ve değerlendirilmesi gerektiğini göstermektedir.

g. Büyük dil modelleri, doğru kullanıldıklarında ve sürekli geliştirildiklerinde, çok dilli tıbbi asistanlar olarak diş hekimliği pratiğinde değerli bir destek aracı olabilir. Ancak mevcut durumda, yapay zekâ destekli sistemlerin klinik kararlarda yardımcı olarak

değerlendirilmesi; insan denetimi ve uzman görüşünün vazgeçilmez olduğu unutulmamalıdır.

h. Bu çalışmada kullanılan soru sayısının sınırlı olması ve modellerin zamanla güncellenebilir yapıları, değerlendirme sonuçlarını etkileyebilecek faktörler arasındadır. İlerleyen dönemlerde daha geniş örneklemlerle ve uzun zaman aralıklarını kapsayan çalışmalar yapılması önerilmektedir.

REFERANSLAR

[1] S. Wadhwa ve S. Kapila, "TMJ Disorders: Future Innovations in Diagnostics and Therapeutics", *Journal of Dental Education*, c. 72, sy 8, ss. 930-947, Ağu. 2008, doi: 10.1002/j.0022-0337.2008.72.8.tb04569.x.

[2] A. C. Adam Erden ve D. Karakiş, "Temporomandibular eklem düzensizlikleri teşhisinde kullanılan görüntüleme yöntemleri", *Mersin Üniversitesi Sağlık Bilimleri Dergisi*, c. 15, sy 1, ss. 110-127, Nis. 2022, doi: 10.26559/mersinsbd.903504.

[3] M. Yayman ve Prof. Dr. S. Akman, "Temporomandibular Eklem Seslerini Değerlendirme Yöntemleri", *Selcuk Dental Journal*, c. 10, sy 3, ss. 600-604, Ara. 2023, doi: 10.15311/selcukdentj.1235656.

[4] Ö. Miloğlu, "Temporomandibular Eklem Disfonksiyonu Olan Hastalardaki Kondiller Kemik Değişikliklerinin İnternal Düzensizlik (Disk Deplasmani) İle Olan ilişkisinin İncelenmesi," PhD dissertation, Diş. Hek. Fak. Oral Diagnoz ve Rad. ABD., Atatürk Üni., Erzurum, Türkiye, 2009.

[5] K. Baba, Y. Tsukiyama, M. Yamazaki, ve G. T. Clark, "A review of temporomandibular disorder diagnostic techniques", *The Journal of Prosthetic Dentistry*, c. 86, sy 2, ss. 184-194, Ağu. 2001, doi: 10.1067/mpr.2001.116231.

[6] S. Levitt, "Predictive value of the TMJ scale in detecting clinically significant symptoms of temporomandibular disorders", *Journal of craniomandibular disorders : facial & oral pain*, c. 4, ss. 177-85, Şub. 1990.

[7] N. Demirkol, M. Demirkol, ve A. Üşümez, "The use of low-level laser therapy in temporomandibular joint disorders: Temporomandibular eklem rahatsızlıklarında düşük doz lazer tedavisinin kullanımı", *European Journal of Therapeutics*, c. 21, sy 3, Art. sy 3, 2015, doi: 10.5455/GMJ-30-165547.

[8] R. L. Gauer ve M. J. Semidey, "Diagnosis and treatment of temporomandibular disorders", *Am Fam Physician*, c. 91, sy 6, ss. 378-386, Mar. 2015.

[9] Doğa Hospital. "Radyolojide Yapay Zeka Kullanımı: Geleceğin Teşhis Yöntemleri". Erişim: 05 Ocak 2025. [Çevrimiçi]. Erişim adresi: https://dogahospital.com/radyolojide-yapay-zeka-kullanimi/

[10] M. Binbay. (2024). “Yapay Zeka ile Kanserlerin Erken Teşhisi”. Erişim: 05 Ocak 2025. [Çevrimiçi]. Erişim adresi: https://muratbinbay.com/yapay-zeka-ile-kanserlerin-erken-teshisi

[11] Yesil Science. “Sağlıkta Yapay Zeka Uygulaması”. Erişim: 05 Ocak 2025. [Çevrimiçi]. Erişim adresi: https://yesilscience.com/tr/yapay-zeka-ve-saglik/

[12] M. K. Yekedüz ve F. T. Eminoğlu, “Nadir Hastalıklarda Yapay Zekâ Uygulamaları”, *Türkiye Klinikleri Eczacılık Bilimleri - Özel Konular*, c. 4, sy 2, ss. 33-38, 2024.

[13] E. Karal ve M. Turan, “Hekime Tanı Koymada Yardımcı, Yapay Zekâ Destekli Hastalık Tespit Uzmanı”, *European Journal of Science and Technology*, Haz. 2021, doi: 10.31590/ejosat.945518.

[14] X. Yang *vd.*, “Application of large language models in disease diagnosis and treatment”, *Chinese Medical Journal (Engl)*, c. 138, sy 2, ss. 130-142, Oca. 2025, doi: 10.1097/CM9.0000000000003456.

[15] S. Palla, “Chapter 6 - Anatomy and Pathophysiology of the Temporomandibular Joint”, içinde *Functional Occlusion in Restorative Dentistry and Prosthodontics*, I. Klineberg ve S. E. Eckert, Ed., Mosby, 2016, ss. 67-85. doi: 10.1016/B978-0-7234-3809-0.00006-1.

[16] K. Brett, C. Wells, A. Sinclair, H. Tenenbaum, B. Freeman, ve C. Spry, “Anatomical Structures”, içinde *Interventions for Temporomandibular Joint Disorder: An Overview of Systematic Reviews [Internet]*, Canadian Agency for Drugs and Technologies in Health, 2018. Erişim: 29 Kasım 2023. [Çevrimiçi]. Erişim adresi: https://www.ncbi.nlm.nih.gov/books/NBK544534/

[17] P. Gosling, “DORLAND’S ILLUSTRATED MEDICAL DICTIONARY”, *Australas Chiropr Osteopathy*, c. 11, sy 2, s. 65, Tem. 2003.

[18] X. Alomar *vd.*, “Anatomy of the Temporomandibular Joint”, *Seminars in Ultrasound, CT and MRI*, c. 28, sy 3, ss. 170-183, Haz. 2007, doi: 10.1053/j.sult.2007.02.002.

[19] F. Koç Direk ve S. Canbay Durmaz, “Temporomandibular Eklem ve Klinik Anatomisi”, içinde *Anatomiye Klinik ve Morfolojik Yaklaşımlar*, F. Koç Direk, Ed., Özgür Yayınları, 2024. doi: 10.58830/ozgur.pub626.c2696.

[20] D. Durna ve A. B. Yilmaz, “Temporomandibular Eklem Disfonksiyonlu Bireylerde Kondile Ait Kemik Değişikliklerinin Dental Volumetrik Tomografi İle

Değerlendirilmesi", *Atatürk Üniversitesi Diş Hekimliği Fakültesi Dergisi*, ss. 1-1, Eki. 2020, doi: 10.17567/ataunidfd.757797.

[21] "The Glossary of Prosthodontic Terms: Ninth Edition", *The Journal of Prosthetic Dentistry*, c. 117, sy 5S, ss. e1-e105, May. 2017, doi: 10.1016/j.prosdent.2016.12.001.

[22] M.A. Ruffer. (1913). "Studies in the palaeopathology of Egypt". Open Library. Erişim: 03 Aralık 2023. [Çevrimiçi]. Erişim adresi: https://openlibrary.org/books/OL6638860M/Studies_in_the_palaeopathology_of_Egypt

[23] F. Adams. (1886). "The genuine works of Hippocrates". Internet Archive. Erişim: 03 Aralık 2023. [Çevrimiçi]. Erişim adresi: https://archive.org/details/genuineworkship02hippgoog

[24] C. McNeill, "History and evolution of TMD concepts", *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, c. 83, sy 1, ss. 51-60, Oca. 1997, doi: 10.1016/S1079-2104(97)90091-3.

[25] D. J. Goodfriend, "Abnormalities of the Mandibular Articulation**Read at the Seventy-Fifth Annual Session of the American Dental Association in conjunction with the Chicago Centennial Dental Congress, Aug. 9, 1933.", *The Journal of the American Dental Association (1922)*, c. 21, sy 2, ss. 204-218, Şub. 1934, doi: 10.14219/jada.archive.1934.0035.

[26] "Ramfjord, S.P. and Ash, M.M. (1983) Occlusion. 3rd Edition, WB Saunders, Philadelphia. - References - Scientific Research Publishing". Erişim: 03 Aralık 2023. [Çevrimiçi]. Erişim adresi: https://www.scirp.org/(S(lz5mqp453edsnp55rrgjct55.))/reference/referencespapers.aspx?referenceid=1951090

[27] N. A. Shore, "Occlusal equilibration and temporomandibular joint dysfunction", *(No Title)*, Erişim: 03 Aralık 2023. [Çevrimiçi]. Erişim adresi: https://cir.nii.ac.jp/crid/1130000796325295744

[28] W. C. Schulte ve R. R. Rhyne, "Synovial chondromatosis of temporomandibular joint. Report of a case", *Journal of Oral Surgery, Oral Medicine, Oral Pathology, and Oral Radiology*, c. 28, sy 6, ss. 906-913, Ara. 1969, doi: 10.1016/0030-4220(69)90347-8.

[29] R. H. Griffiths, "Report of the President's Conference on the Examination, Diagnosis, and Management of Temporomandibular Disorders", *The Journal of the American*

Dental Association, c. 106, sy 1, ss. 75-77, Oca. 1983, doi: 10.14219/jada.archive.1983.0020.

[30] E. Schiffman *vd.*, "Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) for Clinical and Research Applications: Recommendations of the International RDC/TMD Consortium Network* and Orofacial Pain Special Interest Group†", *Journal of Oral Facial Pain Headache*, c. 28, sy 1, ss. 6-27, Oca. 2014, doi: 10.11607/jop.1151.

[31] C. McNeill, "Management of temporomandibular disorders: concepts and controversies", *The Journal of Prosthetic Dentistry*, c. 77, sy 5, ss. 510-522, May. 1997, doi: 10.1016/s0022-3913(97)70145-8.

[32] T. I. Suvinen, P. C. Reade, K. R. Hanes, M. Könönen, ve P. Kemppainen, "Temporomandibular disorder subtypes according to self-reported physical and psychosocial variables in female patients: a re-evaluation", *Journal of Oral Rehabilitation*, c. 32, sy 3, ss. 166-173, Mar. 2005, doi: 10.1111/j.1365-2842.2004.01432.x.

[33] C. McNeill, "Evidence-based TMD guidelines", *Journal of Orofacial Pain*, c. 11, sy 2, s. 93, 1997.

[34] K. Oral, B. B. Küçük, B. Ebeoğlu, ve S. DiNçer, "Etiology of temporomandibular disorder pain", 2009.

[35] C. S. Greene, "Etiology of temporomandibular disorders", *Seminars in Orthodontics*, c. 1, sy 4, ss. 222-228, Ara. 1995, doi: 10.1016/S1073-8746(95)80053-0.

[36] J. P. Okeson, "Temporomandibular Disorders: Etiology and Classification", içinde *TMD and Orthodontics*, S. Kandasamy, C. S. Greene, D. J. Rinchuse, ve J. W. Stockstill, Ed., Cham: Springer International Publishing, 2015, ss. 19-36. doi: 10.1007/978-3-319-19782-1_2.

[37] S. F. Dworkin *vd.*, "Epidemiology of signs and symptoms in temporomandibular disorders: clinical signs in cases and controls", *The Journal of the American Dental Association*, c. 120, sy 3, ss. 273-281, Mar. 1990, doi: 10.14219/jada.archive.1990.0043.

[38] R. J. De Kanter *vd.*, "Prevalence in the Dutch adult population and a meta-analysis of signs and symptoms of temporomandibular disorder", *Journal of Dental Research*, c. 72, sy 11, ss. 1509-1518, Kas. 1993, doi: 10.1177/00220345930720110901.

[39] R. Eraslan ve K. Kiliç, "Temporomandibular eklem iç düzensizliğinin cinsiyet, yaş, eğitim durumu, iş durumu ve medeni durum ile ilişkisinin incelenmesi", *Selcuk Dental Journal*, c. 7, sy 2, Art. sy 2, Ağu. 2020, doi: 10.15311/selcukdentj.483443.

[40] S. F. Dworkin ve L. LeResche, "Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique", *Journal of craniomandibular disorders*, c. 6, sy 4, ss. 301-355, 1992.

[41] J.Okeson. (1996). "Orofacial pain : guidelines for assessment, diagnosis, and managemen". Semantic Scholar. Erişim: 05 Aralık 2023. [Çevrimiçi]. Erişim adresi: https://www.semanticscholar.org/paper/Orofacial-pain-%3A-guidelines-for-assessment%2C-and-Okeson/98275d5de995c7a76b396bf3c6f7678f9c84fb1d

[42] O. Güneş, "Temporomandibular Eklem İnternal Bozukluklarinda, Minimal İnvaziv Tedavinin Etkileri Üzerine Retrospektif Bir Araştırma," PhD. Dissertation, Diş. Hek. Fak. Ağız, Diş ve Çene Cerrahisi ABD., Ankara Üni.,Ankara,Türkiye, 2015.

[43] İ. Albayrak Gezer, "Temporomandibular eklem rahatsızlıklarının sınıflandırılması, tanı ve tedavisi", *Genel Tıp Derhisi*, c. 26, sy 1, ss. 34-34, Nis. 2016, doi: 10.15321/GenelTipDer.2016118200.

[44] C. Dağ, N. Özalp, ve M. Dağ, "Temporomandibular Düzensizlikler: Tanı ve Tedavi", 2011.

[45] M. Rantala, "Temporomandibular disorders and related psychosocial factors in non-patients : A survey and a clinical follow-up study based on the RDC/TMD".

[46] B. P. Singh *vd.*, "Occlusal interventions for managing temporomandibular disorders", *Cochrane Database of Systematic Reviews*, Kas. 2017, doi: 10.1002/14651858.CD012850.

[47] R. Ohrbach, "Diagnostic Criteria for Temporomandibular Disorders: Assessment Instruments". 2014.

[48] S. Sacco, "Pain comorbidities – understanding and treating the complex patient: IASP Press, 2012", *The Journal of Headache and Pain*, c. 14, sy 1, s. 38, Nis. 2013, doi: 10.1186/1129-2377-14-38.

[49] E. L. Truelove, E. E. Sommers, L. LeResche, S. F. Dworkin, ve M. Von Korff, "Clinical diagnostic criteria for TMD. New classification permits multiple diagnoses", *The Journal of the American Dental Association*, c. 123, sy 4, ss. 47-54, Nis. 1992, doi: 10.14219/jada.archive.1992.0094.

[50] M. T. John, S. F. Dworkin, ve L. A. Mancl, "Reliability of clinical temporomandibular disorder diagnoses", *Pain*, c. 118, sy 1-2, ss. 61-69, Kas. 2005, doi: 10.1016/j.pain.2005.07.018.

[51] İ. Karadağ, "Temporomandibular Eklem Rahatsizliklarinin Teşhisinde Temporomandibular Eklem Rahatsizliklari Araştirma Tanı Kriterleri ve Eklem Titreşim Analizi Yöntemlerinin Karşılaştırilması," PhD dissertation, Protetik Diş Tedavisi ABD., Gazi Üni., Ankara, Türkiye, 2022.

[52] S. Polat *vd.*, "Diagnostic Criteria for Temporomandibular Disorders: Assessment Instruments (Turkish) Temporomandibuler Düzensizlikler için Teşhis Kriterleri: Değerlendirme Araçları Turkish translation by", *Temporomandibuler Düzensizlikler için Teşhis Kriterleri: Değerlendirme Araçları*, c. www.rdc-tmdinternational.org, May. 2016.

[53] H. G. Aydin ve G. Pekkan, "Temporomandibular Rahatsızlıklarda Klinik Değerlendirmeler ve Tanı", *Turkiye Klinikleri Journal Prosthodontics-Special Topics*, c. 9, sy 2, ss. 10-18, 2023.

[54] G. Ulay ve F. N. Pekiner, "Temporomandibular eklem disfonksiyonlu bir grup hastada klinik bulguları", 2019.

[55] S. Suca ve C. Akçaboy, "KAS FASYA FONKSİYONU BOZUKLUĞU (MPD)", 1985.

[56] E. Yüce ve I. D. Ş. Yamaner, "Miyofasiyal Ağrılı Hastalarda Botulinum Toksin-A Enjeksiyonunun Yaşam Kalitesi Üzerindeki Etkilerinin Değerlendirilmesi".

[57] E. Altun, F. Yaşar, ve D. Kaymak, "Temporomandibular Eklem Görüntüleme Yöntemleri", *Turkiye Klinikleri Journal of Oral Maxillofac Surgery-Special Topics*, c. 10, sy 2, ss. 16-27, 2024.

[58] A. E. Gökce, "Temporomandibular Eklem Disfonksiyonu Olan Kadınlarda Ağrı Şiddetinin Fonksiyonel ve Ruhsal Durum Parametreleri ile İlişkisinin İncelenmesi," M.S Thesis, FTR ABD., Pamukkale Üni., Pamukkale, Türkiye, 2023.

[59] M. S. Demirsoy ve N. Akbulut, "Temporomandibuler Eklem Disfonksiyonu:Redüksiyonsuz Disk Deplasmanı / Artrosentez", *Atatürk Üniversitesi Diş Hekimliği Fakültesi Dergisi*, ss. 1-1, Nis. 2020, doi: 10.17567/ataunidfd.519370.

[60] A. N. Alagöz, S. Şirin, ve S. D. Bünül, "Temporomandibular dysfunction in patients with tension type headache", *Kocaeli Medical Journal*, c. 9, sy 3, ss. 79-85, 2020, doi: 10.5505/ktd.2020.48608.

[61] R. De Leeuw ve G. D. Klasser, *Orofacial pain: guidelines for assessment, diagnosis, and management*. Quintessence Publishing Company, Incorporated Hanover Park, IL, USA, 2018. Erişim: 07 Aralık 2023. [Çevrimiçi]. Erişim adresi: https://scholar.archive.org/work/m47ij6u33jbwvfxnot2hiwxbwi/access/wayback/http://www.stomaeduj.com/wp-content/uploads/2019/07/15-BR-2-2015-Orofacial-Pain.pdf

[62] S. Murakami, A. Takahashi, H. Nishiyama, M. Fujishita, ve H. Fuchihata, "Magnetic resonance evaluation of the temporomandibular joint disc position and configuration", *Dentomaxillofac Radiol*, c. 22, sy 4, ss. 205-207, Kas. 1993, doi: 10.1259/dmfr.22.4.8181648.

[63] W. M. Talaat, O. I. Adel, ve S. Al Bayatti, "Prevalence of temporomandibular disorders discovered incidentally during routine dental examination using the Research Diagnostic Criteria for Temporomandibular Disorders", *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, c. 125, sy 3, ss. 250-259, Mar. 2018, doi: 10.1016/j.oooo.2017.11.012.

[64] R. W. Katzberg, P. L. Westesson, R. H. Tallents, ve C. M. Drake, "Anatomic disorders of the temporomandibular joint disc in asymptomatic subjects", *Journal of Oral Maxillofacial Surgery*, c. 54, sy 2, ss. 147-153; discussion 153-155, Şub. 1996, doi: 10.1016/s0278-2391(96)90435-8.

[65] M. Ahmad *vd.*, "Research diagnostic criteria for temporomandibular disorders (RDC/TMD): development of image analysis criteria and examiner reliability for image analysis", *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, c. 107, sy 6, ss. 844-860, Haz. 2009, doi: 10.1016/j.tripleo.2009.02.023.

[66] M. M. Tasaki, P. L. Westesson, A. M. Isberg, Y. F. Ren, ve R. H. Tallents, "Classification and prevalence of temporomandibular joint disk displacement in patients and symptom-free volunteers", *American Journal of Orthodontics and Dentofacial Orthopedics*, c. 109, sy 3, ss. 249-262, Mar. 1996, doi: 10.1016/s0889-5406(96)70148-8.

[67] A. L. Young, “Internal derangements of the temporomandibular joint: A review of the anatomy, diagnosis, and management”, *The Journal of Indian Prosthodontic Society*, c. 15, sy 1, ss. 2-7, 2015, doi: 10.4103/0972-4052.156998.

[68] Ö. Sancaktar ve N. Yanikoğlu, “Temporomandibular Eklem Hastaliklarında Okluzal Splint Çeşitlerinin Kullanımı : Derleme”, *Atatürk Üniversitesi Diş Hekimliği Fakültesi Dergisi*, ss. 1-1, Eki. 2021, doi: 10.17567/ataunidfd.743161.

[69] S. Sener ve F. Akgänlü, “MRI characteristics of anterior disc displacement with and without reduction”, *Dentomaxillofacial Radiology*, c. 33, sy 4, ss. 245-252, Tem. 2004, doi: 10.1259/dmfr/17738454.

[70] R. L. Poluha, G. D. L. T. Canales, Y. M. Costa, E. Grossmann, L. R. Bonjardim, ve P. C. R. Conti, “Temporomandibular joint disc displacement with reduction: a review of mechanisms and clinical presentation”, *Journal of Applied Oral Science.*, c. 27, s. e20180433, 2019, doi: 10.1590/1678-7757-2018-0433.

[71] S. Sembronio, A. M. Albiero, C. Toro, M. Robiony, ve M. Politi, “Is there a role for arthrocentesis in recapturing the displaced disc in patients with closed lock of the temporomandibular joint?”, *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, c. 105, sy 3, ss. 274-280; discussion 281, Mar. 2008, doi: 10.1016/j.tripleo.2007.07.003.

[72] D. Manfredini, L. Guarda-Nardini, E. Winocur, F. Piccotti, J. Ahlberg, ve F. Lobbezoo, “Research diagnostic criteria for temporomandibular disorders: a systematic review of axis I epidemiologic findings”, *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, c. 112, sy 4, ss. 453-462, Eki. 2011, doi: 10.1016/j.tripleo.2011.04.021.

[73] G. Dimitroulis, M. McCullough, ve W. Morrison, “Quality-of-life survey comparing patients before and after discectomy of the temporomandibular joint”, *Journal of Oral and Maxillofacial Surgery*, c. 68, sy 1, ss. 101-106, Oca. 2010, doi: 10.1016/j.joms.2009.07.092.

[74] M. F. Dolwick, “Temporomandibular Joint Surgery for Internal Derangement”, *Dental Clinics of North America*, c. 51, sy 1, ss. 195-208, Oca. 2007, doi: 10.1016/j.cden.2006.10.003.

[75] D. Diraçoğlu *vd.*, “Arthrocentesis versus nonsurgical methods in the treatment of temporomandibular disc displacement without reduction”, *Oral Surgery, Oral*

Medicine, Oral Pathology, Oral Radiology, and Endodontology, c. 108, sy 1, ss. 3-8, Tem. 2009, doi: 10.1016/j.tripleo.2009.01.005.

[76] G. Y. Yavuz, "Temporomandibular Eklemin Redüksiyonsuz Disk Deplasmaninin Tedavisinde Metilprednizolon Asetat, Sodyum Hyaluronat ve Tenoksikamin Etkilerinin Karşılaştırılması," PhD. dissertation, Ağız, Diş Çene Cerrahisi ABD., Atatürk Üni., Erzurum, Türkiye, 2014.

[77] M. Yaltırık, A. Palancıoğlu, M. Koray, ve C. T. Turgut, "Temporomandibular joint disorders and diagnosis", *Yeditepe Dental Journal*, c. 13, sy 2, ss. 43-50, 2017, doi: 10.5505/yeditepe.2017.07078.

[78] R. N. Jamison, G. J. Fanciullo, ve J. C. Baird, "Usefulness of pain drawings in identifying real or imagined pain: accuracy of pain professionals, nonprofessionals, and a decision model", *The Journal of Pain*, c. 5, sy 9, ss. 476-482, Kas. 2004, doi: 10.1016/j.jpain.2004.08.004.

[79] S. M. Shaffer, J.-M. Brismée, P. S. Sizer, ve C. A. Courtney, "Temporomandibular disorders. Part 1: anatomy and examination/diagnosis", *Journal of Manual & Manipulative Therapy*, c. 22, sy 1, ss. 2-12, Şub. 2014, doi: 10.1179/2042618613Y.0000000060.

[80] A. Benko ve C. Sik Lányi, "History of Artificial Intelligence":, içinde *Encyclopedia of Information Science and Technology, Second Edition*, M. Khosrow-Pour, D.B.A., Ed., IGI Global, 2009, ss. 1759-1762. doi: 10.4018/978-1-60566-026-4.ch276.

[81] A. M. TURING, "I.—Computing Machineri and Intellegence", *Mind*, c. LIX, sy 236, ss. 433-460, Eki. 1950, doi: 10.1093/mind/LIX.236.433.

[82] M. Haenlein ve A. Kaplan, "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence", *California Management Review*, c. 61, sy 4, ss. 5-14, Ağu. 2019, doi: 10.1177/0008125619864925.

[83] N. Gupta, "Artificial Neural Network," 2020.

[84] Perform Yazılım. (2023). "Large Language Model (Büyük Dil Modeli) Nedir?". Erişim: 13 Aralık 2023. [Çevrimiçi]. Erişim adresi: https://performyazilim.medium.com/large-language-model-b%C3%BCy%C3%BCk-dil-modeli-nedir-9e842c842a77

[85] Elasticsearch. (2024). "What is a Large Language Model?". Erişim: 17 Aralık 2023. [Çevrimiçi]. Erişim adresi: https://www.elastic.co/what-is/large-language-models

[86] M. Javaid, A. Haleem, ve R. P. Singh, "ChatGPT for healthcare services: An emerging stage for an innovative perspective", *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, c. 3, sy 1, s. 100105, Şub. 2023, doi: 10.1016/j.tbench.2023.100105.

[87] M. Javaid, A. Haleem, R. P. Singh, S. Khan, ve I. H. Khan, "Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system", *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, c. 3, sy 2, s. 100115, Haz. 2023, doi: 10.1016/j.tbench.2023.100115.

[88] A. Nazir ve Z. Wang, "A comprehensive survey of ChatGPT: Advancements, applications, prospects, and challenges", *Meta-Radiology*, c. 1, sy 2, s. 100022, Eyl. 2023, doi: 10.1016/j.metrad.2023.100022.

[89] R. Raj, A. Singh, V. Kumar, ve P. Verma, "Analyzing the potential benefits and use cases of ChatGPT as a tool for improving the efficiency and effectiveness of business operations", *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, c. 3, sy 3, s. 100140, Eyl. 2023, doi: 10.1016/j.tbench.2023.100140.

[90] Microsoft 365 Copilot. (2025). "Microsoft 365 için Microsoft Copilot'a Genel Bakış". Microsoft Copilot, ABD. Erişim: 22 Temmuz 2024. [Çevrimiçi]. Erişim adresi: https://learn.microsoft.com/tr-tr/copilot/microsoft-365/microsoft-365-copilot-overview

[91] Microsoft 365 Copilot. (Şubat 2025). "Microsoft 365 için Microsoft Copilot'a Genel Bakış". Microsoft Copilot, ABD. Erişim: 22 Temmuz 2024. [Çevrimiçi]. Erişim adresi: https://learn.microsoft.com/tr-tr/copilot/microsoft-365/microsoft-365-copilot-overview

[92] Microsoft Copilot., "Viva Learning'de Copilot Tanıtımı - Microsoft Desteği". Microsoft Copilot. ABD. Erişim: 22 Temmuz 2024. [Çevrimiçi]. Erişim adresi: https://support.microsoft.com/tr-tr/topic/viva-learning-de-copilot-

[93] "Microsoft Copilot | Microsoft AI". Erişim: 22 Temmuz 2024. [Çevrimiçi]. Erişim adresi: https://www.microsoft.com/tr-tr/copilot

[94] Google Gemini. (2023, Aralık). "Introducing Gemini: our largest and most capable AI model", Google. Erişim: 20 Nisan 2025. [Çevrimiçi]. Erişim adresi: https://blog.google/technology/ai/google-gemini-ai/

[95] Open AI ChatGPT (2023). “ChatGPT”. Erişim: 22 Temmuz 2024. Open AI. [Çevrimiçi]. Erişim adresi: https://chatgpt.com

[96] Tracxn, (Mart). “Docus - 2025 Company Profile - Tracxn”. Erişim: 08 Nisan 2025. [Çevrimiçi]. Erişim adresi: https://tracxn.com/d/companies/docus/__AkraO5bm0_hU3FzL5zWmtWiO_4pADe1SbF7rmvomhlg

[97] Anthropic (2024). “Claude”. Anthropic. ABD. Erişim: 22 Temmuz 2024. [Çevrimiçi]. Erişim adresi: https://claude.ai/chat/96a37872-bff9-44bb-b167-adebdaee44fb

[98] Perplexity AI. “Perplexity”. Erişim: 22 Temmuz 2024. [Çevrimiçi]. Erişim adresi: https://www.perplexity.ai/

[99] T. Shan, F. R. Tay, ve L. Gu, “Application of Artificial Intelligence in Dentistry”, *J Dent Res*, c. 100, sy 3, ss. 232-244, Mar. 2021, doi: 10.1177/0022034520969115.

[100] S. B. Khanagar *vd.*, “Developments, application, and performance of artificial intelligence in dentistry – A systematic review”, *Journal of Dental Sciences*, c. 16, sy 1, ss. 508-522, Oca. 2021, doi: 10.1016/j.jds.2020.06.019.

[101] P. Bouletreau, M. Makaremi, B. Ibrahim, A. Louvrier, ve N. Sigaux, “Artificial Intelligence: Applications in orthognathic surgery”, *Journal of Stomatology Oral and Maxillofacial Surgery*, c. 120, sy 4, ss. 347-354, Eyl. 2019, doi: 10.1016/j.jormas.2019.06.001.

[102] F. C. Setzer *vd.*, “Artificial Intelligence for the Computer-aided Detection of Periapical Lesions in Cone-beam Computed Tomographic Images”, *Journal of Endodontics*, c. 46, sy 7, ss. 987-993, Tem. 2020, doi: 10.1016/j.joen.2020.03.025.

[103] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, ve M. K. Khan, “Medical Image Analysis using Convolutional Neural Networks: A Review”, *Journal of Medical Systems*, c. 42, sy 11, s. 226, Eki. 2018, doi: 10.1007/s10916-018-1088-1.

[104] B. Kula, A. Kula, F. Bagcier, ve B. Alyanak, “Artificial intelligence solutions for temporomandibular joint disorders: Contributions and future potential of ChatGPT”, *Korean Journal of Orthodontics*, c. 55, sy 2, ss. 131-141, Mar. 2025, doi: 10.4041/kjod24.106.

[105] J.-H. Lee, D.-H. Kim, S.-N. Jeong, ve S.-H. Choi, “Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural

network algorithm", *Journal of Periodontal & Implant Science*, c. 48, sy 2, ss. 114-123, Nis. 2018, doi: 10.5051/jpis.2018.48.2.114.

[106] S. Raith *vd.*, "Artificial Neural Networks as a powerful numerical tool to classify specific features of a tooth based on 3D scan data", *Comput Biol Med*, c. 80, ss. 65-76, Oca. 2017, doi: 10.1016/j.compbiomed.2016.11.013.

[107] R. Ohrbach, "Diagnostic Criteria for Temporomandibular Disorders: Assessment Instruments," 2014.

[108] T. H. Hotta, M. F. R. Vicente, A. C. dos Reis, O. L. Bezzon, C. Bataglion, ve A. Bataglion, "Combination therapies in the treatment of temporomandibular disorders: a clinical report", *The Journal of Prosthetic Dentistry*, c. 89, sy 6, ss. 536-539, Haz. 2003, doi: 10.1016/s0022-3913(03)00077-5.

[109] W. J. Binder, M. F. Brin, A. Blitzer, ve J. M. Pogoda, "Botulinum toxin type A (BOTOX) for treatment of migraine", *Dis Mon*, c. 48, sy 5, ss. 323-335, May. 2002, doi: 10.1053/mda.2002.24423.

[110] E. Schiffman, G. Anderson, J. Fricton, K. Burton, ve K. Schellhas, "Diagnostic criteria for intraarticular T.M. disorders", *Community Dent Oral Epidemiol*, c. 17, sy 5, ss. 252-257, Eki. 1989, doi: 10.1111/j.1600-0528.1989.tb00628.x.

[111] S. F. Dworkin, J. Sherman, L. Mancl, R. Ohrbach, L. LeResche, ve E. Truelove, "Reliability, validity, and clinical utility of the research diagnostic criteria for Temporomandibular Disorders Axis II Scales: depression, non-specific physical symptoms, and graded chronic pain", *Journal of Orofacial Pain*, c. 16, sy 3, ss. 207-220, 2002.

[112] Z. Kanjee, B. Crowe, ve A. Rodman, "Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge", *Journal of the American Medical Association*, c. 330, sy 1, ss. 78-80, Tem. 2023, doi: 10.1001/jama.2023.8288.

[113] B. N. Karahan, E. Emekli, ve M. A. Altın, "Artificial Intelligence-Based Chatbots' Ability to Interpret Mammography Images: A Comparison of Chat-GPT 4o and Claude 3.5", *European Journal of Therapeutics*, c. 31, sy 1, ss. 28-34, Şub. 2025, doi: 10.58600/eurjther2599.

[114] A. Ríos-Hoyo, N. L. Shan, A. Li, A. T. Pearson, L. Pusztai, ve F. M. Howard, "Evaluation of large language models as a diagnostic aid for complex medical cases",

Frontiers in Medicine, c. 11, s. 1380148, Haz. 2024, doi: 10.3389/fmed.2024.1380148.

[115] G. K. Gupta, A. Singh, S. V. Manikandan, ve A. Ehtesham, "Digital Diagnostics: The Potential of Large Language Models in Recognizing Symptoms of Common Illnesses", *AI*, c. 6, sy 1, s. 13, Oca. 2025, doi: 10.3390/ai6010013.

[116] A. Moglia, K. Georgiou, P. Cerveri, L. Mainardi, R. M. Satava, ve A. Cuschieri, "Large language models in healthcare: from a systematic review on medical examinations to a comparative analysis on fundamentals of robotic surgery online test", *Artif Intell Rev*, c. 57, sy 9, s. 231, Ağu. 2024, doi: 10.1007/s10462-024-10849-5.

[117] M. Giuffrè *vd.*, "Systematic review: The use of large language models as medical chatbots in digestive diseases", *Aliment Pharmacol Ther*, c. 60, sy 2, ss. 144-166, Tem. 2024, doi: 10.1111/apt.18058.

[118] C. Fang *vd.*, "How does ChatGPT-4 preform on non-English national medical licensing examination? An evaluation in Chinese language", *PLOS Digit Health*, c. 2, sy 12, s. e0000397, Ara. 2023, doi: 10.1371/journal.pdig.0000397.

[119] L. Chen, M. Zaharia, ve J. Zou, "How is ChatGPT's behavior changing over time?", 31 Ekim 2023, *arXiv*: arXiv:2307.09009. doi: 10.48550/arXiv.2307.09009.

[120] E. Goh *vd.*, "ChatGPT Influence on Medical Decision-Making, Bias, and Equity: A Randomized Study of Clinicians Evaluating Clinical Vignettes", *medRxiv*, s. 2023.11.24.23298844, Kas. 2023, doi: 10.1101/2023.11.24.23298844.

[121] T. Savage, A. Nayak, R. Gallo, E. Rangan, ve J. H. Chen, "Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine", *npj Digital Medicine*, c. 7, sy 1, s. 20, Oca. 2024, doi: 10.1038/s41746-024-01010-1.

[122] A. A. Tierney *vd.*, "Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation", *NEJM Catalyst*, c. 5, sy 3, s. CAT.23.0404, Şub. 2024, doi: 10.1056/CAT.23.0404.

[123] T. H. Kung *vd.*, "Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models", *PLOS Digital Health*, c. 2, sy 2, s. e0000198, Şub. 2023, doi: 10.1371/journal.pdig.0000198.

[124] J. W. Ayers *vd.*, "Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum", *JAMA Internal*

Medicine, c. 183, sy 6, ss. 589-596, Haz. 2023, doi: 10.1001/jamainternmed.2023.1838.

[125] Y. H. Yeo *vd.*, "Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma", *Clinical and Molecular Hepatology*, c. 29, sy 3, ss. 721-732, Tem. 2023, doi: 10.3350/cmh.2023.0089.

[126] I. Zheleiko, "Natural language processing in lifelong learning choices : a case of Finland", 2023, Erişim: 18 Mart 2025. [Çevrimiçi]. Erişim adresi: https://lutpub.lut.fi/handle/10024/165104

[127] C. A. Gao *vd.*, "Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers", *npj Digital Medicine*, c. 6, sy 1, s. 75, Nis. 2023, doi: 10.1038/s41746-023-00819-6.

[128] R. Tsang, "Practical Applications of ChatGPT in Undergraduate Medical Education", *J Med Educ Curric Dev*, c. 10, s. 23821205231178449, 2023, doi: 10.1177/23821205231178449.

[129] J. Clusmann *vd.*, "The future landscape of large language models in medicine", *Communications Medicine*, c. 3, sy 1, s. 141, Eki. 2023, doi: 10.1038/s43856-023-00370-1.

[130] R. Ali *vd.*, "Performance of ChatGPT and GPT-4 on Neurosurgery Written Board Examinations", *Neurosurgery*, c. 93, sy 6, ss. 1353-1365, Ara. 2023, doi: 10.1227/neu.0000000000002632.

[131] R. Ali *vd.*, "Performance of ChatGPT, GPT-4, and Google Bard on a Neurosurgery Oral Boards Preparation Question Bank", *Neurosurgery*, c. 93, sy 5, ss. 1090-1098, Kas. 2023, doi: 10.1227/neu.0000000000002551.

[132] P. Giannos, "Evaluating the limits of AI in medical specialisation: ChatGPT's performance on the UK Neurology Specialty Certificate Examination", *BMJ Neurol Open*, c. 5, sy 1, s. e000451, Haz. 2023, doi: 10.1136/bmjno-2023-000451.

[133] Y. Huang *vd.*, "Benchmarking ChatGPT-4 on a radiation oncology in-training exam and Red Journal Gray Zone cases: potentials and challenges for ai-assisted medical education and decision making in radiation oncology", *Frontiers in Oncology*, c. 13, s. 1265024, 2023, doi: 10.3389/fonc.2023.1265024.

[134] J. Kasai, Y. Kasai, K. Sakaguchi, Y. Yamada, ve D. Radev, “Evaluating GPT-4 and ChatGPT on Japanese Medical Licensing Examinations”, 05 Nisan 2023, *arXiv*: arXiv:2303.18027. doi: 10.48550/arXiv.2303.18027.

[135] D. Jang, T.-R. Yun, C.-Y. Lee, Y.-K. Kwon, ve C.-E. Kim, “GPT-4 can pass the Korean National Licensing Examination for Korean Medicine Doctors”, 2023, doi: 10.48550/ARXIV.2303.17807.

[136] N. Oh, G.-S. Choi, ve W. Y. Lee, “ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models”, *Annals of Surgical Treatment and Research*, c. 104, sy 5, ss. 269-273, May. 2023, doi: 10.4174/astr.2023.104.5.269.

[137] L. Passby, N. Jenko, ve A. Wernham, “Performance of ChatGPT on Specialty Certificate Examination in Dermatology multiple-choice questions”, *Clinical and Experimental Dermatology*, c. 49, sy 7, ss. 722-727, Haz. 2024, doi: 10.1093/ced/llad197.

[138] J. Smith, P. M. Choi, ve P. Buntine, “Will code one day run a code? Performance of language models on ACEM primary examinations and implications”, *Emergency Medicine Australasia*, c. 35, sy 5, ss. 876-878, Eki. 2023, doi: 10.1111/1742-6723.14280.

[139] B. Huo *vd.*, “The performance of artificial intelligence large language model-linked chatbots in surgical decision-making for gastroesophageal reflux disease”, *Surgical Endoscopy*, c. 38, sy 5, ss. 2320-2330, May. 2024, doi: 10.1007/s00464-024-10807-w.

[140] B. Lim *vd.*, “Can AI Answer My Questions? Utilizing Artificial Intelligence in the Perioperative Assessment for Abdominoplasty Patients”, *Aesthetic Plastic Surgery*, c. 48, sy 22, ss. 4712-4724, Kas. 2024, doi: 10.1007/s00266-024-04157-0.

[141] A. Duran, O. Cortuk, ve B. Ok, “Future Perspective of Risk Prediction in Aesthetic Surgery: Is Artificial Intelligence Reliable?”, *Aesthetic Surgery Journal*c. 44, sy 11, ss. NP839-NP849, Eki. 2024, doi: 10.1093/asj/sjae140.

[142] V. Ponzo *vd.*, “Comparison of the Accuracy, Completeness, Reproducibility, and Consistency of Different AI Chatbots in Providing Nutritional Advice: An Exploratory Study”, *JCM*, c. 13, sy 24, s. 7810, Ara. 2024, doi: 10.3390/jcm13247810.

[143] S. Günay, A. Öztürk, ve Y. Yiğit, "The accuracy of Gemini, GPT-4, and GPT-4o in ECG analysis: A comparison with cardiologists and emergency medicine specialists", *The American Journal of Emergency Medicine*, c. 84, ss. 68-73, Eki. 2024, doi: 10.1016/j.ajem.2024.07.043.

[144] J. M. Hoppe, M. K. Auer, A. Strüven, S. Massberg, ve C. Stremmel, "ChatGPT With GPT-4 Outperforms Emergency Department Physicians in Diagnostic Accuracy: Retrospective Analysis", *Journal of Medical Internet Research*, c. 26, s. e56110, Tem. 2024, doi: 10.2196/56110.

[145] S. Takagi, T. Watari, A. Erabi, ve K. Sakaguchi, "Performance of GPT-3.5 and GPT-4 on the Japanese Medical Licensing Examination: Comparison Study", *Journal of Medical Internet Research*, c. 9, s. e48002, Haz. 2023, doi: 10.2196/48002.

[146] M. C. S. Menezes *vd.*, "The potential of Generative Pre-trained Transformer 4 (GPT-4) to analyse medical notes in three different languages: a retrospective model-evaluation study", *The Lancet Digital Health*, c. 7, sy 1, ss. e35-e43, Oca. 2025, doi: 10.1016/S2589-7500(24)00246-2.

[147] A. Harigai *vd.*, "Response accuracy of GPT-4 across languages: insights from an expert-level diagnostic radiology examination in Japan", *Japanese Journal of Radiology*, c. 43, sy 2, ss. 319-329, 2025, doi: 10.1007/s11604-024-01673-6.

[148] D. Lee ve G. Kader, "The Emergence of Strategic Reasoning of Large Language Models", 07 Şubat 2025, *arXiv*: arXiv:2412.13013. doi: 10.48550/arXiv.2412.13013.

[149] R. T. Oğuzman ve Z. Z. Yurdabakan, "Performance of Chat Generative Pretrained Transformer and Bard on the Questions Asked in the Dental Specialty Entrance Examination in Turkey Regarding Bloom's Revised Taxonomy", Current Research in Dental Sciences, c. 34, sy 1, Art. sy 1, Oca. 2024, doi: 10.5152/CRDS.2024.23261.