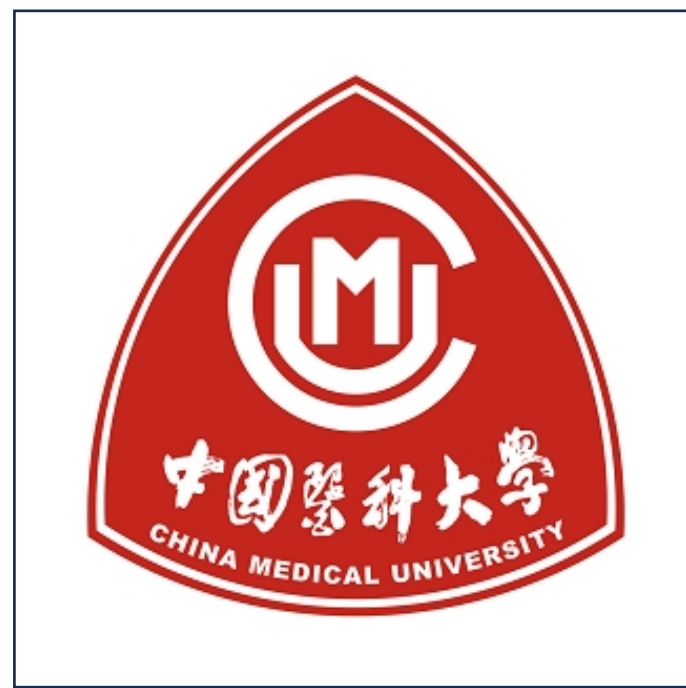


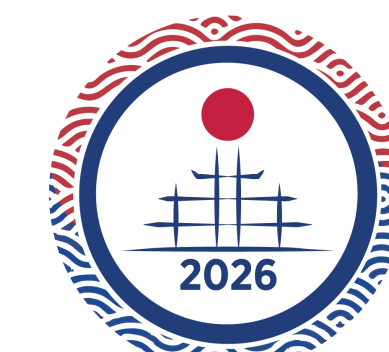


P2.131: Multi-Center Pathologist Consistency and Artificial Intelligence Performance of PD-L1 Scoring in NSCLC

Liang Wang¹, Shujun Xu¹, Xiaojing Lv¹, Fang Han¹, Jian Wang¹

¹Dept. of Pathology, The 1st Hospital and College of Basic Medical Sciences, China Medical University

Correspondence: lwang@cmu.edu.cn



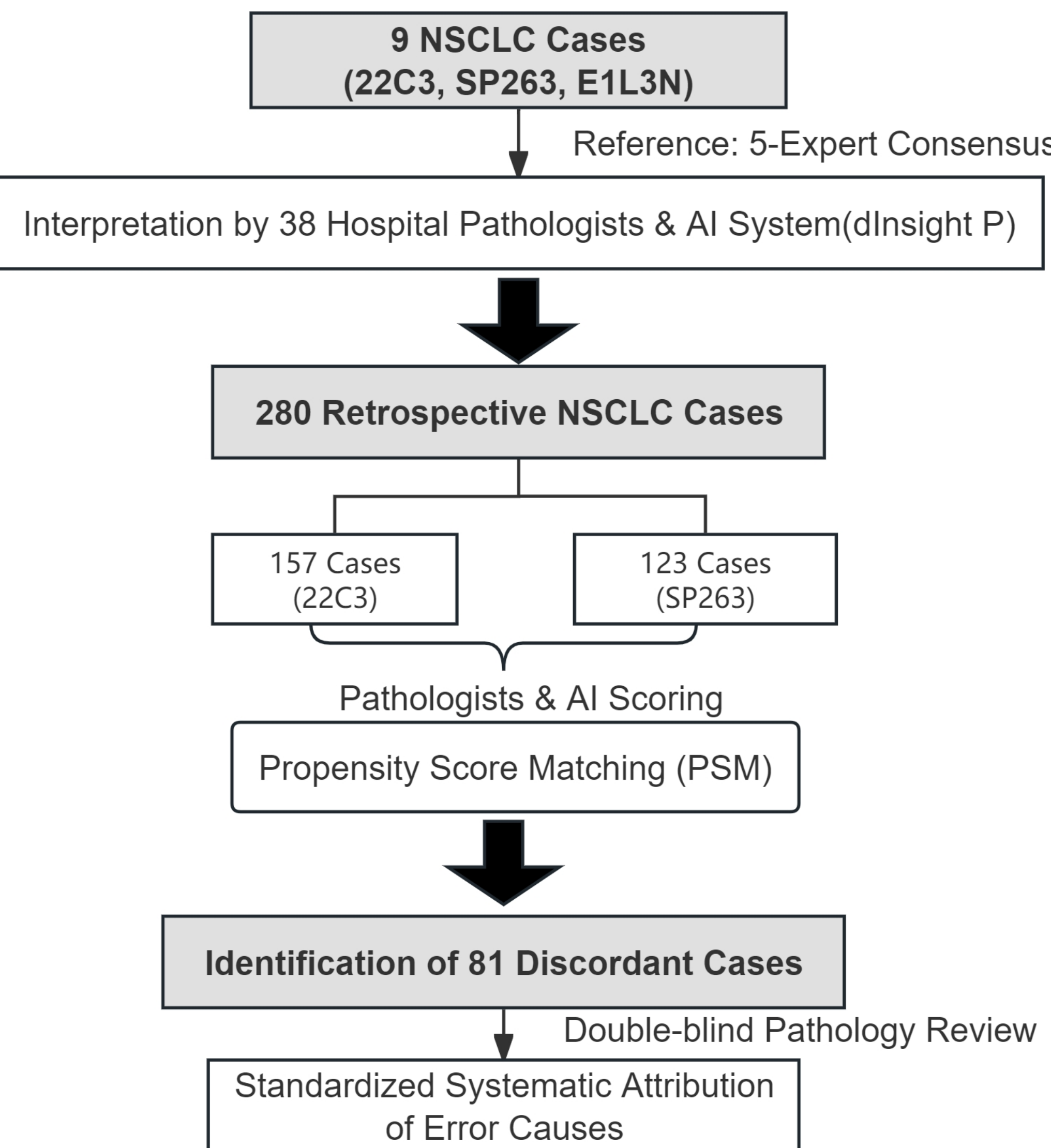
IASLC 2026 World
Conference on Lung Cancer

SEPTEMBER 12 - 15, 2026
SEOUL, REPUBLIC OF KOREA

INTRODUCTION

Accurate evaluation of programmed death-ligand 1 (PD-L1) expression in non-small cell lung cancer (NSCLC) is critical for immunotherapy guidance but remains limited by inter-observer variability. We evaluated inter-institutional agreement among pathologists and assessed the concordance of a National Medical Products Administration (NMPA)-approved artificial intelligence (AI) scoring system, dInsight P, against experienced pathologists across different PD-L1 antibody clones. The study utilized a three-phase design involving pathologists from 38 hospitals to interpret tumor proportion scores (TPS) on cases stained with 22C3, SP263, and E1L3N antibodies, comparing results against an expert consensus standard. Subsequently, the deep learning scoring system was validated in a retrospective cohort of 280 propensity score-matched cases, utilizing a standardized framework to attribute specific causes of discordance. While agreement between participating hospitals and the expert reference standard was almost perfect (intraclass correlation coefficient [ICC] = 0.996), the AI system demonstrated high correlation for continuous TPS scoring (ICC = 0.83) but only moderate agreement for categorical clinical thresholds (Kappa = 0.53). Distinct antibody-specific error patterns were observed: the SP263 assay was associated with AI overestimation due to non-tumor cell misclassification, whereas the 22C3 assay was linked to underestimation caused by tumor cell oversight. Furthermore, specimen type was a significant predictor of AI errors. In conclusion, although human agreement is high within a provincial pathology society, AI performance is heavily influenced by antibody-specific staining characteristics, suggesting that clone-specific calibration is essential to mitigate overestimation risks in SP263 and underestimation in 22C3 assays when scores are close to critical clinical thresholds.

METHODS



DISCUSSION AND CONCLUSION

Table 1. Univariate analysis of clinicopathological factors associated with concordance between AI and pathologist PD-L1 interpretation.

Variables	Total (n = 280)	Discordant (n = 81)	Concordant (n = 199)	P value
Antibody clone, n (%)				0.080
22C3	157	52 (33.12)	105 (66.88)	
SP263	123	29 (23.58)	94 (76.42)	
Specimen type, n (%)				0.025
Biopsy	215	55 (25.58)	160 (74.42)	
Surgical resection	65	26 (40.00)	39 (60.00)	
PD-L1 expression, n (%)				<.001
< 1%	71	33 (46.48)	38 (53.52)	
1%-49%	135	21 (15.56)	114 (84.44)	
≥ 50%	74	27 (36.49)	47 (63.51)	
Histology, n (%)				0.284
Adenocarcinoma	167	53 (31.74)	114 (68.26)	
Squamous cell carcinoma	99	23 (23.23)	76 (76.77)	
Others	14	5 (35.71)	9 (64.29)	

Our study evaluates multi-center PD-L1 interpretation consistency and the real-world performance of an AI-assisted diagnostic system in NSCLC. High inter-institutional agreement is achievable within a provincial pathology society. While the AI system demonstrates reliability in continuous TPS quantification, diagnostic precision is challenged near critical clinical thresholds (1% and 50%). Notably, AI misinterpretation patterns are antibody-dependent: the SP263 assay is prone to overestimation, whereas the 22C3 assay tends toward underestimation. Furthermore, algorithmic discordance is significantly higher in surgical resections than in biopsies. These findings underscore the necessity for antibody-specific algorithm calibration. More importantly for clinical practice, these findings caution thoracic oncologists against blind reliance on automated algorithms in their current state; mandatory human verification in borderline cases remains essential to ensure accurate immunotherapy stratification and prevent clinical mis-triage.

REFERENCES & ACKNOWLEDGEMENTS

Reference

1. Siegel RL, Kratzer TB, Giaquinto AN, et al. Cancer statistics, 2025. CA: A Cancer Journal for Clinicians 2025;75:10–45. doi:10.3322/caac.21871.
2. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med 2016;375:1823–1833. doi:10.1056/NEJMoa1606774.
3. Brahmer JR, Rodríguez-Abreu D, Robinson AG, et al. Health-related quality-of-life results for pembrolizumab versus chemotherapy in advanced, PD-L1-positive NSCLC (KEYNOTE-024): a multicentre, international, randomised, open-label phase 3 trial. Lancet Oncol 2017;18:1600–1609. doi:10.1016/S1470-2045(17)30690-3.
4. Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. N Engl J Med 2012;366:2443–2454. doi:10.1056/NEJMoa1200690.
5. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Updated Analysis of KEYNOTE-024: Pembrolizumab Versus Platinum-Based Chemotherapy for Advanced Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score of 50% or Greater. J Clin Oncol 2019;37:537–546. doi:10.1200/JCO.18.00149.
6. Nosaki K, Saka H, Hosomi Y, et al. Safety and efficacy of pembrolizumab monotherapy in elderly patients with PD-L1-positive advanced non-small-cell lung cancer: Pooled analysis from the KEYNOTE-010, KEYNOTE-024, and KEYNOTE-042 studies. Lung Cancer 2019;135:188–195. doi:10.1016/j.lungcan.2019.07.004.
7. Doroshow DB, Bhalla S, Beasley MB, et al. PD-L1 as a biomarker of response to immune-checkpoint inhibitors. Nat Rev Clin Oncol 2021;18:345–362. doi:10.1038/s41571-021-00473-5.
8. Fan J, Shen X, Wang Y, et al. Biomarkers for immune checkpoint therapy targeting programmed death 1 and programmed death ligand 1. Biomed Pharmacother 2020;130:110621. doi:10.1016/j.biopha.2020.110621.
9. Tsao MS, Kerr KM, Kockx M, et al. PD-L1 Immunohistochemistry Comparability Study in Real-Life Clinical Samples: Results of Blueprint Phase 2 Project. J Thorac Oncol 2018;13:1302–1311. doi:10.1016/j.jtho.2018.05.013.

Artificial Intelligence (AI) Use Disclaimer

No potential conflict of interest was reported by the author(s). Specifically, the authors declare no financial or non-financial conflicts of interest regarding the AI software providers (Hangzhou Dyingjia Technology Co., Ltd.) or the scanner manufacturers (HaideXing Technology Co., Ltd. / Motic).

RESULTS

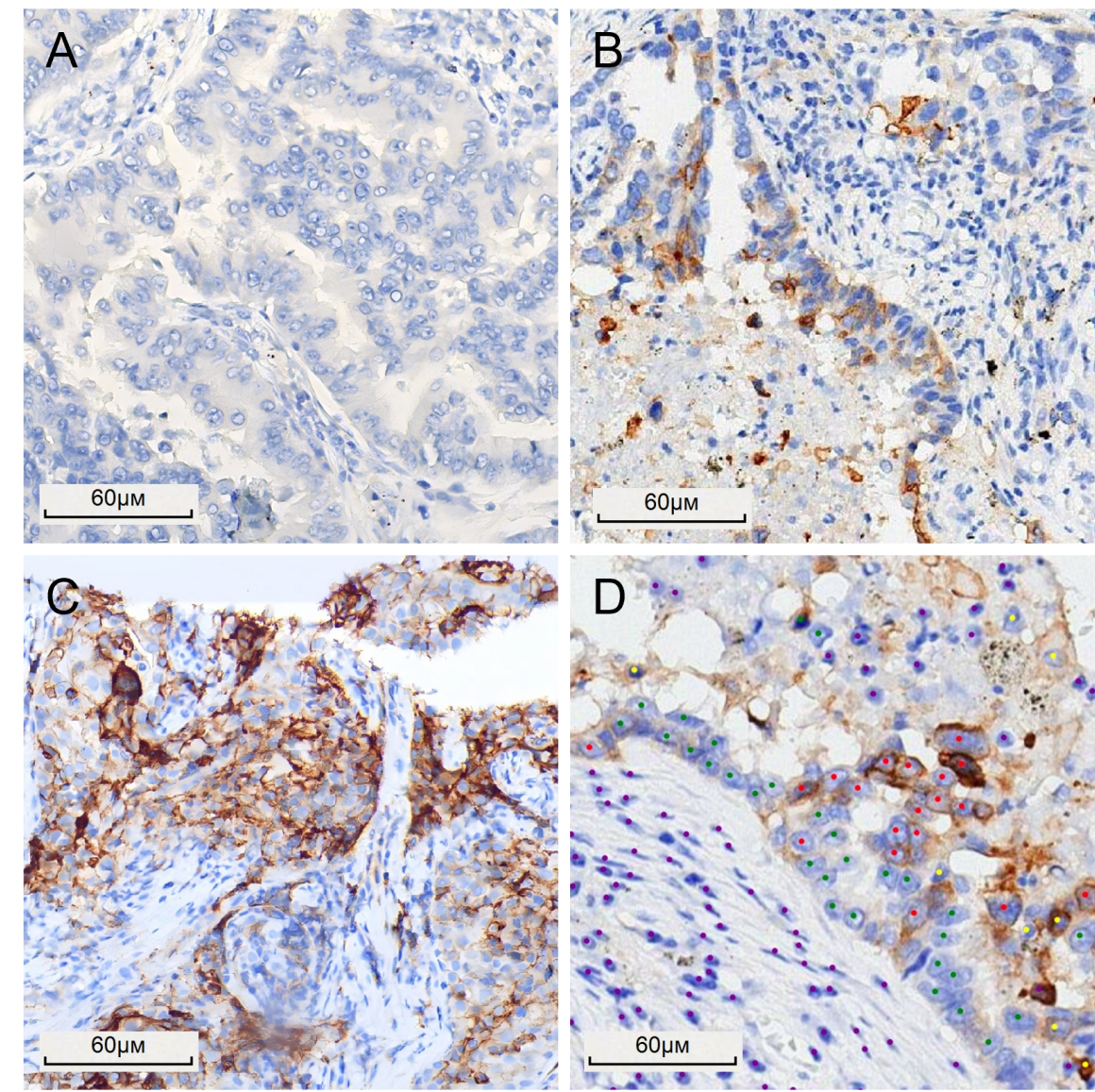


Figure 1. Representative examples of whole slide images (WSI)-based PD-L1 evaluation. (A-C) Manual scoring by pathologist stratified by TPS category: (negative, low, and high). (D) AI-based interpretation visualization: red, positive tumor cells; green, negative tumor cells; yellow, positive non-tumor cells; purple, negative non-tumor cells.

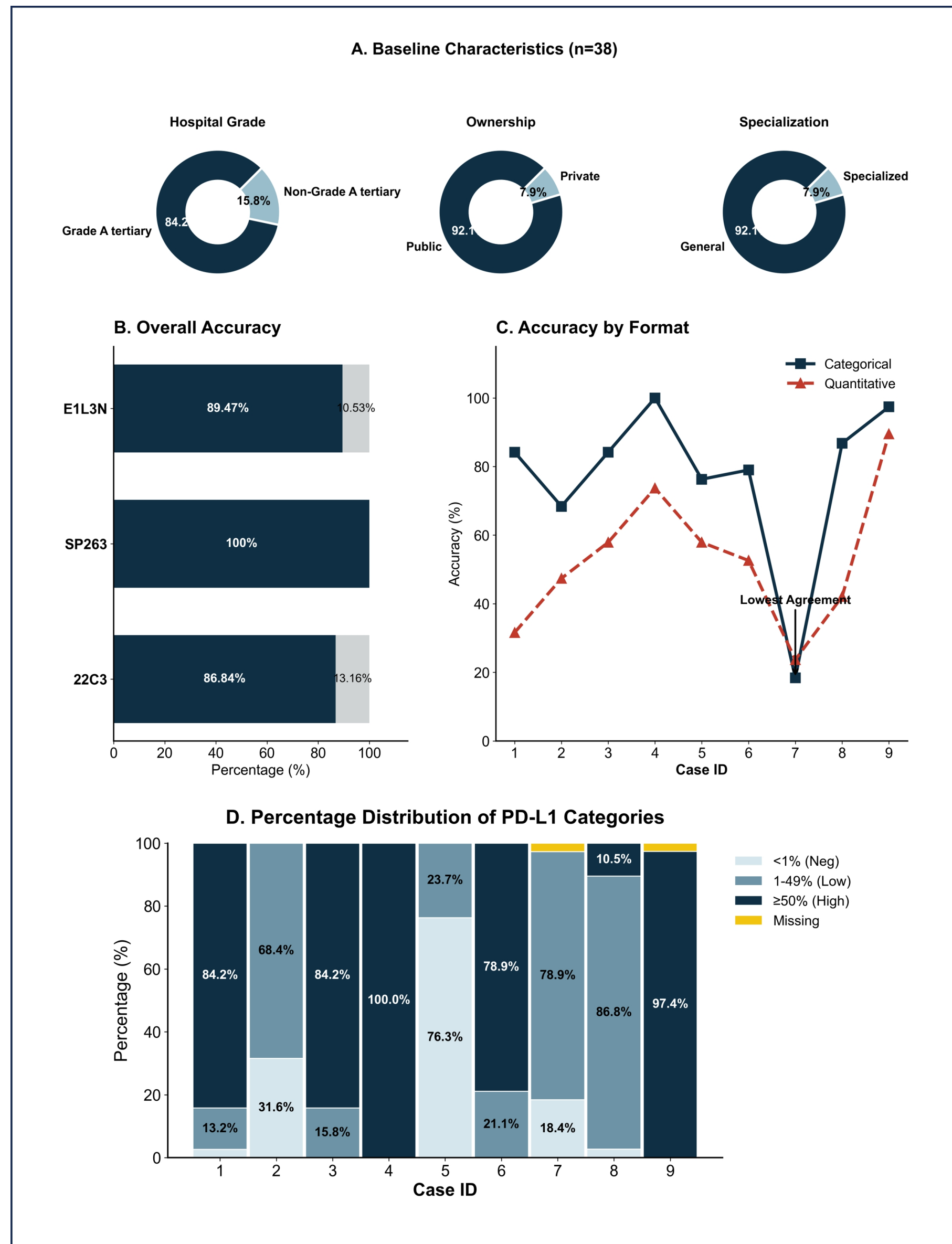


Figure 2. A. Baseline characteristics of participating hospitals. B. Overall performance of PD-L1 interpretation. C. Accuracy of the two assessment formats. D. Interpretation results of 9 Cases by 38 Hospitals.

	human reading				human reading				human reading			
	categ.1	categ.2	categ.3	sum	categ.1	categ.2	categ.3	sum	categ.1	categ.2	categ.3	sum
AI reading	38	17	1	56	37	16	1	54	1	1	0	2
categ.1	33	114	26	173	13	49	19	81	20	65	7	92
categ.2	0	4	47	51	0	3	19	22	0	1	28	29
categ.3	71	135	74	280	50	68	39	157	21	67	35	123
sum												

Figure 3. Confusion matrices illustrating the categorical concordance between AI and pathologist interpretation before propensity score matching (PSM) across three clinical categories (<1%, 1-49%, and ≥50%)

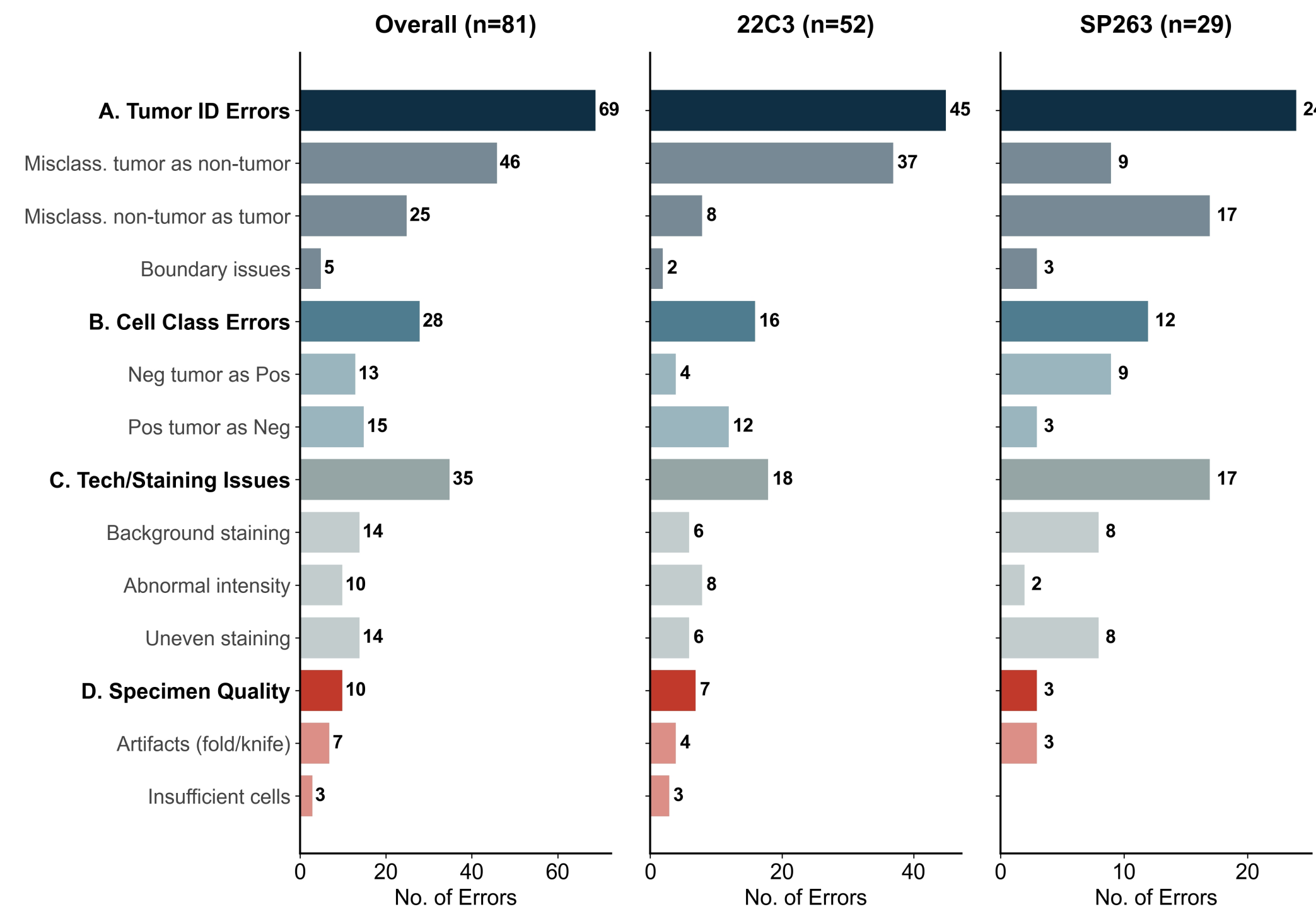


Figure 5. Distribution of specific causes for discordance between AI and pathologist PD-L1 interpretation stratified.

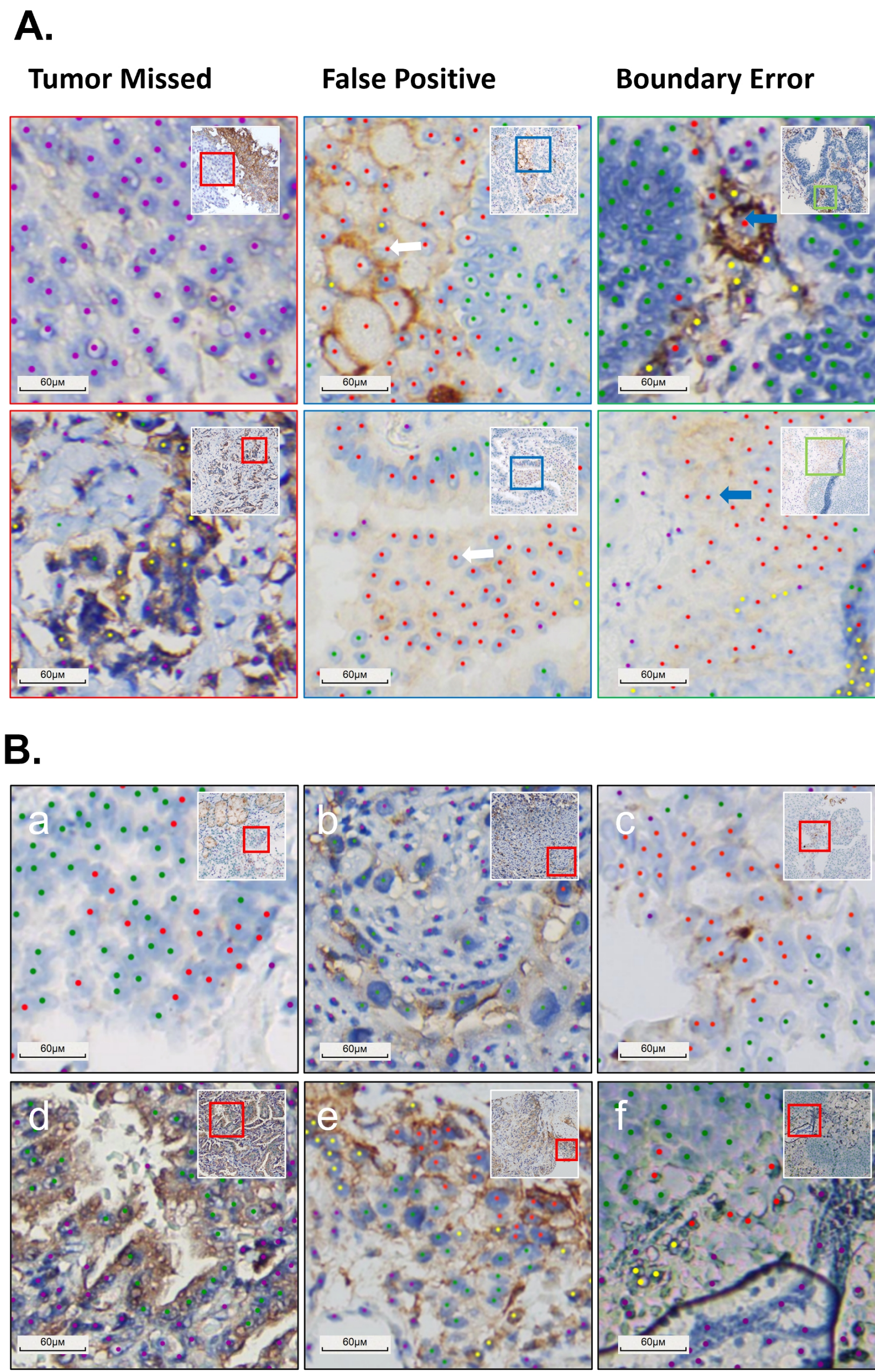


Figure 6. Visualization of AI-assisted cell identification and classification errors. A. Examples of AI tumor identification errors. The left column (Tumor Missed) shows the AI system misclassifying tumor cells as non-tumor cells (indicated by purple and yellow dots). The middle column (False Positive) shows the misclassification of histiocytes as tumor cells (white arrows). The right column (Boundary Error) demonstrates the failure to accurately recognize tumor boundaries, where tumor-associated immune cells were misclassified as tumor cells (blue arrows). B. Examples of cell classification errors. (a) misinterpretation of negative tumor cells as positive (red dots); (b) misinterpretation of positive tumor cells as negative (green dots); (c) non-specific background staining misclassified as positive tumor cells (red dots); (d) excessive staining intensity causing the AI to fail in recognizing positive tumor cells (green dots); (e) uneven staining intensity leading to the misclassification of positive tumor cells as negative (green dots); and (f) specimen quality issues (e.g., air bubbles) leading to the misclassification of optical artifacts as positive tumor cells (red dots).