



Küçük Hücreli Dışı Akciğer Karsinomlarında PD-L1 Skorlamasında Patolog ve Yapay Zekâ Uyumu: Tek Merkezli Validasyon Çalışması

Pathologist–Artificial Intelligence Concordance in PD-L1 Scoring in Non-Small Cell Lung Carcinoma: A Single-Center Validation Study

Sevilay ÖZMEN¹ , Demir Kürşat YILDIZ² , Rabia DEMİRTAŞ¹ , Berk MADEN¹ , Burçin ERGÜL³ , Numan BULUT⁴ , Özge BOZKURT⁵ , Serenay DEMİRCİ⁶ , Arzu AKÇAY⁷ , Ömer ARAZ⁸ , Yakup HİLAL⁹ , Salim Başol TEKİN¹⁰

¹Atatürk Üniversitesi Tıp Fakültesi, Tıbbi Patoloji Anabilim Dalı, Erzurum, Türkiye

²Van Yüzüncü Yıl Üniversitesi Tıp Fakültesi, Tıbbi Patoloji Anabilim Dalı, Van, Türkiye

³Bursa Şehir Hastanesi Tıbbi Patoloji Laboratuvarı, Bursa, Türkiye

⁴Erzurum Şehir Hastanesi Tıbbi Patoloji Laboratuvarı, Erzurum, Türkiye

⁵Ankara Atatürk Sanatoryumu Eğitim ve Araştırma Hastanesi Patoloji Laboratuvarı, Ankara, Türkiye

⁶Iğdır Dr. Nevruz Erez Devlet Hastanesi Tıbbi Patoloji Laboratuvarı, Iğdır, Türkiye

⁷Yeni Yüzyıl Üniversitesi Tıp Fakültesi, Tıbbi Patoloji Anabilim Dalı, İstanbul, Türkiye

⁸Atatürk Üniversitesi Tıp Fakültesi, Göğüs Hastalıkları Anabilim Dalı, Erzurum, Türkiye

⁹Kars Harakani Devlet Hastanesi Patoloji Laboratuvarı, Kars, Türkiye

¹⁰Acıbadem Bursa Hastanesi Onkoloji Kliniği, Bursa, Türkiye

ÖZET

Amaç: Programlanmış ölüm ligandı 1 (PD-L1) tümör proporsiyon skorunun (TPS) klinik eşiklere göre sınıflandırılması değerlendiriciye bağlı değişkenlik gösterebilir. Bu çalışmada, küçük hücreli dışı akciğer karsinomlarında patoloğun ilk değerlendirmesi (KY-1) ile Gemini 3.1 Pro yapay zekâ (YZ) modeli arasındaki uyum ve uyumsuz olgulardaki kör ikinci patolog değerlendirmesi (KY-2) incelendi.

Gereç ve Yöntemler: Tek merkezli retrospektif validasyon çalışmasına 25 arşiv olgusu alındı. Dört mikrometre kalınlığındaki formalinle fikse parafine gömülü kesitler, CAL10 klonu ve ultraView Universal DAB sistemiyle BenchMark ULTRA cihazında boyandı. Olgu başına tüm tümörlü alanları kapsayan, toplam en az 100 canlı tümör hücresi içeren en az 10 Full HD görüntü 20 objektifle kaydedildi. Her Türkçe istem bir kez çalıştırıldı ve tek YZ çıktısı kullanıldı. TPS <%1, %1–49 ve ≥%50 olarak sınıflandırıldı. KY-1–YZ uyumu kappa ve kesin McNemar testleriyle, KY-2 alt kümesi keşifsel olarak değerlendirildi.

Bulgular: KY-1 dağılımı 15 negatif, 6 düşük ve 4 yüksek pozitif; YZ dağılımı 14, 5 ve 6 idi. Tam uyum 16/25 (%64,0), Cohen kappa 0,377 ve karesel ağırlıklı kappa 0,531 bulundu. YZ beş olguda daha yüksek, dört olguda daha düşük kategori verdi. %1 ve %50 eşiklerinde marjinal fark yoktu (p=1,000; p=0,625). KY-2 bulunan sekiz uyumsuz olgunun yedisinde KY-2, KY-1 ile aynı kategorideydi (kappa 0,784); KY-2–YZ kappa -0,366 idi.

Sonuç: KY-1 ile YZ arasında sınırlı-orta düzeyde uyum vardı ve klinik eşiklerde sistematik yön farkı gösterilemedi. KY-2, seçilmiş uyumsuz olguların çoğunda KY-1 sonucunu yineledi. Bağımsız çok okuyuculu konsensüs standardı bulunmadığından YZ doğruluğu hakkında kesin çıkarım yapılamaz; klinik kullanım öncesinde yerel validasyon, standartlaştırılmış görüntüleme ve istem protokolü ile kör uzlaşısı değerlendirilmesi gereklidir.

Anahtar Sözcükler: Küçük hücreli dışı akciğer karsinomu, Programlanmış hücre ölümü ligandı 1, İmmünohistokimya, Gözlemci değişkenliği, Yapay zekâ

(J Curr Pathol. 2026;10(2):96-103)

Geliş Tarihi : 07.08.2026 Kabul Tarihi : 02.09.2026

İletişim: Sevilay Özmen

Atatürk Üniversitesi Tıp Fakültesi, Tıbbi Patoloji Anabilim Dalı,
Erzurum, Türkiye

E-posta: rtekozmen@gmail.com



ABSTRACT

Objective: Classification of programmed death ligand 1 (PD-L1) tumor proportion scores at clinical thresholds may vary by evaluator. This study assessed concordance between the primary pathologist assessment (KY-1) and the Gemini 3.1 Pro artificial intelligence model (YZ), and explored a blinded second pathologist assessment (KY-2) in discordant cases.

Material and Methods: This single-center retrospective validation study included 25 archival cases. Four-micrometer formalin-fixed paraffin-embedded sections were stained with CAL10 and ultraView Universal DAB on a BenchMark ULTRA platform. At least 10 Full HD images covering all tumor-bearing areas and collectively containing at least 100 viable tumor cells were captured per case using a 20 objective. Each Turkish prompt was run once, and a single YZ output was recorded. Scores were categorized as <1%, 1–49%, or ≥50%. KY-1–YZ agreement was assessed with Cohen's kappa and exact McNemar tests; the KY-2 subset was analyzed exploratorily.

Results: KY-1 classified 15 cases as negative, 6 as low-positive, and 4 as high-positive; YZ classified 14, 5, and 6 cases, respectively. Exact agreement was 16/25 (64.0%), Cohen's kappa was 0.377, and quadratically weighted kappa was 0.531. YZ assigned a higher category in five cases and a lower category in four. Marginal differences were absent at the 1% and 50% thresholds ($p=1.000$; $p=0.625$). In eight discordant cases with KY-2 data, KY-2 matched KY-1 in seven cases ($kappa = 0.784$); the KY-2–YZ kappa was -0.366 .

Conclusion: KY-1 and YZ showed limited-to-moderate agreement, without a systematic directional difference at clinical thresholds. KY-2 reproduced KY-1 in most selected discordant cases. Because no independent multi-reader consensus standard was available, YZ accuracy could not be established. Local validation, standardized imaging and prompting, and blinded adjudication are required before clinical implementation.

Keywords: Carcinoma, Non-small-cell lung, Programmed cell death 1 ligand 1 protein, Immunohistochemistry, Observer variation, Artificial intelligence

GİRİŞ

Akciğer kanseri, dünya genelinde kanser insidansı ve mortalitesinin başlıca nedenlerinden biridir. Küçük hücreli dışı akciğer karsinomu (KHDAK), olguların büyük çoğunluğunu oluşturur. Güncel akciğer tümör sınıflaması, doğru alt tiplere için morfoloji, immünohistokimya ve moleküler patolojinin birlikte değerlendirilmesini esas almaktadır (1-3).

İmmün kontrol noktası inhibitörlerinin kullanıma girmesiyle programlanmış ölüm ligandı 1 (PD-L1) testi, KHDAK tedavisinde temel biyobelirteçlerden biri hâline gelmiştir. PD-L1, canlı tümör hücrelerinde herhangi bir şiddette parsiyel veya komplet membranöz boyanma gösteren hücrelerin yüzdesini ifade eden tümör proporsiyon skoru (TPS) ile raporlanmaktadır. <1%, %1–49 ve ≥50 kategorileri, pembrolizumab çalışmalarında tanımlanan ve klinik karar süreçlerinde kullanılan eşiklerle uyumludur (4-9).

TPS görünüşte basit bir ölçüm olsa da yarı nicel görsel tahmine dayanır. Tümör heterojenitesi, örnek türü, boyama kalitesi, nekroz, ezilme artefaktı, immün hücre boyanması ve değerlendirici deneyimi sonucu etkileyebilir. Blueprint ve yöntem karşılaştırma çalışmaları analitik değişkenleri tanımlamış; yeniden üretilebilirlik çalışmalarında uyumun kullanılan yöntem, eşik, örnek ve eğitim bağlamına göre değişebildiği gösterilmiştir (10-16).

YZ sistemleri PD-L1 değerlendirmesinde ikinci gözlemci olarak kullanılabilir. Bununla birlikte YZ çıktısının patolojik değerlendirilmesiyle uyumlu olması, tek başına doğru olduğu anlamına gelmez. Bağımsız ve kör bir referans standardı olmadan daha yüksek ya da daha düşük skorlama eğiliminin kaynağı belirlenemez. Bu çalışmanın amacı, tek merkezli 25 olguluk KHDAK kohortunda arşiv kayıtlarındaki PD-L1 da-

ğılımını tanımlamak, KY-1 ile Gemini 3.1 Pro YZ modeli arasındaki üç kategorili ve eşik temelli uyumu değerlendirmek ve uyumsuz olgularda mevcut KY-2 sonuçlarını keşifsel olarak incelemektir.

GEREÇ ve YÖNTEMLER

Çalışma Tasarımı ve Olgu Seçimi

Bu çalışma tek merkezli, retrospektif ve arşiv temelli bir patoloji validasyon çalışması olarak tasarlandı. Histolojik tanısı doğrulanmış, klinikopatolojik verileri bulunan ve KY-1 ile YZ için tam eşleştirilmiş değerlendirme setine sahip 25 KHDAK olgusu çalışmaya alındı. Olguların tamamında arşiv kayıtlarında nihai PD-L1 TPS kategorisi mevcuttu.

Histolojik tanılar skuamöz hücreli karsinom, adenokarsinom, müsinöz adenokarsinom, pulmoner enterik adenokarsinom, sınıflandırılmayan KHDAK ve büyük hücreli nöroendokrin karsinom olarak gruplandı (2,3).

Histopatolojik Değerlendirme ve PD-L1 İmmünohistokimyası

İmmünohistokimyasal inceleme, formalinle fikse parafine gömülü (FFPE) biyopsi bloklarından hazırlanan 4 µm kalınlığındaki kesitlerde gerçekleştirildi. PD-L1 boyanması, Roche Ventana BenchMark ULTRA IHC/ISH platformunda PDL-1 DAB protokolü (#339; sürüm 3; 03.11.2025) kullanılarak Biocare CAL10 tavşan monoklonal antikoru ile yapıldı. Cihaz içi deparafinizasyonun ardından antijen geri kazanımı ULTRA Cell Conditioning Solution 1 (CC1) ile 95°C'de 64 dakika uygulandı. Primer antikor, cihazda PREP KIT 41 olarak tanımlanan dispenserden 36°C'de 60 dakika inkübe edildi. Saptama ultraView Universal DAB sistemiyle gerçekleştirildi; UV Inhibitor 4 dakika, UV HRP Universal Multimer 8 dakika, UV DAB/

UV DAB H₂O₂ 8 dakika ve UV Copper 4 dakika uygulandı. Kesitler Hematoxylin II ile 8 dakika karşıt boyandı ve Bluing Reagent ile 4 dakika işlem gördü. Her çalışma serisinde dış pozitif kontrol olarak tonsil dokusu kullanıldı; retiküler epitel hücrelerinde beklenen kuvvetli membranöz boyanma izlendi.

TPS, canlı tümör hücrelerinde herhangi bir şiddette ($\geq 1+$) parsiyel veya komplet membranöz boyanma gösteren hücrelerin tüm canlı tümör hücrelerine oranı olarak tanımlandı ve "TPS = (PD-L1 pozitif tümör hücresi sayısı / toplam canlı tümör hücresi sayısı) 100" formülüyle hesaplandı. Değerlendirme en az 100 canlı tümör hücresi üzerinden yapıldı. Yalnız sitoplazmik granüler boyanma, nekrotik hücreler ve immün hücre boyanması TPS hesabına katılmadı (9). Immün hücre (IC) alanı değerlendirmesi yalnızca en az 50 tümör hücresi içeren alanlarda yapıldı ve birincil uyum analizine dahil edilmedi. Sonuçlar negatif ($<1\%$), düşük pozitif ($1-49\%$) ve yüksek pozitif ($\geq 50\%$) olmak üzere üç kategoride sınıflandırıldı.

Mikroskopik Görüntüleme ve Değerlendiriciler

Patologun ilk değerlendirmesi KY-1, YZ değerlendirmesi YZ ve uyumsuz olgulardaki kör ikinci patolog değerlendirme KY-2 olarak kodlandı. KY-2, KY-1–YZ arasında kategori uyumsuzluğu bulunan dokuz olgunun sekizinde, önceki değerlendirme sonuçları gizlenerek gerçekleştirildi; bir uyumsuz olguda KY-2 kaydı yoktu. Mikroskopik alanlar Nikon Eclipse Ni mikroskop ve iPhone 17 akıllı telefon kullanılarak Full HD çözünürlükte, 20 objektif altında kaydedildi. Aynı araştırmacı tarafından her preparatta tüm tümörlü alanları kapsayacak ve toplamda en az 100 canlı tümör hücresi içerecek şekilde olgu başına en az 10 mikroskopik fotoğraf çekildi. Nekroz, yoğun ezilme artefaktı ve odak dışı alanlar görüntü setine alınmadı.

Kaydedilen fotoğraflar Gemini 3.1 Pro YZ modeliyle Türkçe istemler aracılığıyla değerlendirildi. İstemde canlı tümör hücrelerinin belirlenmesi, immün hücrelerin TPS hesabı dışında tutulması, TPS'nin verilen formülle hesaplanması ve sonucun $<1\%$, $1-49\%$ veya $\geq 50\%$ kategorilerinden biriyle bildirilmesi istendi. Her vaka için hazırlanan istem bir kez çalıştırıldı ve elde edilen tek değerlendirme çıktısı analize alındı; yeniden çalıştırma veya çıktı seçimi yapılmadı. KY-1, KY-2 ve YZ sonuçları olgu düzeyinde eşleştirildi ve aynı üç TPS kategorisine dönüştürüldü.

Arşiv kayıtlarındaki nihai TPS kategorileri yalnızca kohortun PD-L1 dağılımını tanımlamak için kullanıldı. Bu kayıtların bağımsız ve kör bir konsensüs süreciyle oluşturulduğu belgelenmediğinden, KY-1, KY-2 veya YZ doğruluğunu hesaplamak için referans standart kabul edilmedi.

Sonlanım Noktaları

Birincil sonlanım noktası, $<1\%$, $1-49\%$ ve $\geq 50\%$ kategorilerinde KY-1–YZ uyumuydu. İkincil sonlanım noktaları 1% ve 50% klinik eşiklerinde ikili uyum, uyumsuzlukların yönü, değerlendiriciye göre kategori dağılımı ve arşiv kayıtlarındaki nihai PD-L1 kategorilerinin klinikopatolojik dağılımıydı.

KY-2 verisi bulunan uyumsuz olgularda KY-1–KY-2 ve KY-2–YZ çiftli uyumu ile üç değerlendiricili uyum keşifsel sonlanım noktaları olarak incelendi.

İstatistiksel Analiz

Ham veri tablosundaki sayısal TPS değerleri hücre biçimi dikkate alınarak yüzdeye dönüştürüldü; örneğin yüzde biçimli 0,10 değeri 10% olarak işlendi. " $<1\%$ " biçimindeki metinsel kayıtlar negatif, $1-49\%$ aralığındaki değerler düşük pozitif ve $\geq 50\%$ değerler yüksek pozitif olarak kodlandı. Sürekli değişkenler ortalama±standart sapma, medyan, çeyrekler arası aralık ve aralıkla; kategorik değişkenler sayı ve yüzdeyle özetlendi. Yaş, nihai TPS kategorileri arasında Kruskal–Wallis testiyle; yüksek ekspresyonun cinsiyet ve ana histolojik gruplarla ilişkisi Fisher kesin testiyle değerlendirildi. Küçük ve seyrek alt gruplar nedeniyle çok değişkenli model kurulmadı.

KY-1–YZ üç kategorili uyumu için ham yüzde uyum, Cohen kappa ve sıralı kategorilere karesel ağırlık veren kappa hesaplandı. Ham uyum için Clopper–Pearson kesin 95% güven aralığı, üç kategorili kappa katsayıları için asimptotik 95% güven aralığı verildi. 1% ve 50% eşiklerinde eşleştirilmiş marjinal farklılık kesin McNemar testiyle incelendi. KY-2 verisi bulunan seçilmiş uyumsuz olgularda KY-1–KY-2 ve KY-2–YZ için ham uyum, Cohen kappa ve karesel ağırlıklı kappa; üç değerlendirici için Fleiss kappa hesaplandı. Bu alt küme uyumsuzluğa göre seçildiği ve yalnız sekiz olgudan oluştuğu için sonuçlar keşifsel ve betimleyici olarak yorumlandı. Kappa katsayıları katı nitel etiketlerden çok büyüklük ve yönleriyle değerlendirildi (17). Analizler Python 3.13.5; NumPy 2.3.5, SciPy 1.17.0, scikit-learn 1.8.0 ve statsmodels 0.14.6 kullanılarak yapıldı. İki yönlü $p<0,05$ anlamlı kabul edildi; alt grup analizleri keşifsel olarak yorumlandı.

Çalışma raporlaması Gözlemsel Çalışmaların Raporlanması'nın Güçlendirilmesi ilkeleri dikkate alınarak hazırlandı (18).

BULGULAR

Klinikopatolojik Özellikler

Toplam 25 olgunun yaş ortalaması $64,5 \pm 7,0$ yıl, medyan yaşı 62 yıl (çeyrekler arası aralık 59–69) ve yaş aralığı 53–78 yıl idi. Olguların 20'si erkek ($80,0\%$), 5'i kadındı ($20,0\%$). En sık histolojik grup skuamöz hücreli karsinomdu ($12/25$; $48,0\%$). Bunu nonmüsinöz adenokarsinom ($8/25$; $32,0\%$), müsinöz adenokarsinom ($2/25$; $8,0\%$), sınıflandırılmayan KHDAK ($2/25$; $8,0\%$) ve büyük hücreli nöroendokrin karsinom ($1/25$; $4,0\%$) izledi (Tablo I).

Arşiv Kayıtlarındaki Nihai PD-L1 Dağılımı

Nihai TPS kayıtlarında 14 olgu ($56,0\%$) negatif, 5 olgu ($20,0\%$) düşük pozitif ve 6 olgu ($24,0\%$) yüksek pozitif. Nonmüsinöz adenokarsinomların $3/8$ 'inde ($37,5\%$) ve skuamöz hücreli karsinomların $3/12$ 'inde ($25,0\%$) yüksek ekspresyon görüldü; iki grup arasındaki fark anlamlı değildi (Fisher kesin testi; $p=0,617$). Müsinöz adenokarsinom, sınıflandırılmayan

KHDAK ve büyük hücreli nöroendokrin karsinom gruplarında yüksek ekspresyon saptanmadı. TPS kategorileri arasında yaş farklı değildi (Kruskal-Wallis $p=0,153$) ve cinsiyet yüksek ekspresyonla ilişkili değildi (Fisher kesin $p=1,000$) (Tablo II).

KY-1-YZ Uyum

Analizde 25 eşleştirilmiş değerlendirme yer aldı. KY-1, 15 olguyu negatif (%60,0), 6 olguyu düşük pozitif (%24,0) ve 4 olguyu yüksek pozitif (%16,0) sınıflandırdı. YZ, 14 olguyu negatif (%56,0), 5 olguyu düşük pozitif (%20,0) ve 6 olguyu yüksek pozitif (%24,0) sınıflandırdı (Tablo III).

Tablo I: Çalışma kohortunun klinikopatolojik özellikleri.

Özellik	Değer
Toplam olgu	25
Yaş, yıl; ortalama $\pm$ SS	64,5 $\pm$ 7,0
Yaş, yıl; medyan (CAA)	62 (59–69)
Yaş aralığı	53–78
Erkek	20 (%80,0)
Kadın	5 (%20,0)
Skumöz hücreli karsinom	12 (%48,0)
Nonmüsinöz adenokarsinom	8 (%32,0)
Müsinöz adenokarsinom	2 (%8,0)
Sınıflandırılmayan KHDAK	2 (%8,0)
Büyük hücreli nöroendokrin karsinom	1 (%4,0)

CAA: Çeyrekler arası aralık, KHDAK: Küçük hücreli dışı akciğer karsinomu, SS: Standart sapma.

Üç kategorili tam uyum 16/25 (%64,0; %95 güven aralığı %42,5–%82,0) idi. Cohen kappa 0,377 (%95 güven aralığı 0,072–0,682), karesel ağırlıklı kappa 0,531 (%95 güven aralığı 0,196–0,866) bulundu. Dokuz uyumsuzluğun beşinde YZ, KY-1'den daha yüksek kategori atadı: iki olgu negatiften düşük pozitif, iki olgu negatiften yüksek pozitif ve bir olgu düşük pozitiften yüksek pozitif geçti. Dört uyumsuzlukta YZ daha düşük kategori verdi: üç olgu düşük pozitiften negatife, bir olgu yüksek pozitiften düşük pozitif geçti.

%1 eşliğinde KY-1'in pozitiflik oranı 10/25 (%40,0), YZ'nin pozitiflik oranı 11/25 (%44,0) idi; ikili uyum 18/25 (%72,0; %95 güven aralığı %50,6–%87,9), kappa 0,426 ve kesin McNemar $p=1,000$ bulundu. %50 eşliğinde KY-1, 4/25 (%16,0), YZ ise 6/25 (%24,0) olguyu yüksek pozitif sınıflandırdı; ham uyum 21/25 (%84,0; %95 güven aralığı %63,9–%95,5), kappa 0,505 ve kesin McNemar $p=0,625$ idi (Tablo IV).

Kör İkinci Patolog Değerlendirmesi (KY-2)

KY-1 ve YZ arasında uyumsuzluk olan dokuz olgu aynı patoloj tarafından kör olarak yeniden değerlendirildi. Bu alt kümede KY-2, 2 olguyu negatif (%22,2), 6 olguyu düşük pozitif (%66,7) ve 1 olguyu yüksek pozitif (%11,1) sınıflandırdı. KY-2, dokuz olgunun yedisinde KY-1 ile aynı kategoriyi verdi; KY-1–KY-2 tam uyumu 7/9 (%77,8; %95 güven aralığı %47,3–%99,7), Cohen kappa 0,784 ve karesel ağırlıklı kappa 0,846 idi. KY-2–YZ tam uyumu 2/9 (%22,2; %95 güven aralığı %0,3–%52,7), Cohen kappa -0,366 ve karesel ağırlıklı kappa -0,290 bulundu. Üç değerlendirmenin aynı kategoride birleştiği olgu yoktu; Fleiss kappa -0,091 idi. KY-2 yedi olguda KY-1'i, bir olguda YZ'yi yineledi (Tablo IV ve Tablo V).

Tablo II: Arşiv kayıtlarındaki nihai PD-L1 TPS dağılımı ve histolojik gruplara göre yüksek ekspresyon.

A. Nihai TPS kategorileri			
Nihai TPS kategorisi	n	%	
Negatif (<%1)	14	56,0	
Düşük pozitif (%1–49)	5	20,0	
Yüksek pozitif (≥%50)	6	24,0	
B. Histolojik gruplara göre yüksek ekspresyon			
Histolojik grup	Toplam n	Yüksek PD-L1, n (%)	Yüksek değil, n (%)
Nonmüsinöz adenokarsinom	8	3 (37,5)	5 (62,5)
Skumöz hücreli karsinom	12	3 (25,0)	9 (75,0)
Müsinöz adenokarsinom	2	0 (0,0)	2 (100,0)
Sınıflandırılmayan KHDAK	2	0 (0,0)	2 (100,0)
Büyük hücreli nöroendokrin karsinom	1	0 (0,0)	1 (100,0)

Yüksek ekspresyon TPS $\geq$ %50 olarak tanımlandı. Adenokarsinom ile skumöz hücreli karsinom karşılaştırması: Fisher kesin $p=0,617$.

KHDAK: Küçük hücreli dışı akciğer karsinomu, PD-L1: Programlanmış ölüm ligandı 1, TPS: tümör proporsiyon skoru.

TARTIŞMA

Bu tek merkezli validasyon çalışmasının temel bulgusu, KY-1 ile Gemini 3.1 Pro YZ modeli arasında klinik olarak kullanılan üç TPS kategorisinde tam uyumun %64,0 olmasıdır. Cohen kappa 0,377, karesel ağırlıklı kappa 0,531 bulundu. Sıralı kategoriler arasındaki uzaklık dikkate alındığında uyum artmakla birlikte, dokuz olguda kategori farklılığı devam etti. Uyuşmazlıklar tek yönlü değildi; YZ beş olguda daha yüksek, dört olguda daha düşük kategori verdi.

PD-L1 eşikleri klinik açıdan önemlidir. KEYNOTE-024, KEYNOTE-010 ve KEYNOTE-042 çalışmaları farklı TPS düzeyleriyle tanımlanan tedavi popülasyonlarını; KEYNOTE-189 ve KEYNOTE-407 ise nonskuamöz ve skuamöz KHDAK'ta kemoterapi-immünoterapi kombinasyonlarının değerini ortaya koymuştur (4-8). Bu nedenle bir olgunun <%1'den %1-49'a

veya %1-49'dan $\geq$ %50'ye taşınması yalnız sayısal bir fark değildir; klinik sınıflandırmayı etkileyebilecek bir kategori değişimidir. Çalışmamızda YZ, %1 eşğinde KY-1'den yalnız bir, %50 eşğinde iki fazla olguyu pozitif sınıflandırdı; ancak eşleştirilmiş marjinal farklar anlamlı değildi.

Patologlar arasında 22C3 klonu ile yapılan çalışmalarda iyi ya da belirgin uyum bildirilebilmekle birlikte, tekrarlanabilirliğin örnek türü, eşik ve eğitim bağlamına duyarlı olduğu gösterilmiştir (13,14,16). Çalışmamızdaki kappa değerleri bu yayınlarla doğrudan karşılaştırılmaz; çünkü burada ikinci değerlendirici bir patolog değil, seçilmiş mikroskopik fotoğrafları Türkçe istemlerle yorumlayan bir YZ modelidir. Sonuçlar, YZ çıktılarının yerel validasyon ve kalibrasyon yapılmadan değişmez bir ölçüm olarak kabul edilmemesi gerektiğini göstermektedir.

Tablo III: Patolog değerlendirmeleri (KY-1 ve KY-2) ile Yapay Zekâ (YZ) sonuçlarının karşılaştırılması.

A. Değerlendiriciye göre kategori dağılımı.				
Değerlendirici	Negatif (<%1)	Düşük pozitif (%1-49)	Yüksek pozitif ($\geq$ %50)	
KY-1	15 (%60,0)	6 (%24,0)	4 (%16,0)	
YZ (Gemini 3.1 Pro)	14 (%56,0)	5 (%20,0)	6 (%24,0)	
KY-2 (n=9)	2 (%22,2)	6 (%66,7)	1 (%11,1)	
B. Üç kategorili KY-1-YZ çapraz tablosu (Satırlar KY-1, sütunlar YZ sonuçlarını göstermektedir).				
KY-1 / YZ	Negatif	Düşük pozitif	Yüksek pozitif	Toplam
Negatif	11	2	2	15
Düşük pozitif	3	2	1	6
Yüksek pozitif	0	1	3	4
Toplam	14	5	6	25

KY-1 patologun ilk değerlendirmesini, KY-2 uyumsuz olgulardaki kör ikinci değerlendirmeyi ve YZ Gemini 3.1 Pro değerlendirmesini temsil eder. KY-2 yüzdeleri yalnız KY-2 verisi bulunan sekiz uyumsuz olgu üzerinden hesaplanmıştır.

Tablo IV: KY-1-YZ birincil uyum analizleri ve KY-2 keşifsel sonuçları.

Analiz	n	Ham uyum, % (%95 GA)	Kappa	Ağırlıklı kappa	McNemar p
KY-1-YZ, üç kategori	25	64,0 (42,5-82,0)	0,377 (0,072-0,682)	0,531 (0,196-0,866)	-
Pozitiflik eşiği: $\geq$ %1	25	72,0 (50,6-87,9)	0,426	-	1,000
Yüksek pozitiflik eşiği: $\geq$ %50	25	84,0 (63,9-95,5)	0,505	-	0,625
KY-1-KY-2, üç kategori	9	77,8 (47,3-99,7)	0,784	0,846	-
KY-2-YZ, üç kategori	9	11,1 (0,3-52,7)	-0,366	-0,290	-
KY-1/KY-2/YZ, üç değerlendirici	9	0,0 (0,0-36,9)	-0,091*	-	-

Ağırlıklı kappa üç sıralı kategori için karesel ağırlıkla hesaplandı. Üç değerlendirici satırdaki değer Fleiss kappa'dır. Eşleştirilmiş p değerleri kesin McNemar testine aittir. Ham uyum güven aralıkları Clopper-Pearson kesin yöntemiyle hesaplandı. KY-2 analizleri, uyumsuzluğa göre seçilmiş sekiz olgudan oluştuğu için keşifselidir.

GA: Güven aralığı; KY-1: Patologun ilk değerlendirmesi; KY-2: Kör ikinci patolog değerlendirmesi; YZ: Yapay zekâ.

Tablo V: Olgu düzeyinde KY-1, YZ ve mevcut KY-2 TPS sonuçları ile kategori uyumu.

Olgu no*	Kaynak notu / özellik	KY-1 TPS	YZ TPS	KY-2 TPS	KY-1-YZ durumu
1	—	<%1 (N)	%2 (DP)	<%1 (N)	YZ
2	—	%10 (DP)	%20 (DP)	—	Uyum
3	İç kontrol negatif	<%1 (N)	<%1 (N)	—	Uyum
4	—	%100 (YP)	%95 (YP)	—	Uyum
5	—	%20 (DP)	<%1 (N)	%10–20 (DP)	YZ
6	—	<%1 (N)	<%1 (N)	—	Uyum
7	Müsinöz	<%1 (N)	<%1 (N)	—	Uyum
8	—	<%1 (N)	<%1 (N)	—	Uyum
9	—	%100 (YP)	%95 (YP)	—	Uyum
10	Hücre bloğu?	%70 (YP)	%10 (DP)	%70–80 (YP)	YZ
11	—	<%1 (N)	<%1 (N)	—	Uyum
12	—	<%1 (N)	%5 (DP)	%1 (DP)	YZ
13	—	%10 (DP)	<%1 (N)	%20 (DP)	YZ
14	—	<%1 (N)	%95 (YP)	%5	YZ
15	—	%100 (YP)	%95 (YP)	—	Uyum
16	—	<%1 (N)	%90 (YP)	<%1 (N)	YZ
17	—	%10 (DP)	%90 (YP)	%10–20 (DP)	YZ
18	—	%10 (DP)	%10 (DP)	—	Uyum
19	—	<%1 (N)	<%1 (N)	—	Uyum
20	—	<%1 (N)	<%1 (N)	—	Uyum
21	—	<%1 (N)	<%1 (N)	—	Uyum
22	İkinci %5	<%1 (N)	<%1 (N)	—	Uyum
23	İkinci <%1	<%1 (N)	<%1 (N)	—	Uyum
24	—	%10 (DP)	<%1 (N)	%20 (DP)	YZ
25	—	<%1 (N)	<%1 (N)	—	Uyum

Olgu numaraları, Excel kayıt sırasına göre preparatı bulunmayan bir kayıt dışlandıktan sonra verilmiştir. TPS değerleri, çalışma sayfasında görüntülenen yüzde biçimleri esas alınarak aktarılmıştır. Kaynak veri tablosunda olgu düzeyinde yaş, cinsiyet ve histolojik alt tip eşleştirme bulunmadığından yalnız mevcut not/özellik alanı gösterilmiştir. N: negatif (<%1); DP: düşük pozitif (%1–49); YP: yüksek pozitif (≥%50); YZ: YZ'nin KY-1'den daha yüksek kategori vermesi; YZ: YZ'nin KY-1'den daha düşük kategori vermesi; —: veri yok.

Blueprint ve yöntem karşılaştırma projeleri, PD-L1 testinde antikör klonu, platform ve analitik bağlamın önemini göstermiştir (10-12,15). Benzer biçimde YZ değerlendirmesinde de yalnız model adını bildirmek yeterli değildir. Bu çalışmada kesit kalınlığı, antijen geri kazanımı, antikör inkübasyonu, saptama sistemi, dış kontrol, görüntü sayısı, alan seçimi, büyüme, çözünürlük, istem çalıştırma sayısı ve çıktı kuralı standartlaştırılarak ayrıntılı biçimde tanımlandı. Model sürümü veya görüntüleme koşulları değiştiğinde yeniden validasyon yapılması gerekir.

YZ'nin KY-1'den daha yüksek kategori verdiği olgular yeniden gözden geçirildiğinde, özellikle bazı yüksek pozitif YZ değerlendirmelerinde alveoler makrofaj yoğunluğunun belirgin olduğu görüldü. TPS hesabına immün hücreler katılmamasına rağmen, seçilmiş mikroskopik fotoğraflarda makrofaj boyanmasının tümör hücresi membran boyanmasıyla karışması YZ skorunu artırmış olabilir. Alan seçimi ve görüntü kalitesi de karşılaştırmayı etkileyen unsurlar olabilir. Bununla birlikte YZ dört olguda KY-1'den daha düşük kategori verdiğinden, makrofaj karışıklığı bütün uyumsuzlukları açıklayan tek bir mekanizma olarak kabul edilemez.

Çift yönlü uyumsuzluklar; tümör heterojenitesi, seçilmiş alanların temsiliyeti, sınır değere yakın görsel tahminler, immün hücre-tümör hücresi ayrımı veya hem patolog hem YZ kaynaklı sınıflandırma değişkenliği ile ilişkili olabilir. Bağımsız, kör ve çok okuyuculu bir konsensüs standardı bulunmadığından bu olasılıkların hangisinin baskın olduğu belirlenemez. Gelecek çalışmalarda her olgunun en az iki deneyimli torasik patolog tarafından bağımsız değerlendirilmesi, uyumsuzlukların üçüncü patolog ya da panel tarafından karara bağlanması ve YZ çıktısının bu referans standardına kör tutulması gerekir.

Kör ikinci patolog değerlendirmesi, seçilmiş uyumsuz olgularda önemli ek bilgi sağladı. KY-2, dokuz olgunun yedisinde KY-1 kategorisini yineledi ve KY-1-KY-2 ağırlıklı kappa 0,846 idi; bir olguda bir üst kategoriye çıktı, yalnız bir olguda YZ ile aynı kategoriye ulaştı. Bu örüntü, ilk patolog değerlendirmesinin söz konusu alt kümede tekrarlanabilir olduğunu düşündürmektedir. Ancak KY-2 yalnız uyumsuzluklara göre seçilmiş dokuz olguda bulunduğundan, bu değerler tüm kohorta ait intraobserver uyumu temsil etmez ve YZ'nin hatalı olduğunu tek başına kanıtlamaz.

Kohortta yüksek ekspresyon hem nonmüsinöz adenokarsinom hem skuamöz hücreli karsinom grubunda görüldü; diğer histolojik alt gruplarda yüksek ekspresyon saptanmadı. Bununla birlikte küçük alt gruplar bir ile iki olgudan oluştuğundan bu dağılımlar yalnız betimleyici ve hipotez üretici kabul edilmelidir. Çalışmanın güçlü yönleri, tam eşleştirilmiş 25 olguluk kohortun kullanılması, klinik eşiklerin ayrı değerlendirilmesi, uyumsuzluk yönünün açık biçimde gösterilmesi, sekiz uyumsuz olguda kör ikinci patolog değerlendirmesinin bulunması ve olgu düzeyindeki sonuçların ayrıntılı olarak sunulmasıdır.

Sınırlılıklar

Çalışmanın başlıca sınırlılıkları örneklemin küçük olması ve çalışmanın tek merkezli, retrospektif tasarımıdır. Araştırmaya planlanan süre içinde yalnızca 25 olgu dahil edilebilmiştir. Bu durum kappa tahminlerinde geniş güven aralıklarına neden olmuştur. Yalnız bir patolog ile bir YZ modeli karşılaştırılmış ve bağımsız konsensüs referans standardı oluşturulmamıştır; bu nedenle duyarlılık, özgüllük veya doğruluk hesaplanamaz. Patolog tarafından ikinci kör değerlendirme (KY-2) tüm olgulara uygulanmadığından genel intraobserver tekrarlanabilirlik olarak değerlendirilemez. YZ değerlendirmesi lamdaki kesitin tüm yüzeyinin taranmış görüntülerine değil, akıllı telefonla kaydedilmiş seçilmiş mikroskopik fotoğraflara dayandığından alan seçimi sonuçları etkilemiş olabilir. Her olgu için istemin yalnız bir kez çalıştırılması modelin yanıt tekrarlanabilirliğinin ölçülmesine izin vermemiştir. Evre, sigara öyküsü, moleküler değişiklikler, tedavi, yanıt ve sağkalım verileri bulunmadığından prediktif veya prognostik çıkarım yapılamaz.

SONUÇ

Bu 25 olguluk KHDAK validasyonunda KY-1 ile Gemini 3.1 Pro YZ modeli arasında PD-L1 TPS kategorileri bakımından tam uyum %64,0; Cohen kappa 0,377 ve karesel ağırlıklı kappa 0,531 bulundu. Uyumsuzluklar çift yönlüydü ve %1 ile %50 klinik eşiklerinde eşleştirilmiş marjinal fark saptanmadı. Kör KY-2 değerlendirmesi, mevcut sekiz uyumsuz olgunun yedisinde KY-1 sonucunu yineledi. Bununla birlikte bağımsız ve çok okuyuculu konsensüs referans standardı bulunmadığından bu örüntü YZ'nin doğruluğunu ya da hatasını kesin olarak göstermez. Klinik uygulama için yerel kalibrasyon, model sürümünün sabitlenmesi, standartlaştırılmış görüntüleme ve istem protokolü, patolog denetimi ve kör uzlaşısı değerlendirilmesi gereklidir.

Gelecek çalışmalarda YZ değerlendirmesinin birden çok kez çalıştırılarak tekrarlanabilirliğinin gözden geçirilmesi, çalışmanın tümü taranmış lamalarda gerçekleştirilmesi, olgu sayısının daha fazla olması, çalışmanın birden çok patolog tarafından kör değerlendirme yapılarak yürütülmesi, sınırlılıkları azaltacaktır.

Teşekkür

Patoloji arşivinin düzenlenmesi ve immünohistokimyasal preparatların değerlendirmeye hazırlanmasındaki katkıları için Atatürk Üniversitesi Tıp Fakültesi Tıbbi Patoloji teknik ekibine teşekkür ederiz.

Bildirimler

Etik Kurul Onayı: Çalışma, Atatürk Üniversitesi Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu tarafından 21.02.2024 tarih ve 1/92 sayılı kararla onaylandı. Bilgilendirilmiş onam alınması, retrospektif ve anonimleştirilmiş tasarım nedeniyle Etik Kurul tarafından gerekli bulunmadı.

Araştırma Desteği: Bu çalışma için finansal destek beyan edilmemiştir.

Veri Güvenliği ve Verilerin Sorgulanması: Anonimleştirilmiş veri seti ve analiz kodu, etik ve kurumsal kısıtlar çerçevesinde makul talep üzerine sorumlu yazardan sağlanabilir.

Yapay Zekâ Kullanımı: Gemini 3.1 Pro, araştırma kapsamında YZ değerlendirmesi olarak kullanılmıştır. Makalenin dil, anlatım ve biçim düzenlenmesinde OpenAI ChatGPT'den destek alınmıştır. Tüm istatistiksel sonuçlar, bilimsel yorumlar ve nihai metin yazarlar tarafından doğrulanmış ve onaylanmıştır.

Çıkar Çatışması: Yazarlar bu çalışma ile ilgili çıkar çatışmaları olmadığını beyan etmişlerdir.

YAZAR KATKILARI

Çalışmanın fikri veya tasarımı: SÖ
Veri toplama: SÖ, YH, RD, BE, NB, ÖB, SD, ÖA, SBT
Veri analizi ve yorumlama: KY, BM, AA
Makale taslağının hazırlanması: BM
Makalenin kritik revizyonu: KY
Tüm yazarlar sonuçları gözden geçirmiş ve makalenin son hâlini onaylamıştır.

KAYNAKLAR

- Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;74(3):229-263. PMID: 38572751. doi:10.3322/caac.21834
- Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, Dacic S, Jain D, Kerr KM, Lantuejoul S, Noguchi M, Papotti M, Rekhtman N, Scagliotti G, van Schil P, Sholl L, Yatabe Y, Yoshida A, Travis WD. The 2021 WHO Classification of Lung Tumors: impact of advances since 2015. *J Thorac Oncol.* 2022;17(3):362-387. PMID: 34808341. doi:10.1016/j.jtho.2021.11.003
- WHO Classification of Tumours Editorial Board. Thoracic Tumours. 5th ed. Lyon, France: International Agency for Research on Cancer; 2021.
- Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Cssi T, Fülöp A, Gottfried M, Peled N, Tafreshi A, Cuffe S, O'Brien M, Rao S, Hotta K, Leiby MA, Lubiniecki GM, Shentu Y, Rangwala R, Brahmer JR; KEYNOTE-024 Investigators. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med.* 2016;375(19):1823-1833. PMID: 27718847. doi:10.1056/NEJMoa1606774
- Herbst RS, Baas P, Kim DW, Felip E, Pérez-Gracia JL, Han JY, Molina J, Kim JH, Arvis CD, Ahn MJ, Majem M, Fidler MJ, de Castro G Jr, Garrido M, Lubiniecki GM, Shentu Y, Im E, Dolled-Filhart M, Garon EB. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet.* 2016;387(10027):1540-1550. PMID: 26712084. doi:10.1016/S0140-6736(15)01281-7
- Mok TSK, Wu YL, Kudaba I, Kowalski DM, Cho BC, Turna HZ, Castro G Jr, Srimuninnimit V, Laktionov KK, Bondarenko I, Kubota K, Lubiniecki GM, Zhang J, Kush D, Lopes G; KEYNOTE-042 Investigators. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet.* 2019;393(10183):1819-1830. PMID: 30955977. doi:10.1016/S0140-6736(18)32409-7
- Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, Domine M, Clingan P, Hochmair MJ, Powell SF, Cheng SY, Bischoff HG, Peled N, Grossi F, Jennens RR, Reck M, Hui R, Garon EB, Boyer M, Rubio-Viqueira B, Novello S, Kurata T, Gray JE, Vida J, Wei Z, Yang J, Raftopoulos H, Pietanza MC, Garassino MC; KEYNOTE-189 Investigators. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med.* 2018;378(22):2078-2092. PMID: 30207917. doi:10.1056/NEJMoa1801005
- Paz-Ares L, Luft A, Vicente D, Tafreshi A, Gümüş M, Mazières J, Hermes B, Çay Şenler F, Cssi T, Fülöp A, Rodríguez-Cid J, Wilson J, Sugawara S, Kato T, Lee KH, Cheng Y, Novello S, Halmos B, Li X, Lubiniecki GM, Piperdi B, Kowalski DM; KEYNOTE-407 Investigators. Pembrolizumab plus chemotherapy for squamous non-small-cell lung cancer. *N Engl J Med.* 2018;379(21):2040-2051. PMID: 30280635 doi: 10.1056/NEJMoa1810865.
- Agilent Technologies. PD-L1 IHC 22C3 pharmDx Interpretation Manual: Non-Small Cell Lung Cancer. Agilent Technologies. Accessed July 5, 2026. https://www.agilent.com/cs/library/usermanuals/public/29158_pd-l1-ihc-22C3-pharmdx-nslc-interpretation-manual.pdf
- Hirsch FR, McElhinny A, Stanforth D, Ranger-Moore J, Jansson M, Kulangara K, Richardson W, Towne P, Hanks D, Vennapusa B, Mistry A, Kalamegham R, Averbuch S, Novotny J, Rubin E, Emancipator K, McCaffery I, Williams JA, Walker J, Longshore J, Tsao MS, Kerr KM. PD-L1 immunohistochemistry assays for lung cancer: results from phase 1 of the Blueprint PD-L1 IHC Assay Comparison Project. *J Thorac Oncol.* 2017;12(2):208-222. PMID: 27913228. doi:10.1016/j.jtho.2016.11.2228
- Tsao MS, Kerr KM, Kockx M, Beasley MB, Borczuk AC, Botling J, Bubendorf L, Chirieac L, Chen G, Chou TY, Chung JH, Dacic S, Lantuejoul S, Mino-Kenudson M, Moreira AL, Nicholson AG, Noguchi M, Pelosi G, Poleri C, Russell PA, Sauter J, Thunnissen E, Wistuba I, Yu H, Wynes MW, Pintilie M, Yatabe Y, Hirsch FR. PD-L1 immunohistochemistry comparability study in real-life clinical samples: results of Blueprint Phase 2 Project. *J Thorac Oncol.* 2018;13(9):1302-1311. PMID: 29800747. doi:10.1016/j.jtho.2018.05.013
- Hendry S, Byrne DJ, Wright GM, Young RJ, Sturrock S, Cooper WA, Fox SB. Comparison of four PD-L1 immunohistochemical assays in lung cancer. *J Thorac Oncol.* 2018;13(3):367-376. PMID: 29175115. doi:10.1016/j.jtho.2017.11.112
- Cooper WA, Russell PA, Cherian M, Duhig EE, Godbolt D, Jessup PJ, Khoo C, Leslie C, Mahar A, Moffat DF, Sivasubramanian V, Faure C, Reznichenko A, Grattan A, Fox SB. Intra- and interobserver reproducibility assessment of PD-L1 biomarker in non-small cell lung cancer. *Clin Cancer Res.* 2017;23(16):4569-4577. PMID: 28420726. doi:10.1158/1078-0432.CCR-17-0151
- Chang S, Park HK, Choi YL, Jang SJ. Interobserver reproducibility of PD-L1 biomarker in non-small cell lung cancer: a multi-institutional study by 27 pathologists. *J Pathol Transl Med.* 2019;53(6):347-353. PMID: 31656061. doi:10.4132/jptm.2019.09.29
- Butter R, 't Hart NA, Hooijer GKJ, Monkhorst K, Speel EJ, Theunissen P, Thunnissen E, Von der Thüsen JH, Timens W, van de Vijver MJ. Multicentre study on the consistency of PD-L1 immunohistochemistry as predictive test for immunotherapy in non-small cell lung cancer. *J Clin Pathol.* 2020;73(7):423-430. PMID: 31822512. doi:10.1136/jclinpath-2019-205993
- Yuan P, Guo C, Lin L, Guo L, Zhang F, Ying J. The reproducibility of histopathologic assessments of programmed cell death-ligand 1 using companion diagnostics in NSCLC. *JTO Clin Res Rep.* 2021;2(2):100102. PMID: 34589980. doi:10.1016/j.jtocrr.2020.100102
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174. PMID: 843571. doi:10.2307/2529310
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet.* 2007;370(9596):1453-1457. PMID: 25046131. doi:10.1016/S0140-6736(07)61602-X