


RESEARCH ARTICLE OPEN ACCESS

GloPath: An Entity-Centric Foundation Model for Glomerular Lesion Assessment and Clinicopathological Insights

Qiming He^{1,2,3} | Jing Li⁴ | Tian Guan² | Yifei Ma⁵ | Zimo Zhao⁵ | Yanxia Wang⁴ | Hongjing Chen⁶ | Yingming Xu² | Shuang Ge^{2,7} | Yexing Zhang⁸ | Yizhi Wang² | Xinrui Chen² | Lianghui Zhu² | Yiqing Liu² | Qingxia Hou⁹ | Shuyan Zhao⁹ | Xiaoqin Wang⁹ | Lili Ma⁹ | Peizhen Hu⁴ | Qiang Huang^{2,3} | Zihan Wang² | Zhiyuan Shen² | Junru Cheng² | Siqi Zeng² | Jiurun Chen^{2,7} | Zhen Song⁷ | Chao He⁵ | Zhe Wang⁴ | Yonghong He² 

¹Interdisciplinary Institute for Medical Engineering, Fuzhou University, Fuzhou, China | ²Institute of Biopharmaceutical and Health Engineering, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China | ³Medical Optical Technology R&D Center, Research Institute of Tsinghua, Guangzhou, China | ⁴Department of Pathology, School of Basic Medicine and Xijing Hospital, Fourth Military Medical University, Xi'an, China | ⁵Department of Engineering Science, University of Oxford, Oxford, UK | ⁶Department of Pathology, Ankang Central Hospital, Ankang, China | ⁷Peng Cheng Laboratory, Shenzhen, China | ⁸Jinan Inspur Data Technology Co., Ltd., Jinan, China | ⁹Department of Pathology, Xijing Hospital, Fourth Military Medical University, Xi'an, China

Correspondence: Chao He (chao.he@eng.ox.ac.uk) | Zhe Wang (zhwang@fmmu.edu.cn) | Yonghong He (heyh@sz.tsinghua.edu.cn)

Received: 17 October 2025 | **Revised:** 6 March 2026 | **Accepted:** 6 March 2026

Keywords: foundation model | glomerular lesion | glomerular pathology | pretraining | renal disease

ABSTRACT

Glomerular pathology is central to the diagnosis and prognosis of renal diseases, yet the heterogeneity of glomerular morphology and fine-grained lesion patterns remain challenging for current AI approaches. We present **GloPath**, an entity-centric foundation model trained on over one million glomeruli extracted from 14 049 renal biopsy specimens using multi-scale and multi-view self-supervised learning. GloPath addresses two major challenges in nephropathology: glomerular lesion assessment and clinicopathological insights discovery. For lesion assessment, GloPath was benchmarked across three independent cohorts on 52 tasks—including lesion recognition, grading, few-shot classification, and cross-modality diagnosis—outperforming state-of-the-art methods in 42 tasks (80.8%). In the large-scale real-world study, it achieved an ROC-AUC of 91.51% for lesion recognition, demonstrating strong robustness in routine clinical settings. For clinicopathological insights, GloPath systematically revealed statistically significant associations between glomerular morphological parameters and clinical indicators across 224 morphology–clinical variable pairs, demonstrating its capacity to connect tissue-level pathology with patient-level outcomes. Together, these results position GloPath as a scalable and interpretable platform for glomerular lesion assessment and clinicopathological discovery, representing a step toward clinically translatable AI in renal pathology.

1 | Introduction

Kidney disease affects over 850 million people worldwide and is projected to become the fifth leading cause of death by

2040 [1, 2]. Renal pathology remains the gold standard for definitive diagnosis and staging, with glomeruli serving as the most critical functional and diagnostic units [3–7]. Lesion assessment within glomeruli provides essential information for disease

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Advanced Science* published by Wiley-VCH GmbH

classification, prognosis estimation, and therapeutic decision-making. Beyond lesion recognition and grading, clinicopathological insights—derived from linking glomerular morphology with patient-level variables such as laboratory indices and clinical outcomes—offer unique opportunities to uncover disease mechanisms and inform precision medicine in nephrology.

However, glomeruli are highly heterogeneous entities. Their morphology reflects complex interactions among specialized cells, basement membranes, and mesangial matrix, and varies widely across disease processes such as sclerosis, inflammation, and capillary loop remodeling. Subtle yet diagnostically significant changes often require both fine-grained recognition of microstructural alterations and contextual interpretation of their spatial organization. This heterogeneity makes manual evaluation challenging and underscores the need for computational approaches that explicitly model glomeruli as the fundamental pathological entities of the kidney.

Artificial intelligence (AI) has shown promise in computational pathology, but current approaches fall short in glomerular pathology. Most models are slide-level or patch-based and lack explicit entity-centric representation learning. As a result, they struggle to capture disease-relevant microstructural features, encode long-range dependencies within glomeruli, and generalize across cohorts or modalities [8–16]. These limitations hinder their utility in clinically important tasks such as few-shot learning, multi-cohort validation, and cross-modality adaptation.

To address these challenges, we developed GloPath, an entity-centric foundation model specifically designed for glomerular pathology (Figure 1). In contrast to conventional patch-based pretraining that treats local image regions as independent samples, GloPath adopts glomeruli as the fundamental learning units, with each training instance corresponding to a biologically defined pathological entity automatically localized from whole-slide images. Pre-trained on over one million glomeruli extracted from 14 049 renal biopsy specimens using self-supervised multi-scale and multi-view learning, GloPath learns robust semantic representations of glomerular morphology with improved interpretability. We demonstrate its broad applicability across two key domains: glomerular lesion assessment, where GloPath achieves superior performance across diverse recognition, grading, and generalization tasks; and clinicopathological insights, where it reveals novel associations between glomerular morphology and patient-level clinical indicators. Together, these contributions highlight GloPath as a scalable and clinically translatable platform for computational nephropathology.

2 | Results

2.1 | An Overview of GloPath

Figure 1 provides an overview of GloPath. To enable accurate and scalable glomerular lesion assessment and clinicopathological analysis, we first assembled a large-scale, multicenter, multi-staining renal pathology dataset. Pathological images were collected from seven independent cohorts, and glomerular entities were automatically extracted from whole-slide images (WSIs) using a pretrained detection model. In total, 1 017 879

glomeruli were obtained across four staining modalities—H&E, PAS, Masson’s trichrome (MT), and PASM (Figure 1a).

We then conducted self-supervised pretraining with multi-scale and multi-view constraints to obtain GloPath (Figure 1b; Section 4.2.2). This pretraining strategy enables the model to capture fine-grained cellular morphology alongside larger-scale structural organization, while preserving the integrity of glomerular entities and their contextual tissue features.

We systematically evaluated GloPath across two major domains. For glomerular lesion assessment (Figure 1c), GloPath was fine-tuned on multiple datasets and demonstrated strong performance in lesion recognition, lesion grading, cross-modality diagnosis, and few-shot classification. The model was further deployed in a real-world clinical setting for open-set lesion recognition, where it exhibited robust generalization beyond closed-set training distributions. For clinicopathological insights (Figure 1d), GloPath was used to extract quantitative glomerular morphological parameters, which were subsequently associated with clinical variables from two independent cohorts, revealing meaningful links between glomerular morphology and patient-level clinical indicators.

2.2 | An Overview of Lesion Assessment

Glomerular lesions constitute the core of glomerular pathology and are critical determinants of kidney disease classification, progression, and clinical outcomes. Leveraging an entity-centric pretraining strategy, GloPath was applied to four major lesion assessment tasks: lesion recognition, grading, cross-modality diagnosis, and few-shot classification. Across all 52 tasks, GloPath consistently achieved state-of-the-art performance. Across all 52 tasks, GloPath consistently achieved leading performance, demonstrating clear advantages over general-purpose pathology foundation models with substantially larger parameter counts, larger training datasets, and more diverse modalities (Figure 2; Section S4.3.1).

Figure 2a summarizes the best-performing models for each of the 52 lesion assessment tasks. Key observations include:

- Overall superiority: GloPath attained the highest performance on 42 out of 52 tasks (80.8%), substantially exceeding the performance of UNI, demonstrating efficient parameter specificity despite a smaller model size and training volume.
- Performance of lesion recognition and grading: For tasks on internal cohorts, GloPath led in 21 lesion recognition tasks and 6 grading tasks, highlighting its capacity to capture fine-grained entity features, consistent with the use of multi-scale and multi-view feature constraints.
- Cross-modality generalization: GloPath achieved the overall performance lead in cross-modality diagnosis tasks, indicating that its learned glomerular representations generalize across different staining modalities.
- Few-shot learning: On 13 external or cross-modality few-shot tasks, GloPath outperformed all baselines, supporting the robustness and specificity of its pretrained features.

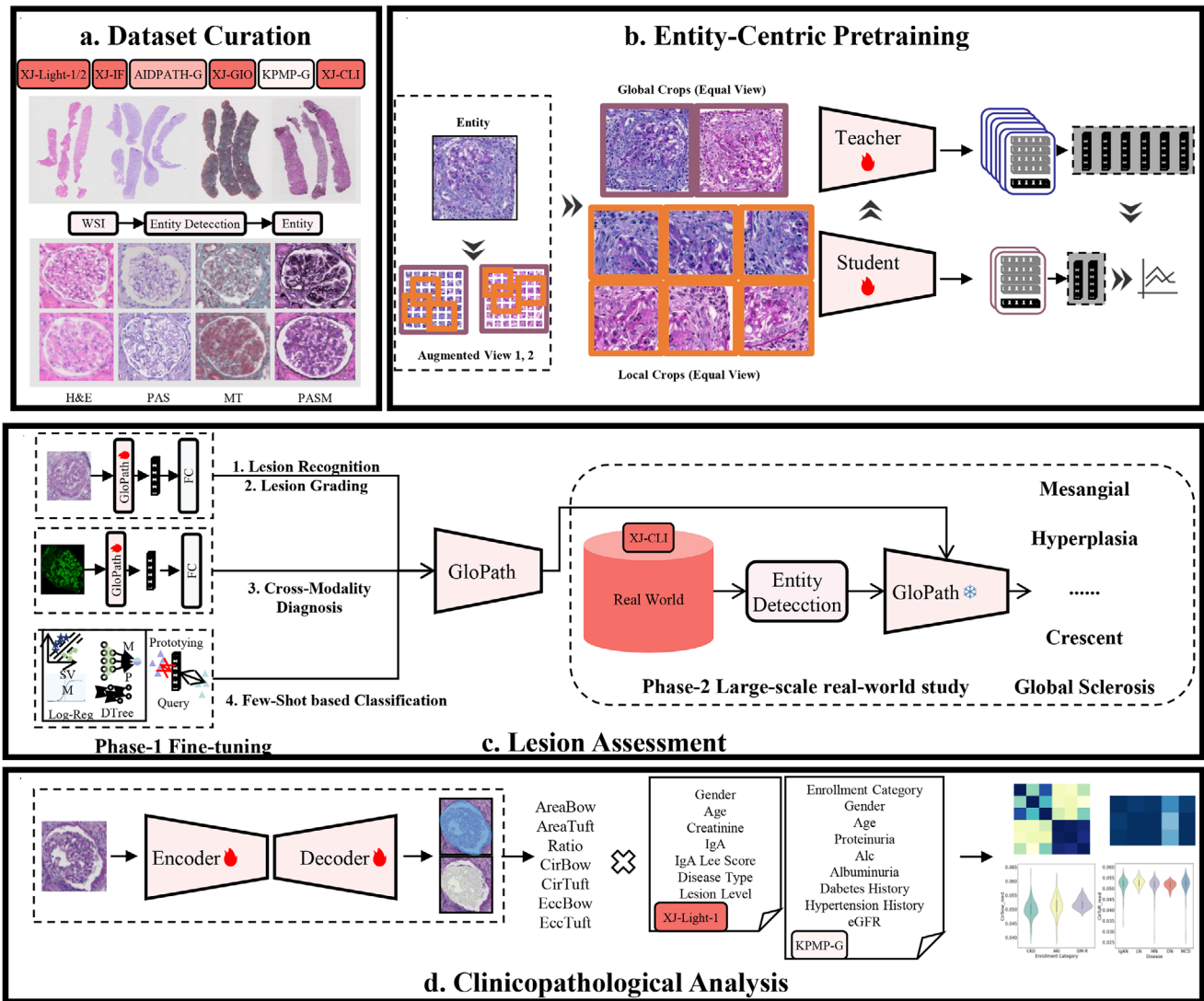


FIGURE 1 | Overview of GloPath. (a) Dataset curation. The pathological images are obtained from seven cohorts, including XJ-Light-1, XJ-Light-2, XJ-IF, AIDPATH-G, XJ-GIO and KPMP-G, where WSI is processed by entity detection. (b) Entity-centric self-supervised pretraining based on multi-scale multi-view. (c) Lesion assessment. (d) Clinicopathological correlation analysis.

Figure 2b provides further quantitative evidence of GloPath's advantages. First, it exhibits superior generalization, with overall performance significantly higher than all competing models ($p < 0.001$) and a median score approaching 0.95. Second, GloPath demonstrates remarkable data efficiency: compared with the second-best model (UNI), it achieves superior performance using only 1% of the training data and 23.2% of the model parameters, in contrast to methods such as CONCH and PLIP that rely on multimodal data. Third, GloPath shows strong stability, as indicated by the narrowest interquartile range among all models, concentrated in the high-performance region. Notably, the RenalPath model developed in this study ranked second only to UNI among general-purpose models, highlighting the effectiveness of large-scale renal pathology pretraining. The additional performance gains achieved by GloPath over RenalPath can be primarily attributed to entity-level pretraining on explicitly detected glomeruli, rather than differences in network architecture or training strategy, underscoring the importance of aligning pretraining units with disease-relevant pathological entities.

In summary, GloPath establishes a strong new benchmark in glomerular lesion assessment, combining high accuracy, data efficiency, and stability, providing a robust foundation for clinical deployment.

2.2.1 | Performance of Lesion Recognition

Overall, GloPath significantly outperformed all competing models across the 27 lesion recognition tasks (all $p < 0.05$; Figure 3a). Both the median and mean F1 scores of GloPath exceeded 0.95, markedly higher than UNI ($p = 6.407 \times 10^{-7}$). Among the general-purpose foundation models, UNI performed slightly better than CONCH ($p = 0.2386$) and RenalPath ($p = 0.0362$), while PLIP showed only marginal improvements. RandomInit and ImageNet exhibited substantial performance deficits compared with all other models (all $p < 0.05$). Stratified analyses by staining modality further confirmed GloPath's advantage (Figure 3b-d). For PAS, MT, and PASM stainings, GloPath achieved the

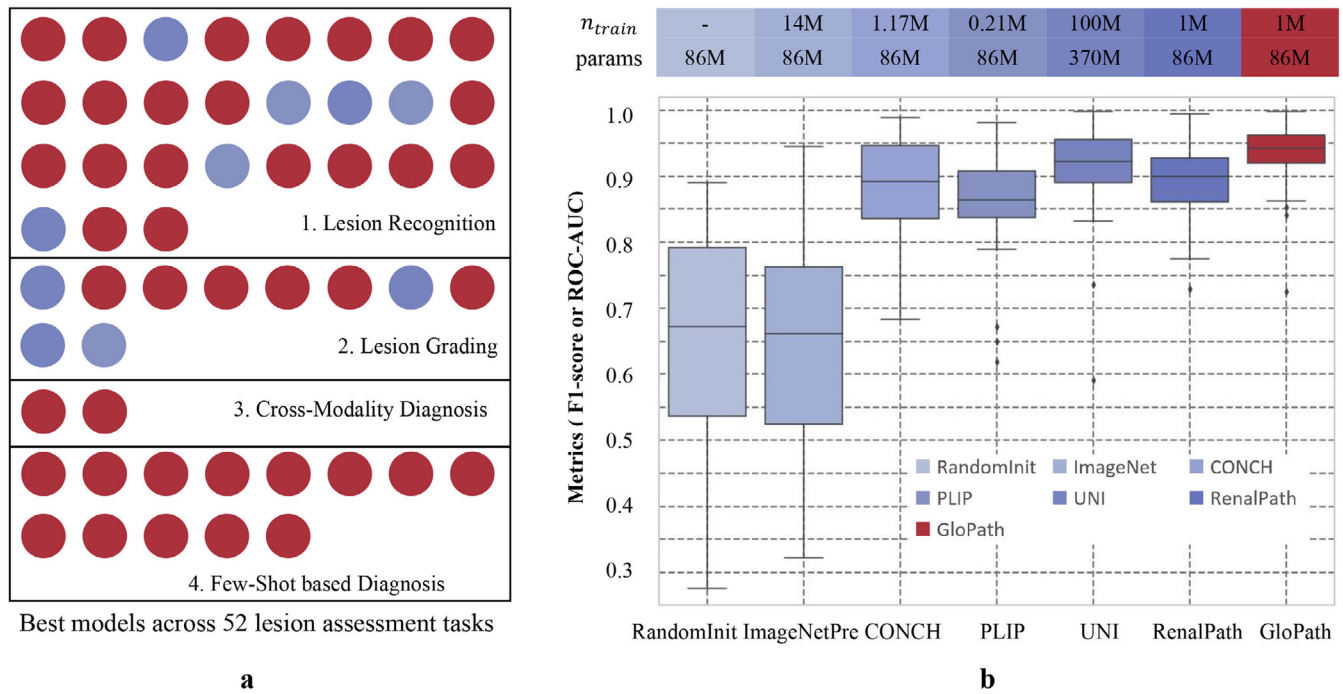


FIGURE 2 | Overview of the performance of lesion assessment. (a) The best model on each lesion assessment task across lesion recognition, lesion grading, cross-modality diagnosis, and few-shot-based diagnosis. Refer to the legend in (b). (b) Comparison of the distribution of the results of lesion recognition.

best results in 10 of 11, 3 of 6, and 8 of 10 lesion categories, respectively. Relative to the next-best model for each lesion, the maximum performance gains reached 7.05% (PAS), 2.53% (MT), and 3.41% (PASM). On average, GloPath exceeded general-purpose pathology foundation models by 7.10%, 3.40%, and 2.86% on PAS, MT, and PASM, respectively ($p = 0.0010$, 0.0313 , and 0.0020). Compared with RenalPath, the corresponding gains were 4.62%, 3.11%, and 3.57% (all $p < 0.05$), suggesting that RenalPath performed slightly better than other general-purpose models but remained inferior to GloPath. The performance gap was even more pronounced against non-pathology-based models. Here, GloPath achieved average gains of 34.63% (PAS), 33.24% (MT), and 24.39% (PASM; all $p < 0.05$). Together, these results highlight not only GloPath's superior diagnostic accuracy across multiple stainings but also its consistent ability to capture a higher proportion of lesion patterns at clinically meaningful performance levels.

GloPath consistently achieved the best performance across PAS, MT, and PASM stainings, with robust recognition accuracy and clinically relevant precision-recall balance (Figure 3b–d, Tables S3–S5).

- PAS:** Figure 3b and Table S3 show that on PAS staining, GloPath achieves an overall lead in recognition performance with an overall F1 score of 0.938. Over 0.95 performance was achieved on PAS(-), Crum, SS, MN, ER, Cre, and GS lesions, and the largest gain compared to other models was achieved on Crum. In terms of the more challenging PR-AUC metrics, GloPath achieved about 0.8 and above performance on lesions such as MH, PAS(+), PAS(-), Cre and GS, which fully demonstrates that GloPath has already reached a fairly reliable

performance for clinical application in these major lesion modes. In comparison, RenalPath and UNI only achieve about 0.8 and above performance on more typical lesions of MH, PAS(-), and GS, while only the latter of the non-pathology-foundation models RandomInit and ImageNetPre achieves more than 0.8 PR-AUC on GS.

- MT:** Figure 3c and Table S4 show that GloPath has the best overall performance on MT staining, with an average F1 score of 0.965 on the diagnostic task for each lesion, and achieves more than 0.97 performance on lesion recognition for SS, SFN, Cre and GS. Compared to other models, GloPath achieves the highest maximum gain in SFN. In terms of PR-AUC, GloPath is able to achieve more than 0.75 performance on MH and GS, which indicates that PR-AUC is able to better balance precision and recall in a clinical setting. In comparison, among the other models, only UNI, PLIP and RenalPath are able to reach the corresponding level on GS.
- PASM:** Figure 3d and Table S5 show that on PASM, GloPath embodies the best performance on the vast majority of lesions, with an average F1 score of 0.951, and achieves more than 0.97 performance on lesions such as PAS(-), SFN, Cre and GS. In terms of PR-AUC, GloPath is the best performer, being able to achieve more than 80% performance on lesions such as PAS(-), Cre, and GS, a number that is one and two more than UNI and RenalPath, respectively. This shows that GloPath has been able to have strong performance for clinical applications.

Grad-CAM-based visualization further underscored GloPath's interpretability (Figure 3b–d). Compared with UNI, GloPath localized lesion regions more accurately and comprehensively. For instance, in PAS–MH, GloPath highlighted thylakoid stromal

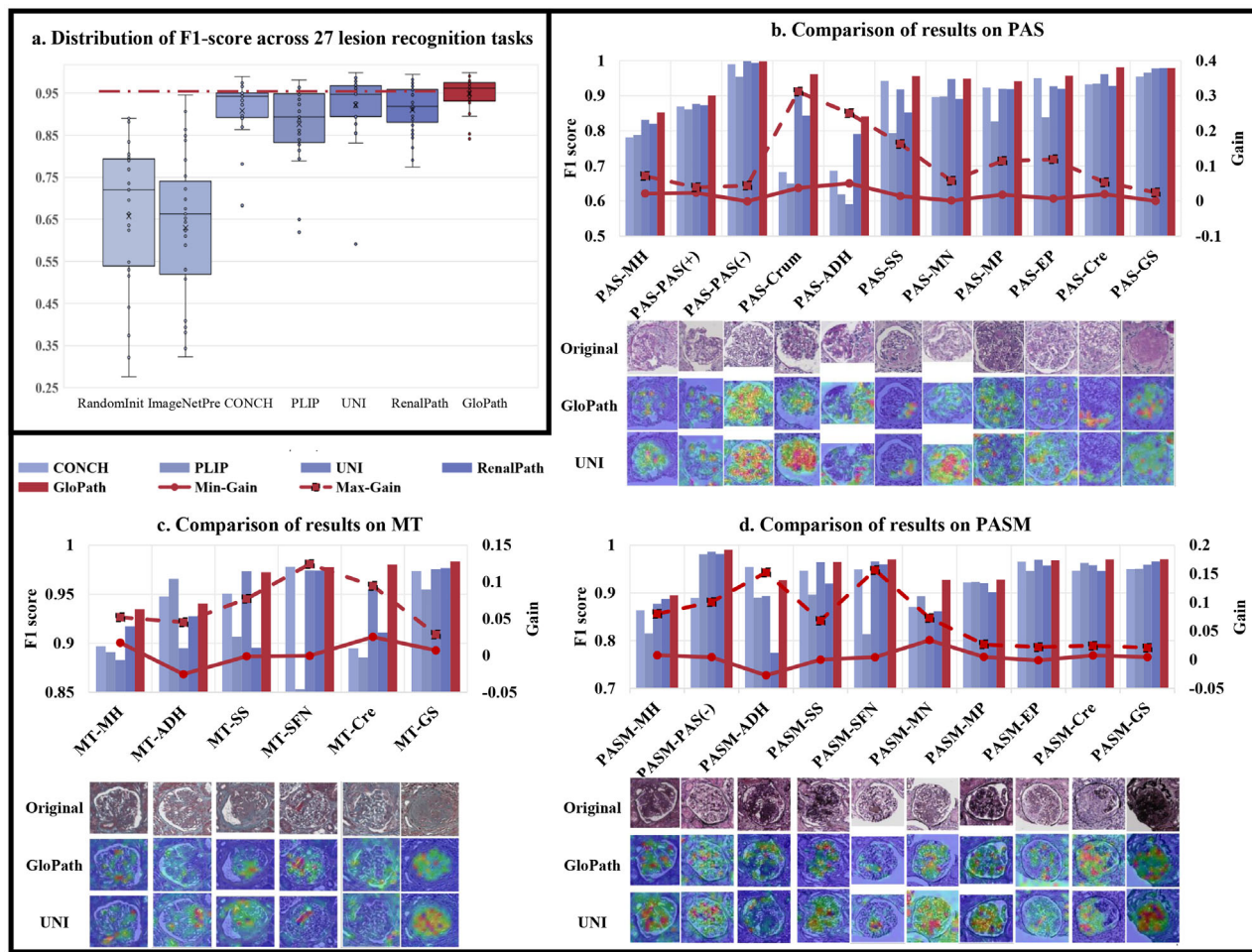


FIGURE 3 | Comparison of results of glomerular lesion analysis. (a) Comparison of the distribution of F1 score of lesion recognition. (b–d) Comparison of results on PAS, MT and PASM stainings. The primary vertical axis corresponds to the F1 score of each model. The secondary vertical axis corresponds to Min-Gain and Max-Gain, which are, respectively, the minimum and maximum performance gains of GloPath over all other models. Below the chart are the visualization results, where rows 1–3 are original images and the results of GloPath and UNI, respectively.

hyperplasia and mesangial proliferation; in MT-SS, it focused precisely on segmental sclerosis regions; and in PASM-Cre, it captured mural epithelial cell proliferation with high fidelity. These interpretable attention maps provide mechanistic plausibility for GloPath’s superior diagnostic accuracy and support its potential for clinical application.

2.2.2 | Performance of Lesion Grading

Lesion grading provides a more fine-grained assessment of disease progression and carries substantial diagnostic value in renal pathology [17–20]. This task also enables direct evaluation of model parameter specificity across diverse morphological subtypes. We benchmarked GloPath and six comparative models on 10 lesion grading tasks (Table S12).

GloPath achieved the best overall performance, leading in 6 of 10 lesion types, while UNI and CONCH ranked first in 3 and 1 tasks, respectively. On average, GloPath outperformed RenalPath by 1.80% ($p = 0.0039$), general-purpose foundation models by 2.71% ($p = 0.0039$), and non-pathology models by 12.96% ($p = 0.0020$).

Performance gains varied by staining. Compared with RenalPath, general-purpose pathology models, and non-pathology models, GloPath achieved average improvements of 1.15%, 1.70%, and 11.28% on PAS; 1.36%, 3.87%, and 19.40% on MT; and 2.65%, 3.13%, and 11.42% on PASM, respectively. Notably, the largest relative gain was observed for MT and PASM staining, while RenalPath consistently outperformed the general-purpose models.

Lesion-specific analyses further highlighted the fine-grained recognition capacity of GloPath. Relative to RenalPath, general-purpose, and non-pathology models, GloPath achieved average gains of 0.82%, 2.75%, and 18.33% on MH; 2.82%, 3.94%, and 13.03% on Cre; and up to 2.69%, 1.80%, and 11.43% on EP. These results underscore the specificity of GloPath in capturing subtle morphological variations across lesion subtypes, while again indicating that RenalPath performs better overall than general-purpose pathology foundation models.

2.2.3 | Performance of Cross-Modality Diagnosis

Immunofluorescence is indispensable for renal pathology, allowing precise classification of immune-mediated glomerular

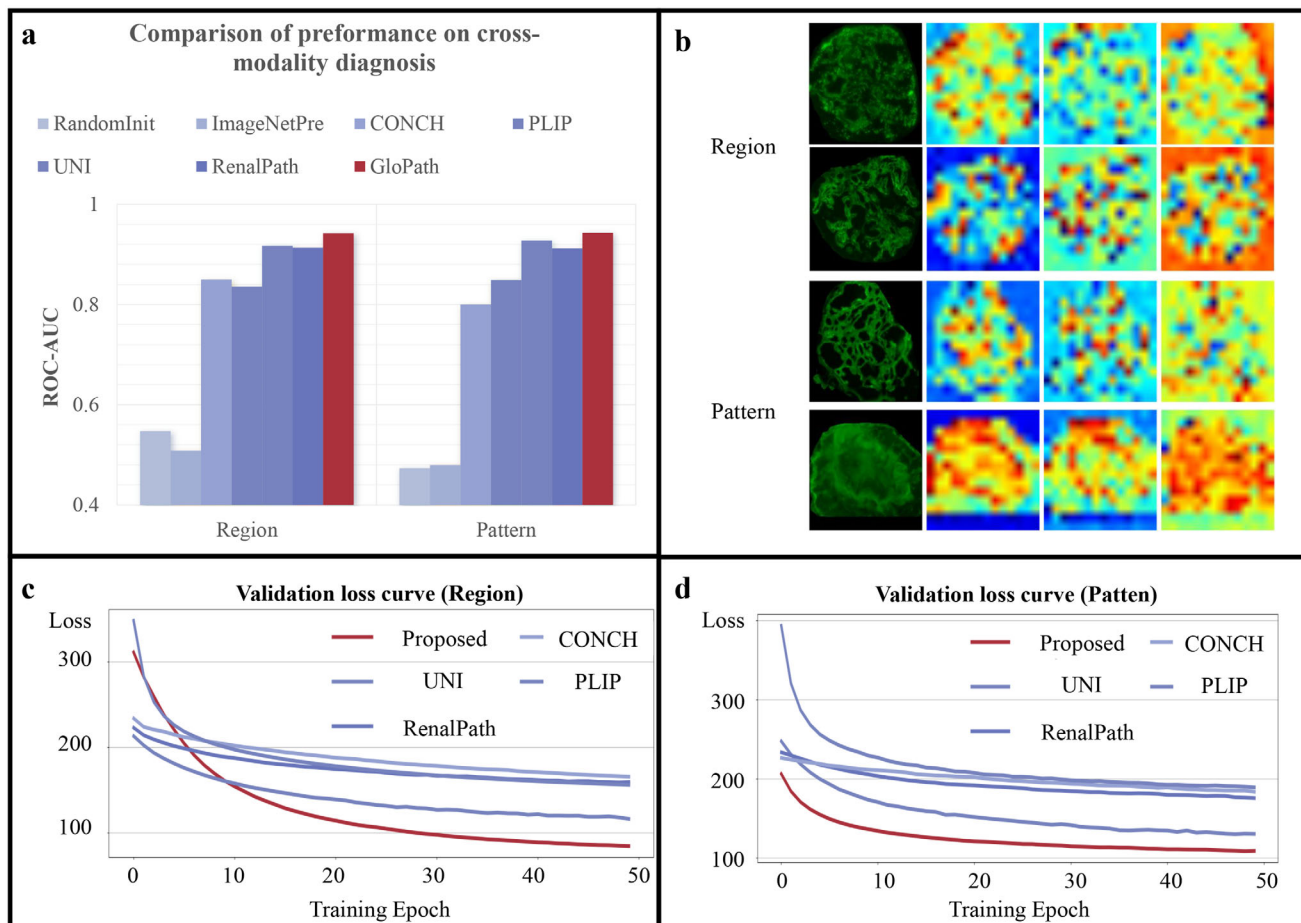


FIGURE 4 | Comparison of results of cross-modality diagnosis. (a) Comparison of performance on cross-modality diagnosis. (b) Visualization of attention map of GloPath. Row 1-2 indicate region classification and row 3-4 indicate pattern classification. Column 1 indicates the original IF images and column 2-4 indicate the attention map from three attention heads. (c, d) Validation loss curves for region and pattern respectively.

diseases by visualizing immune complex deposition [21–23]. Although fluorescence and bright-field images differ markedly in appearance, they share a common glomerular morphological framework. We therefore evaluated GloPath’s ability to generalize across modalities by testing its performance on immunofluorescence (IF)-based lesion classification.

Under full supervision, GloPath achieved the best performance in both deposition region and deposition pattern classification (Figure 4a; Table S34). Relative to the suboptimal model UNI, GloPath yielded gains of 2.50% and 1.57%, respectively. Larger improvements were observed over PLIP and CONCH (9.94% and 11.87%), as well as over RenalPath (2.86% and 3.14%). Compared with non-pathology models, GloPath demonstrated substantial gains of 41.44% and 46.72%. Across evaluation metrics, GloPath consistently achieved the highest performance, with an average ROC-AUC of 88.66% for deposition region classification, exceeding deposition pattern classification by 0.51%.

Visualization further supported the interpretability of cross-modality performance. Attention maps (Figure 4b) revealed that different attention heads captured complementary structural features—including glomerular interiors, boundaries, and fluorescent deposition signals—while the integrated attention mechanism localized lesions in a visually consistent manner.

Training dynamics also reflected efficient domain adaptation: although initial loss values were higher for deposition region classification, GloPath showed the fastest decline and reached the lowest loss within 10 epochs (Figure 4c). For deposition pattern classification, GloPath maintained consistently lower loss values than all other models, supporting its capability to capture glomerular-level pathology across imaging modalities (Figure 4d). To qualitatively assess cross-modality representation alignment, Uniform Manifold Approximation and Projection (UMAP) was used to visualize feature embeddings from brightfield and IF images. The embeddings showed substantial overlap after IF adaptation, suggesting that IF adaptation reduces modality-induced divergence at the feature level (Figure S13).

2.2.4 | Performance of Few-Shot Based Classification

We first established a binary classification task on the AIDPATH-G dataset to detect the presence of glomerulosclerosis [17, 24]. While full supervision with GloPath demonstrated clear advantages, this setting alone was insufficient to fully validate its representational efficiency (Table S6). To further evaluate GloPath’s feature expressiveness, we compared few-shot performance across multiple classifiers—including SVM, Linear

Regression (LR), Multi-Layer Perceptron (MLP), Random Forest (RF), and Prototype Learning (PTL)—using limited positive and negative samples (Figure 5).

Across all classifiers, GloPath consistently achieved the best performance (Figure 5a). With SVM, GloPath maintained the lead for all sample sizes ($k = 1$ to 100), with the greatest relative advantage at lower k . For instance, with only one positive and one negative sample, GloPath achieved 79.45% ROC-AUC, a 7.60% gain over the next-best model (Table S7, $p = 0.0371$). Similar trends were observed with LR, where GloPath outperformed competitors across all sample sizes, yielding 0.29%–7.93% improvements over the suboptimal model (all $p < 0.05$) and 1.57%–17.68% improvements over RenalPath (Table S8). GloPath also led consistently in MLP, while in RF the performance gains were modest ($\leq 2.24\%$) yet stable across sample sizes (Tables S9 and S10). In PTL, GloPath's advantage was most pronounced at small k , with gains of up to 21.89% relative to non-foundation models (Table S11).

Feature visualization further highlighted GloPath's strong representations. Without downstream fine-tuning, t-SNE plots (Figure 5b) showed that GloPath and PLIP separated normal and sclerotic glomeruli most clearly, while RandomInit and ImageNetPre failed to produce discernible boundaries. Attention maps (Figure 5c) revealed complementary focus among attention heads: some emphasized glomerular interiors, others the surrounding context or Bowman's capsule, suggesting biologically interpretable partitioning of structural features during pretraining.

Few-shot classification on IF images posed additional challenges, requiring cross-modality feature transfer and rapid adaptation from limited labeled data. Even under these stringent conditions, GloPath maintained robust performance (Figure 5d). With LR, GloPath achieved maximum gains of 37.69% and 42.40% for deposition region and deposition pattern classification, respectively, with average gains of 24.86% and 31.49% across k values (Tables S35 and S36). With MLP, GloPath again achieved the best performance at all k . Notably, at $k = 1$, GloPath achieved ROC-AUC values of 66.61% and 66.53% for region and pattern classification, outperforming UNI by 4.08% and 4.75% despite UNI's substantially larger pretraining scale. As k increased, GloPath's relative margin narrowed but remained significant. Performance advantages over RF and PTL were smaller but still consistent across modalities and tasks (Tables S39–S42).

To further assess the stability and data efficiency of different models under few-shot settings, we analyzed the performance distributions across 10 independent repeated experiments at each k . As shown in the box plots (Figure S3-few-if-conf), GloPath achieved comparatively higher performance at low k values and reduced performance variance, particularly in low-data regimes on AIDPATH-G (Figure S3). In contrast, competing methods exhibited larger interquartile ranges and greater variability across repeated runs, suggesting higher sensitivity to data sampling. These results further indicate the relatively data-efficient and stable learning behavior of GloPath.

Together, these results demonstrate that GloPath achieves superior few-shot generalization in both bright-field and fluorescence

modalities. Its robust feature representations enable rapid and accurate classification from extremely limited labeled data, highlighting strong potential for deployment in real-world clinical scenarios where annotations are scarce.

2.2.5 | Large-Scale Real-World Study

To evaluate the clinical utility of GloPath beyond curated datasets, we constructed XJ-CLI, comprising 13 749 unfiltered PAS-stained glomeruli collected under routine clinical workflows. On this large-scale real-world cohort, GloPath achieved a mean ROC-AUC of 0.913 (as shown in Table S43), representing only a 3.39% decrease compared with the curated test set (XJ-Light-2).

At the lesion level, GloPath maintained high performance across diverse pathological entities. In XJ-CLI, seven lesions achieved ROC-AUCs greater than 0.90, including MH (0.897), Crum (0.983), SS (0.920), MN (0.919), MP (0.982), EP (0.999), and GS (0.997). Importantly, for clinically critical lesions such as MH, Cre, and GS, ROC-AUCs reached 0.897, 0.954, and 0.997, respectively. Notably, GloPath even outperformed its internal benchmark on MP (0.982 vs. 0.969) and EP (0.999 vs. 0.974), underscoring its robustness in clinical settings. Direct comparison with XJ-Light-2 further highlighted the generalization capacity of GloPath. While certain categories, such as PAS(+), Crum, and ADH, showed moderate decreases (–9.9%, –15.2%, and –14.9%, respectively), performance remained stable or improved for others. The overall pattern indicates that GloPath retains strong discriminative ability under real-world data variability, achieving clinically actionable accuracy.

To characterize the failure modes of GloPath, representative false-negative (FN) and false-positive (FP) cases were analyzed for each lesion type in the XJ-CLI cohort (Figure S12). Overall, errors were observed across both common and lesion-specific patterns. Among common error modes, false-positive predictions in membranous nephropathy (MN; FP, $p = 0.7686$) and mesangial proliferation (MP; FP, $p = 0.9433$) frequently occurred in glomeruli with relatively homogeneous but non-specific staining. Biopsy-related artifacts further contributed to misclassification, particularly in cases exhibiting elongated glomerular morphology and intensified staining. Such artifacts were associated with increased error rates, including MN FP ($p = 0.9636$), PAS(–) FN ($p = 0.7891$), and segmental sclerosis (SS) FN ($p = 0.5964$), as well as over-prediction in SS FP cases. Lesion-specific errors were mainly concentrated in histologically borderline or confusable abnormalities. FP cases in ADH ($p = 0.8186$) and GS ($p = 0.9293$) showed morphological features overlapping with adjacent structures, leading to incorrect positive predictions. FN cases in mesangial hypercellularity (MH; $p = 0.3555$) were characterized by mild or early-stage changes, resulting in intermediate prediction scores. Similarly, FN cases in crescentic lesions (Cre) demonstrated partial parietal epithelial proliferation that did not reach diagnostic thresholds, contributing to under-recognition by the model.

To further assess the interpretability of GloPath, we conducted a quantitative analysis of attention–lesion alignment on the XJ-CLI cohort. Specifically, 10 Cre and 10 GS cases were ran-

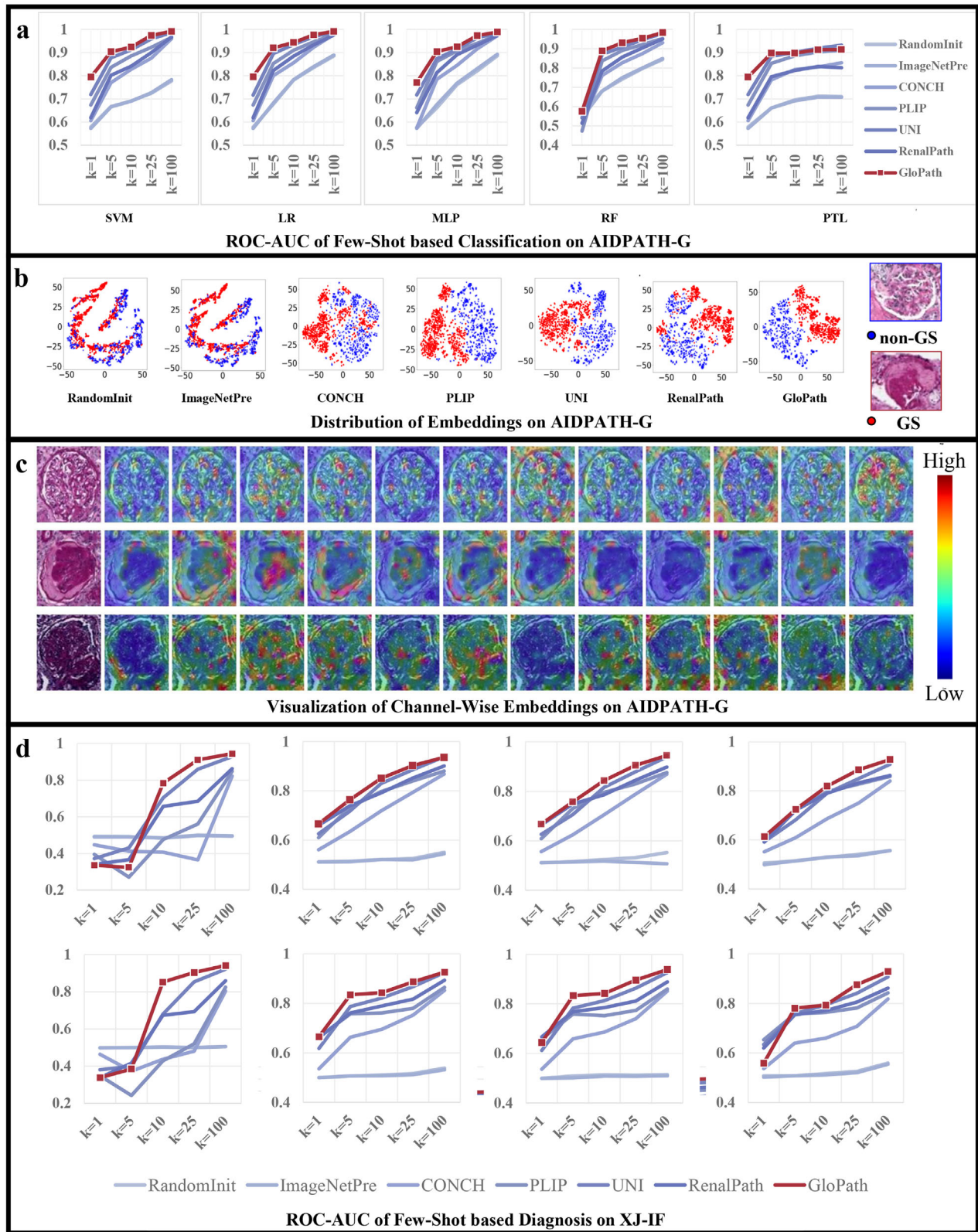


FIGURE 5 | Comparison of results of few-shot-based classification. (a) Performance of few-shot based classification on AIDPATH-G. (b) Distribution of embedding on AIDPATH-G using t-SNE. (c) Visualization of Channel-wise embedding on AIDPATH-G. (d) Results of few-shot for IF image classification. Row 1-2 indicate results for region, pattern, respectively. Column 1-4 indicate results for different methods, including LR, MLP, RF, and PTL, respectively.

domly selected, and experienced nephrologists provided coarse bounding-box annotations to localize the dominant lesion regions within each glomerulus. This box-level annotation strategy was adopted to reduce annotation burden and inter-observer variability, while aligning with the classification-focused design of the task. Attention heatmaps were then used to compare attention intensities inside versus outside the annotated lesion boxes. Statistical analysis revealed that attention values within lesion-localized regions were significantly higher than those in non-lesion regions (0.5118 ± 0.0351 vs. 0.3261 ± 0.0438 for Cre and 0.4907 ± 0.0534 vs. 0.1816 ± 0.0439 for GS), with the distributional difference confirmed by the KS test ($D = 1$ and $p < 0.0001$). Representative examples and quantitative summaries are shown in Figure S14. These results provide quantitative evidence that GloPath attention aligns with expert-annotated pathological regions, supporting the interpretability of attention maps beyond qualitative visualization.

In summary, the large-scale real-world study on XJ-CLI demonstrates that GloPath delivers rapid and reliable lesion recognition in routine clinical settings. The model consistently achieves high performance across 11 types of glomerular lesions, with most misclassifications arising in borderline cases or under suboptimal histological conditions, such as staining variability or biopsy-related artifacts. Importantly, these errors are largely confined to diagnostically ambiguous regions rather than overt lesion phenotypes, underscoring the robustness of GloPath's learned representations. Together, these findings highlight GloPath's strong translational potential, providing a foundation for timely, scalable, and clinically actionable computational pathology in real-world practice.

2.3 | Clinicopathological Insights

Quantitative analysis of glomerular morphology parameters in pathology images offers a more objective and reproducible assessment than traditional subjective evaluation [25–29]. Integrating these computationally derived morphological features with clinical data enables the identification of potential clinico-pathologic correlations that may serve as predictive biomarkers for disease progression [8, 30–32]. Furthermore, pathomics-driven data mining based on morphological segmentation facilitates high-throughput, data-driven exploration of renal pathology, providing mechanistic insights into disease pathogenesis and progression, and potentially guiding more precise diagnostic and therapeutic strategies.

In this study, GloPath enabled accurate quantification of glomerular morphology (Figure 6a), providing a robust foundation for pathomics analyses. Two cohorts with partial clinical information, XJ-Light-1 and KPMP-G, were utilized, and three models—GloPath, UNI, and RenalPath—were evaluated. The Kolmogorov–Smirnov (KS) and Kruskal–Wallis (KW) tests were employed to assess statistical differences in pathological phenotype distributions across clinical parameter groupings, thereby quantifying clinico-pathologic associations [33, 34].

As shown in Figure 6b,c, morphological parameters quantified by GloPath yielded the most robust pathomics-based insights. In the XJ-Light-1 cohort, GloPath identified significant correlations

in 37 of 98 pairs, matching UNI and outperforming RenalPath by 27.6% (Table S14). In the KPMP-G cohort, GloPath detected significant associations in 43 of 126 pathophysiological–clinical pairs, exceeding UNI by 7 pairs and RenalPath by 8 pairs (Table S15). Among these, correlations were classified as extremely significant (***, $n = 7$), significant (**, $n = 12$), and moderate (*, $n = 24$), underscoring the model's capacity to uncover clinically relevant morphological patterns.

2.3.1 | Morphological Segmentation

Quantitative glomerular morphology provides a reproducible basis for clinicopathological analyses, overcoming the subjectivity and labor-intensiveness of manual assessments [24, 35–46]. Here, we applied GloPath to segment two key glomerular structures, Bowman's capsule and tuft, across three cohorts: XJ-Light, XJ-GIO, and KPMP-G. Segmentor [47] was used with GloPath and six comparative models.

As shown in Figure S1, GloPath consistently achieved the highest IoU. On XJ-Light-1, segmentation of Bowman's capsule reached superior accuracy compared with UNI, despite using only 1% of the training data and 23.2% of the parameters. On XJ-GIO, tuft segmentation demonstrated similar gains, with IoU improvements of 8.22%–9.17% over general-purpose models PLIP and CONCH, 1.22%–1.23% over RenalPath, and 8.23%–15.10% over ImageNetPre and RandomInit. Segmentation performance for tuft was generally higher than for Bowman's capsule, with differences up to 7.21% on PLIP.

Visual inspection further illustrates these results. Figure S2a shows Bowman's capsule segmentation on XJ-Light, where GloPath accurately delineates the structure with minimal errors, while UNI and RenalPath occasionally mislabel edge regions (e.g., row 3). Figure S2b shows tuft segmentation on PAS-stained XJ-GIO glomeruli, including challenging cases such as absent Bowman's capsule (row 1) or small tufts (row 2). GloPath maintained accurate segmentation, whereas UNI and RenalPath showed under- or over-segmentation, and PLIP exhibited noticeable edge over-segmentation. Figure 6a shows Bow and tuft segmentation on KPMP-G; qualitative analysis indicates GloPath reliably segments extreme cases, such as extrusion deformation with incomplete Bowman's capsule (column 1, 3) and tuft with a small area (column 8), outperforming UNI and RenalPath in challenging scenarios.

In addition, we compared GloPath with the state-of-the-art medical image segmentation framework nnU-Net [48] under the same data splits. nnU-Net achieved IoU scores of 0.851 for Bowman's capsule and 0.905 for tuft segmentation, representing a strong task-specific upper bound. Notably, GloPath achieved comparable performance while relying on a unified entity-centric backbone rather than a fully supervised, task-specialized architecture.

These results demonstrate that GloPath enables accurate and robust glomerular morphological quantification across cohorts and staining conditions, establishing a solid foundation for downstream pathomics and clinicopathological correlation analyses.

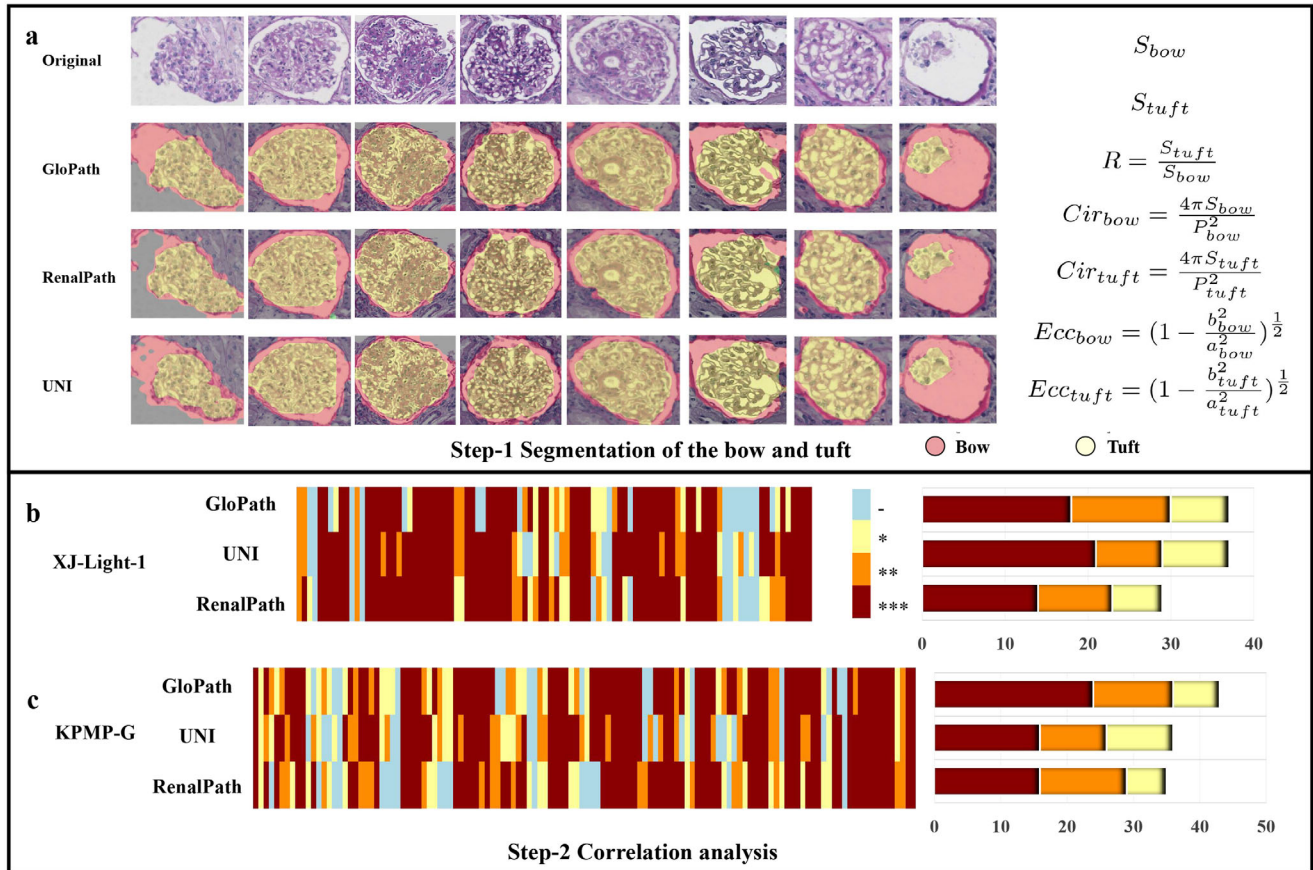


FIGURE 6 | Overview of clinicopathological correlation analysis. a, Visualization of bow and tuft segmentation and morphological parameter definition. b-c, Comparison of distribution of significance levels on XJ-Light-1 and KPMP-G.

2.3.2 | Insights on XJ-Light-1

In the XJ-Light-1 cohort, GloPath revealed robust associations between glomerular morphological features and clinical parameters (Figure 7a,c,e; Tables S16– S22). Notably, male patients exhibited larger glomerular areas than females, consistent with prior evidence suggesting estrogen’s protective role in limiting compensatory hypertrophy (Table S16) [49]. Age negatively correlated with glomerular roundness, indicating progressive structural simplification, likely due to capillary loop collapse and basement membrane changes (Table S17) [50].

Disease-specific patterns were also evident. In IgA nephropathy, glomerular area exhibited a progressive downward shift with disease severity, consistent with a transition from early compensatory hypertrophy to subsequent sclerosis (Tables S19 and S20). Across nephropathy subtypes, glomerular area distributions showed distinct yet partially overlapping profiles: diabetic nephropathy (DN) and membranous nephropathy (MN) tended toward larger glomerular areas, suggestive of compensatory expansion; lupus nephritis (LN) demonstrated pronounced heterogeneity, spanning both hypertrophic and atrophic morphologies; whereas IgAN and minimal change disease (MCD) displayed relatively modest area variation, consistent with early mesangial alterations or largely preserved glomerular structure (Figure 7g-left; Table S21). Comparable distributional trends were observed for tuft-related morphological features, including area, roundness, and eccentricity, as well as for the tuft-to-Bowman’s

capsule ratio, all of which varied in association with age, serum creatinine, injury severity, and disease category (Figure 7g-right). A comprehensive summary of statistically significant pathomics–clinical associations is provided in Figure S5.

To strengthen the clinical insights, we performed multivariate regression analysis (Tables S23 and S24). After adjustment for sex, age, and disease type (IgAN, LN, MN, DN, and MCD), selected GloPath-derived glomerular morphological features were still significantly associated with clinical outcomes. For example, smaller ratio of area of tuft to that of bow was independently associated with higher serum creatinine levels ($p = 0.0050$ for Proposed_CirTuft_mean and $p = 0.0006$ for Proposed_CirTuft_med). Larger eccentricity was also significantly associated with higher severity of injury ($p = 0.0244$ for Proposed_EccTuft_mean and $p = 0.0126$ for Proposed_EccTuft_med). Collectively, these findings suggest that GloPath-derived glomerular morphological features are significantly associated with renal function, lesion severity, and disease status, even after accounting for conventional clinical covariates.

2.3.3 | Insights on KPMP-G

In the KPMP-G cohort, encompassing a wider spectrum of clinical parameters, GloPath identified pronounced differences in glomerular morphology across patient groups (Figure 7b,d,f; Table S25– S33). CKD patients exhibited the largest glomerular

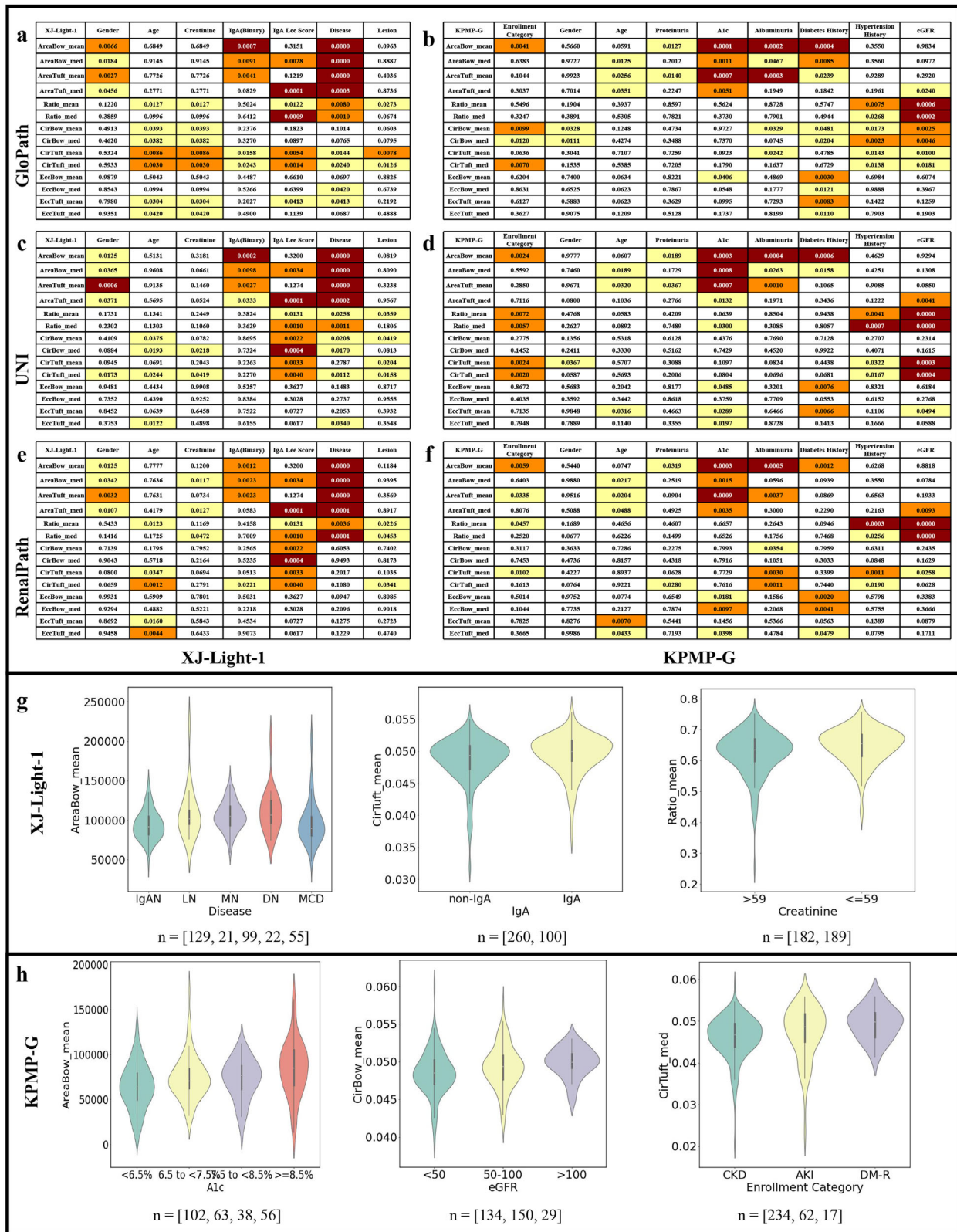


FIGURE 7 | Details of clinicopathological correlation analysis. (a-f) The p-values of correlation analysis. (a-f) refer to GloPath, UNI and RenalPath, respectively. The left and right column refer to XJ-Light-1 and KPMP-G respectively. (g, h) show examples from XJ-Light-1 and KPMP-G.

areas and lowest circularity, indicative of compensatory hypertrophy due to nephron loss, whereas diabetes without renal involvement (DM-R) patients had the smallest areas and highest roundness, reflecting preserved structure. AKI patients showed signs of atrophy consistent with acute ischemic or inflammatory injury (Figure 7h-right; Table S25).

Additional clinicopathological trends were observed across demographic and functional variables. Sex-associated differences in glomerular roundness were detectable, potentially reflecting sex-related hemodynamic or structural variations, while age-related patterns revealed a non-linear distribution, with middle-aged patients exhibiting relatively larger glomerular areas and younger and elderly groups showing reduced values (Tables S26 and S27). Functional and metabolic indicators also demonstrated consistent distributional associations with glomerular morphology. Proteinuria and HbA1c levels were associated with right-shifted glomerular area distributions, suggestive of compensatory hypertrophic responses under metabolic or filtration stress (Figure 7h-left; Tables S28 and S29). In contrast, albuminuria was linked to increased heterogeneity in glomerular area, indicating variable structural responses across disease states (Table S30). Clinical history further modulated morphological patterns: diabetes and hypertension were associated with coordinated changes in both glomerular area and roundness, while eGFR showed a positive association with roundness, consistent with the notion that preserved glomerular geometry supports filtration efficiency (Figure 7h-middle; Tables S31–S33). Tuft-level morphology largely mirrored these distributional trends, with tuft-to-Bowman’s capsule ratios varying across hypertension and eGFR subgroups. A comprehensive summary of pathomics-clinical associations is provided in Figure S5.

2.3.4 | Summary of Clinicopathological Insights

In addition to statistical significance, we quantified effect sizes for the identified associations using ϵ^2 (KW test) and the D statistic (KS test) (Tables S16–S33). Across both cohorts, most statistically significant clinicopathological associations exhibited small-to-moderate effect sizes (KW $\epsilon^2 < 0.1$ and KS $D \approx 0.1 - 0.2$). This pattern aligns with prior clinicopathological studies, in which individual histological or clinical variables typically explain only a limited fraction of phenotypic variance due to the indirect, non-linear, and multifactorial relationships between tissue-level pathology and patient-level clinical manifestations. In renal disease specifically, clinical phenotypes arise from the combined effects of structural lesions, molecular alterations, compensatory mechanisms, systemic factors, and their interactions, rather than from isolated pathological features. The observed associations, together with their consistent effect directions, support their robustness and clinical relevance despite the modest magnitude of individual effects.

Collectively, these analyses highlight the complex, multidimensional relationships between glomerular morphological phenotypes and clinical parameters across cohorts. Sex and age exert consistent effects, with males and middle-aged individuals exhibiting larger glomerular areas and structural variations indicative of compensatory or degenerative changes. Chronic conditions—including CKD, diabetes, and hy-

pertension—further modulate glomerular size, roundness, and tuft-to-Bowman ratios, reflecting disease-specific adaptive or pathological remodeling. Moreover, the interplay among these factors—such as age-related vulnerability combined with chronic comorbidities—likely contributes to the observed phenotypic heterogeneity. Markers of renal function and injury, including proteinuria, albuminuria, and eGFR, correlate with structural alterations, emphasizing the functional relevance of these morphometric features.

These findings demonstrate that GloPath can systematically capture subtle yet clinically relevant glomerular changes, enabling high-resolution pathomics analyses. By integrating computationally derived morphology with patient-level clinical data, this framework provides a powerful approach for elucidating disease mechanisms, identifying potential biomarkers, and guiding personalized nephrology care. While the mechanistic underpinnings require further investigation, the results underscore the potential of GloPath to advance next-generation clinicopathological research and multi-factorial, data-driven insights in renal pathology.

3 | Discussion

In this study, we present GloPath, a domain-specific, entity-centric foundation model for glomerular pathology, enabling accurate lesion assessment and clinicopathological analysis across diverse cohorts. GloPath demonstrates robust recognition of atypical and variable lesions, clinically relevant performance on representative lesions, and effective generalization across external and cross-modality datasets. GloPath excels in recognizing atypical lesions with limited samples or heterogeneous morphology (e.g., PAS-Crum, PASM-ADH), likely due to its multi-scale, multi-view feature-constrained pretraining that encodes low signal-to-noise patterns within broader contexts. It also handles lesions with broad variability or multiple pathological grades (e.g., PAS-SS, MT-SS, PASM-MH), reflecting its capacity to integrate local and global morphological cues. For both common and diagnostically challenging lesions, GloPath achieves clinically reliable recognition performance. Importantly, multi-head attention maps and t-SNE visualizations of GloPath provide interpretable insights into the structural features driving model predictions, offering a bridge between AI-derived features and human pathology understanding. This demonstrates that GloPath not only quantifies subtle pathological phenotypes but also provides mechanistic insights, supporting translational applications in nephrology research.

Overall, GloPath effectively achieves intrinsic, fine-grained, and cross-modal characterization of glomerular pathology with fewer training samples, reduced model parameters, and limited modalities. A key factor underpinning GloPath’s success is its entity-centric pretraining. By treating detected glomeruli as fundamental units, the model focuses on diagnostic structures rather than heterogeneous tissue regions, embedding strong domain specificity from approximately one million glomerular entities. Multi-scale and multi-view strategies further enrich the representations: multi-scale learning captures features preserved across magnifications, while multi-view integration encodes local morphology alongside global semantic patterns, promoting scale-

invariant and semantically consistent representations. Together, these design choices improve robustness to rare, heterogeneous, and multi-grade lesions.

Despite its encouraging generalization, several practical limitations remain. First, variations in scanners, acquisition settings, and staining protocols pose inherent challenges. Although GloPath is robust across institutions and staining types, most pretraining data come from a limited set of scanners and workflows, leaving subtle color or illumination differences under-represented. While the entity-centric and multi-scale design reduces reliance on low-level appearance features, systematic scanner harmonization (e.g., color normalization or domain adaptation) was not explored, and warrants future work. Second, rare or highly specific glomerulonephritis subtypes are under-represented, limiting performance on uncommon lesions or early transitional forms, especially when they resemble more prevalent patterns. Expanding multi-center collections of rare lesions and incorporating hierarchical or uncertainty-aware labeling could improve reliability in these clinically challenging cases.

Beyond lesion recognition, GloPath enables efficient pathomics-based exploration of clinico-pathological correlations. Across the XJ-Light-1 and KPMP-G cohorts, glomerular area, roundness, and the tuft-to-Bowman's capsule ratio were influenced by sex, age, disease subtype, and renal function indices such as proteinuria and eGFR. Male patients and individuals with CKD exhibited larger glomerular areas and lower circularity, whereas age-related changes were associated with structural simplification. Diabetes history, hypertension, and elevated HbA1c correlated with compensatory hypertrophy or morphological distortion, reflecting underlying pathophysiological processes. Distinct nephropathy subtypes displayed characteristic glomerular patterns: for example, compensatory hypertrophy in CKD and hyperfiltration-induced morphological changes in metabolic disease. These associations are consistent with established pathophysiological theories and reinforce the relevance of integrating computationally derived morphology with clinical data.

Integration of GloPath-derived morphological features with clinical parameters provides a novel layer of clinicopathological insight. While some associations remain at the research stage, they indicate potential for early, low-cost, non-invasive risk stratification using routine clinical data such as age, sex, proteinuria, and eGFR. Although effect sizes are generally small-to-moderate—reflecting the multifactorial, indirect nature of kidney disease—these systematic shifts are biologically interpretable and clinically meaningful. GloPath descriptors complement conventional clinical variables by capturing latent structural changes that may precede overt functional decline. Future multimodal frameworks combining these features with laboratory, imaging, or molecular data could enable personalized diagnostic and prognostic models, reduce reliance on invasive procedures, and enhance nephropathy monitoring [51, 52], ultimately improving patient care through timely, informed decision-making.

From a clinical perspective, GloPath provides rapid, reproducible, and accurate lesion recognition on both internal and real-world cohorts, mitigating inter-observer variability and supporting high-throughput analysis. Integration of morphological descriptors with routine clinical parameters enables subtle yet

systematic detection of structural alterations, complementing conventional risk assessment. Although individual associations generally exhibit small-to-moderate effect sizes, this is consistent with the multifactorial nature of kidney disease, where multiple morphological and systemic factors collectively influence clinical outcomes. Future multimodal frameworks combining GloPath features with laboratory, imaging, or molecular data could facilitate personalized diagnostics and prognostics, reducing reliance on invasive procedures.

A potential concern is that a substantial portion of the pretraining and internal evaluation data originates from a single institution, raising the possibility that the model may inadvertently capture center-specific artifacts such as staining protocols or scanner characteristics. Several aspects of GloPath's design and evaluation mitigate this risk. First, GloPath is pretrained at the entity level, focusing exclusively on automatically detected glomeruli rather than whole-slide regions that may encode background, tissue processing, or scanner-specific cues. This entity-centric formulation constrains the model to learn morphology-driven representations that are more tightly coupled to diagnostic structures. Second, the multi-scale and multi-view pretraining strategy encourages the learning of scale-invariant and semantically consistent features, reducing reliance on low-level appearance patterns that are more susceptible to institutional variation. Importantly, the robustness is empirically supported by its consistent performance on external cohorts (AIDPATH-G and KPMP-G), as well as across different imaging modalities. Together, these results suggest that GloPath captures transferable pathological semantics rather than institution-specific imaging artifacts, supporting its generalizability across diverse clinical settings.

Although this study spans multiple cohorts and staining modalities, additional validation across broader multi-center datasets is warranted. The granularity of morphological annotations remains a constraint, particularly in external cohorts lacking detailed labels. Moreover, while GloPath demonstrates strong predictive performance, further studies linking AI-derived morphological features with long-term clinical outcomes are necessary to fully establish clinical utility. Looking ahead, integrating GloPath into multimodal frameworks that combine histopathology, molecular data, and clinical parameters—particularly by leveraging recent advances in integrating morphology with spatial proteomics and spatial transcriptomics [53, 54], or even training glomeruli-centric or renal-specific phenotype-protein-gene association models—holds promise for enabling more comprehensive disease modeling. Enhancements in self-supervised learning and domain adaptation may further improve model generalizability across rare pathologies and staining protocols. Ultimately, the deployment of domain-specific foundation models such as GloPath represents a significant step toward scalable, interpretable, and clinically actionable AI in renal pathology.

4 | Methods

4.1 | Dataset Curation

High-quality data form the cornerstone of progress in computational renal pathology. The marked heterogeneity of kidney biopsy specimens—arising from variable disease subtypes, stain-

ing protocols, and preparation artifacts—necessitates datasets that are both large-scale and well-curated to support the development of robust AI models. In this study, the construction of a comprehensive chronic kidney disease (CKD) pathology database was critical not only for formulating clinically relevant questions but also for driving algorithmic innovation and enabling rigorous validation in real-world settings.

We assembled seven cohorts, comprising five institutional cohorts (XJ-Light-1, XJ-Light-2, XJ-IF, XJ-GIO, and XJ-CLI) and two publicly available cohorts (AIDPATH-G and KPMP-G). Each cohort was assigned to distinct analytic tasks: pretraining of GloPath and RenalPath leveraged portions of XJ-Light-2 (excluding samples reserved for validation); lesion recognition and grading were performed on XJ-Light-1; cross-modality diagnosis and few-shot classification employed XJ-IF and AIDPATH-G; and clinicopathological correlation analysis drew upon morphological segmentation from XJ-GIO and KPMP-G, with quantitative analysis conducted on XJ-Light-1 and KPMP-G. Detailed characteristics of each cohort, including sample size, staining modality, and annotation strategy, are described below.

4.1.1 | XJ-Light-1 and XJ-Light-2

Recognizing the limitations of existing public renal pathology datasets—namely, limited sample sizes, inconsistent annotations, and insufficient representation of disease heterogeneity—we established a large-scale, high-quality private database in collaboration with the Department of Pathology at Xijing Hospital, a leading tertiary care institution affiliated with the Air Force Medical University (Xi'an, China). This effort involved systematic patient data collection, digitization of pathology slides, preprocessing of whole-slide images, and expert lesion annotation.

- **Patient Data Collection:** Renal biopsy specimens were systematically collected at Xijing Hospital to ensure data integrity and consistency. Each biopsy yielded 12 consecutive sections, with two sections mounted per slide, resulting in six slides per case. Sections were stained using four standard histopathological protocols—Hematoxylin and Eosin (H&E), Periodic Acid-Schiff (PAS), Masson's Trichrome (MT), and Periodic Acid-Silver Methenamine (PASM)—in a fixed sequence (H&E, PAS, MT, PASM, H&E, PAS, H&E, PAS, MT, PASM, H&E, PAS). Supplementary clinical information was integrated, including pathology reports, histological findings, and submission reports with patient history, laboratory tests, and imaging results. In addition, IF images were incorporated to identify antibody depositions within glomeruli.
- **Digitization of Pathology Slides:** All slides were digitized using the SQS-600P slide scanning imaging system (Shenzhen Shengqiang Technology Co., Ltd.), producing high-resolution whole-slide images (WSIs) suitable for computational analysis (Figure S7).
- **Entity Extraction:** WSIs were initially downsampled using OpenSdpic to 5× magnification ($\approx 1.68 \mu\text{m}/\text{pixel}$) and converted into tensor formats compatible with AI models. A pre-trained glomerular detection model from our previous work [46] was used to localize glomeruli within renal tissue,

after which glomerular regions were extracted at 20× magnification ($\approx 0.42 \mu\text{m}/\text{pixel}$) for subsequent analysis.

- **Lesion Annotation:** A dedicated software platform, the Kidney Labeling Tool, was developed to support precise annotation (Figure S8). Using this platform, all glomerular lesion annotations in XJ-Light-1 were independently performed by two well-trained nephrologists (Yanxia Wang, and Xiaoqin Wang) following standardized diagnostic criteria. In cases of disagreement, a senior renal pathologist (Jing Li) with extensive diagnostic experience reviewed the discrepant cases and provided adjudicated consensus labels, which were used for model training and evaluation.

XJ-Light-1 represents the inaugural release of light microscopy images from Xijing Hospital, encompassing CKD cases diagnosed between 2019 and 2021. It includes 870 cases, each represented by six slides, totaling 5,220 slides. Slides were stained with H&E, PAS, MT, and PASM, capturing complementary structural features of glomeruli and tubulointerstitium. Each case was systematically classified into specific CKD subtypes, including diabetic nephropathy (DN), focal segmental glomerulosclerosis (FSGS), anti-neutrophil cytoplasmic antibody-associated glomerulonephritis (AAGN), membranous nephropathy (MN), obesity-related glomerulopathy (ORGN), anti-glomerular basement membrane disease (AGBM), lupus nephritis (LN), endocapillary proliferative glomerulonephritis (EPGN), membranoproliferative glomerulonephritis (MPGN), tubulointerstitial nephritis (TIN), crescentic glomerulonephritis (CrGN), IgA nephropathy (IgAN), and Henoch-Schönlein purpura nephritis (HSP). XJ-Light-1 serves as a foundational dataset for downstream tasks such as lesion identification, classification, and morphological segmentation (Table S44).

XJ-Light-2, by contrast, represents the second collection of light microscopy images from Xijing Hospital, encompassing CKD cases archived between 2022 and 2023. Unlike XJ-Light-1, it was curated chronologically at the slide level without detailed CKD subtype annotation, providing a large-scale resource for representation learning. This dataset comprises 14,049 slides stained with H&E, PAS, MT, and PASM, and was used exclusively for pretraining pathology foundation models, facilitating the development of robust, transferable feature representations for downstream analytic tasks.

To prevent potential data leakage between pretraining and downstream evaluation, the XJ-Light-1 and XJ-Light-2 datasets were constructed from strictly non-overlapping patient cohorts. All dataset splits were performed at the patient level, ensuring that no slides or specimens from the same patient appear in both datasets. This temporal and patient-level separation guarantees the independence of pretraining and downstream fine-tuning and evaluation.

4.1.2 | XJ-IF

The XJ-IF dataset comprises 3198 glomerular IF images collected at Xijing Hospital during routine diagnostic evaluations. Images were acquired by renal subspecialty pathologists at a resolution of $0.84 \mu\text{m}/\text{pixel}$ (10× magnification), allowing clear visualization

of deposition patterns and regional distributions. Among these, 1722 images were annotated with deposition region or deposition pattern subclasses, while 1476 remain unannotated, thus providing both labeled and unlabeled data for diverse analytic tasks (Tables S45 and S46). Representative examples are provided in the supplementary materials and illustrated in Figure S9.

4.1.3 | XJ-GIO

The XJ-GIO dataset was derived from glomerular images in the XJ-Light-1 cohort and constructed through a hybrid approach combining automated and expert annotation. Initially, the pre-trained Glo-In-One-v2 model [55] was applied to segment Bowman’s capsules and glomerular tufts. These preliminary masks were then manually refined by senior annotators to ensure accuracy and consistency. The resulting dataset contains 2675 high-quality annotations of Bowman’s capsules and tufts, serving as a robust resource for downstream morphological and structural analyses.

4.1.4 | AIDPATH-G

The AIDPATH-G dataset originates from the AIDPATH project [17, 24], which sought to advance digital pathology tools for glomerulosclerosis assessment. Renal biopsy samples were collected at Hospital Universitari de Girona Dr. Josep Trueta (Spain), Hospital de la Plana (Spain), and the National Pathology Center (Lithuania). All samples were PAS-stained and digitized at 20× magnification, producing 47 whole-slide images (WSIs). From these WSIs, 2340 glomerular images were extracted, evenly divided between normal (n = 1170) and sclerotic (n = 1170) glomeruli. Glomerular contours were annotated, cropped into individual entities, and stored in PNG format. To increase dataset diversity, augmentation techniques such as rotation, vertical flipping, and color jittering were applied. In this study, we refer to this curated dataset as AIDPATH-G. Representative examples are shown in Figure S10.

4.1.5 | KPMP-G

The Kidney Precision Medicine Project (KPMP) is a multi-institutional, longitudinal initiative aimed at advancing the mechanistic understanding of chronic kidney disease (CKD) and acute kidney injury (AKI) through integrated clinical, molecular, and digital pathology data. For this study, we leveraged a subset of the KPMP dataset comprising 1352 renal tissue slides with corresponding clinical metadata. From these slides, 28 242 glomeruli were systematically identified and extracted for downstream analyses, enabling clinicopathological correlation studies. Representative examples from KPMP-G are presented in Figure S11.

4.1.6 | XJ-CLI

The XJ-CLI dataset was curated to evaluate the clinical applicability of GloPath. It was derived from the open diagnostic

stream at the Department of Pathology, Xijing Hospital, reflecting unfiltered real-world biopsy practice. This dataset comprises 695 PAS-stained renal biopsy slides collected during routine clinical workflows, subsequently digitized into WSIs. Glomerular entities were automatically extracted using an entity detection pipeline, after which lesion annotations were generated following the same consensus-based workflow as that used on XJ-Light-1: two nephrologists (Yanxia Wang, and Xiaoqin Wang) independently annotated each glomerular entity, and a senior renal pathologist (Jing Li) adjudicated discrepant cases to produce the final reference labels. By capturing the heterogeneity and complexity of everyday clinical cases, XJ-CLI provides a stringent, forward-looking testbed for validating model robustness and translational value in real-world clinical environments (Table S2).

4.2 | Model Training

4.2.1 | Problem setting

Let $D_{pre} = \{X_i\}_{i=1}^N$ be a pretraining large-scale glomerular pathology image dataset, where denotes the first image. The goal of this study is to compute the mapping function $f : \mathcal{X} \rightarrow \mathcal{Z}$, where \mathcal{X} is the image space and \mathcal{Z} is the feature space. In the self-supervised pre-training phase, a training loss function \mathcal{L}_{pre} is defined to train the model to learn useful feature representations from glomerular pathology images, which can be expressed as:

$$\mathcal{L}_{pre}(\theta) = \mathbb{E}_{\mathcal{X} \sim D_{pre}} [\ell(f_{\theta}(X), g(X))] \quad (1)$$

Where f_{θ} is a pre-trained model with parameter θ , $g(X)$ is an auxiliary function used to generate the objective for the self-supervised task, and ℓ is a function of the loss of the contrast employed by the self-supervision, which is used to measure the discrepancy between the model output and the objective.

After training is completed, the optimal model f_{θ^*} is obtained, where $\theta^* = \operatorname{argmin}_{\theta} \mathbb{L}_{pre}(\theta)$. On the downstream task, let $D_{down} = \{(X_i, Y_i)\}_{i=1}^M$ be the dataset used for model performance validation, where (X_i, Y_i) denotes the first sample and its corresponding label. The goal is to fine-tune the model parameters so as to ensure that the model can be adapted to downstream tasks, which can be expressed as:

$$\theta_{down}^* = \operatorname{argmin}_{\theta} \mathbb{L}_{down}(\theta) = \operatorname{argmin}_{\theta} \mathbb{E}_{x, y \sim D_{down}} [\ell^{\downarrow}(f_{\theta^*}^{\downarrow}(x), y)] \quad (2)$$

where $f_{\theta^*}^{\downarrow}$ is the fine-tuned model and ℓ^{\downarrow} is a loss function that measures the difference between output probability and the ground truth.

4.2.2 | Entity-Centric Self-Supervised Pretraining

Glomeruli are intrinsically entity-based structures, in which local components (e.g., endothelial cells, basement membranes, podocytes) function synergistically to maintain blood filtration. From a pathological perspective, localized damage accumulates to impair overall glomerular function. This consistency between

global and local semantics provides a biologically grounded rationale for designing entity-centric self-supervised pretraining.

Rather than relying on reconstruction-based methods [56–58], we employed contrastive learning [59–63], which has proven effective in computational pathology [64, 65]. Specifically, we adopted the DINO framework, utilizing a teacher-student architecture with exponential moving average (EMA) updates. Both networks share a ViT-Base backbone. The teacher network maintains stable representations by slowly updating its weights using EMA, while applying a sharpening operation to prevent representation collapse, thereby stabilizing the student’s convergence. In the context of GloPath, this self-supervised pretraining allows the model to learn robust, invariant representations of glomerular pathology. By maximizing the similarity between augmented views of the same glomerulus, such as different lesions or whole glomerulus morphology, the model learns to capture fine-grained pathological features that are crucial for accurate diagnosis and disease progression analysis.

During pretraining, our framework introduces multi-scale and multi-view constraints (Figure 1b). Random augmentations generate global views (whole-glomerulus morphology) and local views (substructural or lesion-level features), enabling the model to integrate multi-scale pathological cues. Formally, given a glomerulus image I , we define two augmentation operators: a global-view augmentation \mathcal{A}_G and a local-view augmentation \mathcal{A}_L . The global-view augmentation \mathcal{A}_G preserves the entire glomerular entity and applies stochastic appearance transformations, while the local-view augmentation \mathcal{A}_L additionally incorporates random spatial cropping to capture substructural details using random cropping scales. In contrast to \mathcal{A}_G , which operates on the full glomerulus image, \mathcal{A}_L samples a spatial crop with a random scale factor $s \sim \mathcal{U}(s_{\min}, s_{\max})$. Based on these operators, we generate two global views $\{g^{(1)}, g^{(2)}\}$ via $\mathcal{A}_G(I)$, representing different whole-glomerulus views with varying appearance perturbations, and six local views $\{l^{(k)}\}_{k=1}^6$ via $\mathcal{A}_L(I)$, capturing diverse substructural or lesion-level patterns.

Training is guided by a cross-entropy-based contrastive loss:

$$\mathbb{L}_{pre} = \frac{1}{2n} \sum_i i = 1^n \left(- \sum_{j=1}^m \sum_{k=1}^d g_{m,k} \log(l_{i,k}) \right) \quad (3)$$

where n , m , and d denote the number of local views, global views, and feature dimensions, respectively. This formulation encourages discriminative representation learning and mitigates feature collapse, yielding embeddings suitable for clinical downstream tasks.

4.2.3 | Model Structure

Given the need for effective feature extraction and classification of glomerular pathology images, this study employs the ViT as the foundational structure for self-supervised pretraining [66]. To accommodate the processing of image data, the ViT model divides the input image into multiple fixed-size patches. These local patches are assigned patch tokens and linearly embedded

into a high-dimensional space. For each glomerular pathology image, the classification token (CLS token) is appended, serving as the carrier for the global pathological features of the glomerulus. This study utilizes the ViT-Base model, which comprises 12 Transformer encoder layers, each consisting of multi-head self-attention mechanisms and feed-forward neural networks.

This architecture enables the model to capture long-range dependencies when processing image data, thereby allowing it to identify complex patterns and structures within glomerular pathology images. Specifically, to enhance the ability to learn features at different scales, the ViT-Base model employs multi-head attention mechanisms. This allows the model to simultaneously focus on both local and global interactions between image patches. Since multi-head attention mechanisms can process features in multiple subspaces in parallel, this study leverages this mechanism to augment the model’s capability to capture multi-scale features in glomerular pathology images.

4.2.4 | Parameter Setting

Training was conducted with the following hyperparameters: momentum for EMA encoder = 0.996; temperature = 0.04; mixed-precision (16-bit) training; AdamW optimizer with learning rate 5^{-4} (minimum 1^{-6}) and 10-epoch warm-up. Weight decay was scheduled up to 0.4, with gradient clipping applied. For data augmentation, we adopted global cropping (two entity-level cropping, covering the full glomerulus) and local cropping (six crops; $s_{\min} = 0.4$ and $s_{\max} = 1$), combined with DropPath (rate 0.1). For data augmentation during pretraining, both \mathcal{A}_G and \mathcal{A}_L include a shared set of photometric and geometric transformations, such as random horizontal and vertical flipping, color jittering (brightness, contrast, saturation, hue), grayscale conversion, and Gaussian blur. Batch size was 64 per GPU, and training was performed for 200 epochs. These settings stabilized convergence and promoted generalization, enabling the pretrained GloPath model to deliver robust feature representations for downstream clinical applications.

4.3 | Evaluation

4.3.1 | Benchmark Models

To evaluate the parameter specificity and feature representation ability of the proposed GloPath, we benchmarked it against six representative models spanning three categories: (1) non-pathology foundation models, (2) general-purpose pathology foundation models, and (3) a renal pathology domain-specific foundation model.

The non-pathology foundation models include **RandomInit** and **ImageNetPre**. Both models are based on the ViT-Base architecture, which comprises 86.8 million parameters. RandomInit applies He initialization for linear layers and truncated normal initialization for positional encodings without pretraining [67]. ImageNetPre is pretrained on the large-scale ImageNet dataset [68], enabling transfer of generic visual features but without pathology-specific adaptation.

Three general-purpose pathology foundation models were selected: UNI, PLIP, and CONCH [69–71].

- **UNI** is a ViT-Large-based model (380 million parameters) pretrained with the DINOv2 framework [63] on the Mass-100K dataset, comprising more than 100 million tissue patches across 20 tissue types. This breadth enables rich feature learning, though not tailored to renal pathology.
- **PLIP** adapts the CLIP framework [72] to pathology by jointly training an image encoder and text encoder. Its training set includes 208 414 image–text pairs curated from social media and other sources, supporting multimodal alignment between histopathology images and natural language descriptions.
- **CONCH** also leverages contrastive vision–language learning, trained on 1.17 million image–caption pairs spanning H&E and immunohistochemistry (IHC) slides. This cross-modal setup allows the model to capture associations between morphological patterns and textual descriptors, providing broad but non-specialized pathology representations.

To provide a domain-specific baseline, we developed **RenalPath** using the same ViT-Base architecture and identical self-supervised pretraining strategy as GloPath. Pretraining was conducted on 14 049 renal biopsy WSIs, from which 224×224 image patches were extracted at $0.42 \mu\text{m}/\text{pixel}$, yielding 1 017 879 training samples. These patches encompassed diverse renal tissue components, including glomeruli, interstitial regions, and other renal structures, thereby reflecting the heterogeneity of renal pathology. Importantly, the only difference between RenalPath and GloPath lies in the definition of the pretraining unit and data organization: RenalPath follows a conventional patch-based paradigm over heterogeneous renal regions, whereas GloPath is pretrained exclusively on explicitly detected glomerular entities. This design enables a controlled comparison to assess the impact of entity-level pretraining on downstream performance.

4.3.2 | Evaluation Metrics

To ensure rigorous and clinically meaningful evaluation, task-specific metrics were applied according to dataset characteristics and prediction goals. For lesion recognition (Task A) and lesion grading (Task B), class imbalance posed a major challenge. We therefore prioritized the F1 score, which balances precision and recall, as the primary metric. In lesion recognition, we further incorporated the area under the precision–recall curve (PR-AUC) [73] and Receiver Operating Characteristic Curve (ROC-AUC) as a complementary measure. For cross-modality diagnosis (Task C) and few-shot based classification (Task D), the datasets were relatively balanced. Here, the emphasis was on discriminability across classes and ranking ability. Accordingly, the ROC-AUC was adopted as the primary evaluation metric, reflecting the models' ability to differentiate pathological categories with high sensitivity and specificity.

4.3.3 | Fully Supervised Fine-Tuning

Tasks A (lesion recognition), B (lesion grading), and C (cross-modality diagnosis) were implemented as full-parameter fine-

tuning tasks using a pre-trained glomerular pathology foundation model as the backbone. Task-specific output layers were defined according to clinically relevant categories. For lesion recognition (Task A), the classifier mapped features to binary outputs (lesion vs. non-lesion). For lesion grading (Task B), subclass-specific output nodes reflected established pathological categories, including four outputs for mesangial lesions (normal, mild, moderate, severe) and crescents (normal, cellular, cellular-fibrous, fibrous), and three outputs for membranoproliferative glomerulonephritis and endocapillary proliferation (normal, segmental, global). For cross-modality diagnosis (Task C), fluorescence-based outputs corresponded to capillary vs. mesangial deposition and focal vs. diffuse distribution (two outputs each).

Training was performed with cross-entropy loss and parameter updates by backpropagation. To account for class imbalance in lesion recognition, an imbalanced sampling strategy was applied to balance positive and negative samples within each epoch. To further align the pathology foundation model with fluorescence data, an additional light-weight fine-tuning step was conducted on 1476 unlabeled IF images prior to formal supervised training. Specifically, starting from the brightfield-pretrained GloPath, we performed a self-supervised adaptation using the same pretraining objective, during which only a subset of higher-level layers was updated while the core representation was largely preserved. This procedure resulted in an IF-finetuned GloPath that reduced modality-induced distribution shifts and enabled more consistent embedding alignment between brightfield and IF images.

All models were optimized with the Adam algorithm and trained for 50 epochs with standard data augmentation (random flipping, rotation, shearing, brightness/contrast adjustment) to enhance generalization. Model selection was based on the highest F1 score achieved on the validation set. To assess robustness, lesion recognition models were trained with three independent random seeds, while lesion subclassification and cross-modality tasks were evaluated with three-fold cross-validation. These procedures minimized evaluation bias and ensured reproducibility.

4.3.4 | Few-Shot Based Fine-Tuning

To further validate the model's parameter specificity for glomerular structures, this chapter conducts few-shot based tasks D, targeting the classification of normal and sclerotic glomeruli based on light microscopy images and the classification of deposits based on IF images. Under this setup, on the one hand, with only a very limited number of samples, the model needs to efficiently grasp the morphological essence of different categories based on the few perceived clues. On the other hand, the entity features of the glomeruli extracted by the model are fixed, and only the parameters of the classification end of the model are updated, which poses high demands on the accuracy of the initial feature construction of the model. First, different models extract the CLS token of the glomerular entity as the global deep semantic feature, which is only used as the input feature in the subsequent classification process and its value is no longer updated. Second, different methods are used to construct the classification end to classify the fixed input features. The experiment was repeated 10 times under different dataset splits. This study constructs five

machine learning algorithms to build a mapping from the initial feature space to the specific category space (Figure 1c):

- Support Vector Machine (SVM): The core idea of SVM is to find an optimal hyperplane to separate data of different categories. It uses the radial basis function as the kernel function to enhance the model's nonlinear classification capability.
- Logistic Regression (LR): LR maps the output of the linear combination to the interval (0,1) through the Sigmoid function to achieve probability estimation. The quasi-Newton method is used to optimize the logistic regression process.
- Multilayer Perceptron (MLP): MLP is a feedforward neural network composed of multiple hidden layers, which in this study includes hidden layers with 256 and 128 neurons. MLP uses ReLU as the activation function to introduce nonlinearity and Adam as the adaptive learning rate optimization algorithm, with the maximum number of iterations set to 200.
- Random Forest (RF): RF achieves glomerular classification by constructing multiple decision trees and aggregating their prediction results. This study uses Classification and Regression Trees (CART) to build decision trees and sets the number of decision trees in the forest to 100, which can balance underfitting and overfitting.
- Prototype Learning (PTL): PTL calculates the feature centers of each category based on the known k samples of each category in the training set, and then determines the category of each sample in the test set according to the distance between each sample and the feature centers of each category. Specifically, the feature center is usually obtained by calculating the mean of the samples of each category. During the test phase, for each test sample, the Euclidean distance between it and the feature centers of each category is calculated, and the sample is assigned to the nearest category. The characteristics of prototype learning are its simplicity, efficiency, and ease of understanding and implementation.

Importantly, the datasets used for self-supervised pretraining (XJ-Light-2) and few-shot evaluation (AIDPATH-G and XJ-IF) are fully independent at the patient level. XJ-Light-2 was collected between 2022–2023, whereas XJ-IF was acquired in 2024, and AIDPATH-G is a public dataset from external institutions. No slides or glomeruli from the same patient appear across pretraining and few-shot evaluation, ensuring the absence of data leakage.

4.3.5 | Large-Scale Real-World Study

To evaluate the clinical utility of GloPath in a real-world diagnostic context, we performed lesion-recognition task on the XJ-CLI dataset. For each lesion category, the model initialized with the best-performing weights from prior experiments was deployed without further fine-tuning, thereby ensuring independence between model development and real-world testing. Inference was then conducted across all cases in the XJ-CLI cohort, which consists of de-identified renal biopsy slides collected during routine diagnostic workflow.

Performance was primarily assessed using ROC-AUC, a metric robust to pronounced class imbalance and particularly informative in open-set scenarios where rare but clinically critical lesions may otherwise be underrepresented. By design, this evaluation simulates the conditions under which nephropathologists encounter heterogeneous biopsy material in daily practice. Importantly, all annotations were generated and verified by senior nephropathologists under institutional review board approval, further ensuring the reliability and translational relevance of the evaluation.

4.3.6 | Morphological Segmentation

The global feature representations learned by GloPath encode both overall and local structural information, which can be leveraged to generate fine-grained semantic segmentation of glomerular compartments. In this study, we adopted the Segmenter framework [47], which comprises an encoder and a decoder. The encoder employs the backbone of GloPath (and other models under comparison), with parameters frozen during training to retain the pre-trained feature representations. The decoder follows a MaskTransformer design, capturing global dependencies via multi-head self-attention and applying nonlinear transformations through feed-forward networks. A masking mechanism is applied to constrain attention to specific pixel regions, enhancing the focus on local structural details. In our experiments, the decoder was configured with 12 layers and 12 attention heads, and the model and feed-forward network dimensions were set to 192 and 768, respectively. This configuration allows Segmenter to effectively integrate global context and local features, dynamically weighting them to achieve precise segmentation of complex glomerular structures. Training was performed with a batch size of 32 using the Adam optimizer (learning rate = 1^{-4} ; weight decay = 1^{-4}) for up to 100 epochs.

In addition to comparing Segmentors based on different weights, we also compared GloPath with nnU-Net, the state-of-the-art method for medical image segmentation [48]. To ensure experimental fairness, nnU-Net adopted the exact same dataset split and training/validation strategy as GloPath, without introducing any additional prior information or manual parameter tuning. All model architecture configurations, preprocessing pipelines, and training hyperparameters were automatically determined by nnU-Net's automated planning and configuration mechanism, strictly following its official recommended settings. This ensures that the comparison reflects the intrinsic performance of a specialized supervised segmentation framework rather than benefits from task-specific manual optimization.

The model achieving the highest intersection-over-union (IoU) on the validation set was selected for performance evaluation. To assess robustness and reproducibility, three independent training runs were conducted with identical dataset splits but different random seeds.

4.3.7 | Clinicopathological Correlation Analysis

Morphological alterations in glomeruli, including glomerulosclerosis, mesangial proliferation, and podocyte injury, reflect under-

lying renal pathophysiology and directly influence kidney filtration and metabolic waste clearance. These structural features establish a clinically relevant link between renal histopathology and patient characteristics. In this study, we examined clinicopathological correlations using morphological segmentation derived from GloPath, RenalPath, and UNI. Clinical variables for the XJ-Light-1 and KPMP-G cohorts are summarized in Tables S47 and S48, while pathological indicators were constructed via semantic segmentation of Bowman's capsule and glomerular tuft, entity-level parameter quantification, and case-level aggregation (Table S49).

Segmentor models based on GloPath, RenalPath, and UNI were applied to generate Bowman's capsule and tuft masks for each glomerulus. From these masks, seven glomerular-level parameters were computed, and their median and mean values across all glomeruli in a case were used to obtain 14 aggregated case-level features per sample.

To evaluate the relationships between clinical and pathological indicators, we employed non-parametric statistical tests. The Kolmogorov–Smirnov (KS) test was used to compare the distributions of two independent samples, quantifying differences in their empirical cumulative distribution functions. The Kruskal–Wallis (KW) test was used to assess differences among three or more independent samples based on rank sums. Pathological parameters were treated as continuous variables, while clinical variables were categorized according to relevant clinical criteria. Importantly, because both the KS and KW tests are non-parametric, they do not require the data to satisfy normality assumptions or homogeneity of variance, making them suitable for clinical and pathological datasets that may be skewed or contain outliers. Statistical significance for each pairwise comparison was determined based on the corresponding p-values. The quantified effect sizes using epsilon-squared (ϵ^2) for the KW test and the KS statistic (D) for the KS test are reported to interpret the magnitude of clinicopathological associations.

In addition, to account for potential confounding factors, patient-level multivariate regression analyses were performed in the XJ-Light-1 cohort. For each patient, GloPath-derived glomerular morphological features were aggregated across all detected glomeruli using summary statistics. Three clinically relevant outcomes were analyzed: creatinine as a continuous variable, and lesion severity categorized as non-severe versus severe. Multivariate linear regression was used for continuous outcomes, while multivariate logistic regression was applied for binary outcomes. All models included sex, age, and disease type (IgAN, LN, MN, DN, and MCD) as covariates. Regression analyses were implemented using standard Python statistical libraries with default settings. Statistical significance was assessed using two-sided tests.

4.3.8 | Visualization

To evaluate the interpretability of GloPath and the rationality of its predictions, attention score and feature embedding visualizations were performed. For attention visualization, the multi-head self-attention mechanism in the ViT-Base backbone was utilized. Attention weights for each patch token were

aggregated across all heads and visualized according to their spatial positions, highlighting regions of the glomerulus that contribute most to the model's decision. This approach provides insight into GloPath's focus on lesion-related structures and supports the biological plausibility of its predictions. These visualizations provide saliency-level interpretability by highlighting diagnostically relevant regions, rather than explicit semantic reasoning over predefined pathological concepts. Future work may explore concept-level interpretability by integrating human-defined pathological concepts or structured lesion annotations, enabling a more explicit linkage between model decisions and established diagnostic criteria.

To further assess the discriminability of learned features across lesion categories, t-distributed stochastic neighbor embedding (t-SNE) was applied to the high-dimensional feature representations extracted from the CLS token. t-SNE reduces these features to a two-dimensional space, preserving the relative similarity between samples and enabling intuitive visualization of clustering patterns. This method allows qualitative evaluation of whether the model captures meaningful distinctions among different lesion types. All visualizations were generated using standard Python packages, ensuring reproducibility and clarity of the results.

To qualitatively assess cross-modality feature alignment, feature embeddings extracted from the CLS token were projected into a two-dimensional space using UMAP [74] approach with 50 nearest neighbors and a minimum inter-point distance of 0.1. This configuration balances local continuity with global structure preservation.

4.4 | Statistical Analysis

All statistical analyses were conducted to ensure the scientific rigor and reliability of the reported results. For lesion assessment, paired measurements were analyzed using the two-tailed Wilcoxon signed-rank test, a non-parametric method suitable for assessing differences between related samples without assuming normality. A p-value less than 0.05 was considered indicative of statistical significance.

For clinicopathological correlation analysis, all tests were performed in Python (version 3.12.4) using the `scipy.stats` library. The choice of statistical tests was based on data distribution characteristics and standard methodological assumptions; non-parametric tests were applied when normality assumptions were not met. Continuous variables were first assessed for normality using the Shapiro-Wilk test. When distributions deviated from normality, non-parametric tests were employed. Specifically, the KS test was used to evaluate differences in the empirical distributions of continuous pathological parameters between two groups, while the KW test was applied to compare multiple independent groups. For post-hoc pairwise comparisons following the KW test, Dunn's test with Bonferroni correction was used to control for type I error. Sample size for each statistical analysis is reported in the corresponding figure legends and tables.

All statistical tests were two-sided, with a significance threshold set at $\alpha = 0.05$. Exact p-values are reported wherever applicable.

Summary statistics are presented as mean \pm standard deviation for normally distributed variables, or median with interquartile range for non-normally distributed variables. This approach ensures transparency, reproducibility, and robust interpretation of the associations between model-derived pathological features and clinical variables.

4.5 | Computing Hardware and Software

All experiments were implemented using PyTorch (v2.5.0) with CUDA 12.4 (<https://pytorch.org>). Pretraining was performed on eight NVIDIA A100 GPUs (80 GB each) using a distributed data parallel configuration for multi-GPU, multi-node training. Downstream validation experiments utilized four NVIDIA GeForce RTX 3090 GPUs (20 GB each) and eight NVIDIA A6000 GPUs (40 GB each).

Whole-slide images in SDPC format were processed using the openSdpc library (<https://github.com/WonderLandxD/opensdpc>). Comparative models, including UNI, PLIP, and CONCH, were obtained from their respective public repositories. Image processing tasks were conducted using OpenCV-Python (v4.8.0) and Pillow (v9.5.0), while data manipulation was performed with NumPy (v1.22.4) and Pandas (v1.4.3). Visualization was achieved using Matplotlib (v3.7.2), Seaborn (v0.12.1), and OriginLab 2025.

For model construction, timm (v0.9.7; <https://huggingface.co>) was employed for classification networks, and the Segmenter library (<https://github.com/rstrudel/segmenter>) and nnU-Net2 library (<https://github.com/MIC-DKFZ/nnUNet>) were used for segmentation networks. Clinical-pathological correlation analyses and machine learning-based classifiers (SVM, LR, MLP, and RF) were implemented using scikit-learn (v1.3.2). The multivariate regression analyses were performed using python-statsmodels (<https://pypi.org/project/statsmodels/>). All software versions and hardware configurations are reported to ensure reproducibility.

Author Contributions

Qiming He, Jing Li, Tian Guan, Chao He, Zhe Wang, and Yonghong He conceived and designed the research project and experiments. Chao He, Zhe Wang, and Yonghong He coordinated the project and provided overall guidance. Qiming He, Jing Li, and Tian Guan drafted the manuscript. Qiming He performed data processing and model construction. Jing Li was responsible for dataset construction and annotation, as well as result analysis. Tian Guan completed the polishing of the manuscript. Shuang Ge, Yifei Ma, and Zimo Zhao contributed to the development of part of the code and analysis of experimental results. Yanxia Wang and Hongjing Chen completed dataset construction and partial annotation. Yingming Xu, Yizhi Wang, Xinrui Chen, Lianghui Zhu, Junru Cheng, Yexing Zhang, and Yiqing Liu assisted with some experiments and organized the code. Qingxia Hou, Shuyan Zhao, Xiaoqin Wang, and Lili Ma aided in dataset curation. Qiang Huang, Zihan Wang, Zhiyuan Shen, Siqi Zeng, Peizhen Hu, Jiurun Chen, and Zhen Song provided equipment and resource support.

Acknowledgements

We acknowledge the AIDPATH project for providing access to digital pathology resources that facilitated this study. We also thank the Kidney

Precision Medicine Project (KPMP), which is funded by the grants from the NIDDK, for providing freely available clinical and histology slide data upon which parts of the results in this study are based. The work was supported by National Natural Science Foundation of China (NSFC) (82430062), the Shenzhen Engineering Research Centre (XMHT20230115004), and Tsinghua Shenzhen International Graduate School Cross-disciplinary Research and Innovation Fund Research Plan (JC2024002). The work was also supported by Key Research and Development Program of Shaanxi Province (2024GX-YBXM-123). We also thank the Jilin FuyuanGuan Food Group Co., Ltd.

Qiming He, Jing Li, and Tian Guan contribute equally to this work.

Conflicts of Interest

The authors declare no conflicts of interests.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions. The public external image and clinical data used in this study are available in the KPMP (<https://atlas.kpmp.org/repository>) and AIDPATH (<https://aidpath.eu>). Code and model weights for GloPath can be accessed for academic research purposes at <https://github.com/jiangnanss/Code-for-GloPath>.

References

1. A. Francis, M. N. Harhay, A. C. Ong, et al., "Chronic Kidney Disease and the Global Public Health Agenda: An International Consensus," *Nature Reviews Nephrology* 20, no. 7 (2024): 473.
2. D. Khadzhynov, D. Schmidt, J. Hardt, et al., "The Incidence of Acute Kidney Injury and Associated Hospital Mortality: A Retrospective Cohort Study of Over 100 000 Patients at Berlin's Charité Hospital," *Deutsches Ärzteblatt International* 116, no. 22 (2019): 397.
3. I. V. Murray and M. A. Paolini, *Histology, Kidney and Glomerulus* (2023).
4. A. B. Fogo, "The Glomerulus Reveals Some Secrets," *Nature Reviews Nephrology* 11, no. 11 (2015): 633.
5. M. R. Pollak, S. E. Quaggin, M. P. Hoenig, and L. D. Dworkin, "The Glomerulus: The Sphere of Influence," *Clinical Journal of the American Society of Nephrology* 9, no. 8 (2014): 1461.
6. A. B. Fogo and R. C. Harris, "Crosstalk Between Glomeruli and Tubules," *Nature Reviews Nephrology* 21, 3 (2025): 1–11.
7. G. Clair, H. Soloyan, P. Cravedi, et al., "The Spatially Resolved Transcriptome Signatures of Glomeruli in Chronic Kidney Disease," *JCI Insight* 9, no. 6 (2024): e165515.
8. B. A. Santo, A. Z. Rosenberg, and P. Sarder, "Artificial Intelligence Driven Next-Generation Renal Histomorphometry," *Current Opinion in Nephrology and Hypertension* 29, no. 3 (2020): 265.
9. Z. Zheng, X. Zhang, J. Ding, et al., "Deep Learning-Based Artificial Intelligence System for Automatic Assessment of Glomerular Pathological Findings in Lupus Nephritis," *Diagnostics* 11, no. 11 (2021): 1983.
10. Q. Lei, X. Hou, X. Liu, et al., "Artificial Intelligence Assists Identification and Pathologic Classification of Glomerular Lesions in Patients with Diabetic Nephropathy," *Journal of Translational Medicine* 22, no. 1 (2024): 397.
11. R. Deng, T. Yao, Y. Tang, et al., "Kpis 2024 Challenge: Advancing Glomerular Segmentation from Patch-to Slide-Level," *arXiv Preprint arXiv:2502.07288* (2025).
12. M. Ochi, D. Komura, and S. Ishikawa, "Pathology Foundation Models," *JMA Journal* 8, no. 1 (2025): 121.
13. G. Campanella, S. Chen, M. Singh, et al., "A Clinical Benchmark of Public Self-Supervised Pathology Foundation Models," *Nature Communications* 16, no. 1 (2025): 3640.

14. N. Aben, E. D. de Jong, I. Gatopoulos, et al., "Towards Large-Scale Training of Pathology Foundation Models," *arXiv Preprint arXiv:2404.15217* (2024).
15. J. Lee, J. Lim, K. Byeon, and J. T. Kwak, "Benchmarking Pathology Foundation Models: Adaptation Strategies and Scenarios," *Computers in Biology and Medicine* 190 (2025): 110031.
16. M. Bilal, M. Raza, Y. Altherwy, et al., "Foundation Models in Computational Pathology: A Review of Challenges, Opportunities, and Impact," *arXiv Preprint arXiv:2502.08333* (2025).
17. G. Bueno, L. Gonzalez-Lopez, M. Garcia-Rojo, A. Laurinavicius, and O. Deniz, "Data for Glomeruli Characterization in Histopathological Images," *Data in Brief* 29 (2020): 105314.
18. M. Tinawi, "Update on the Etiology, Classification, and Management of Glomerular Diseases," *Avicenna Journal of Medicine* 10, no. 02 (2020): 61.
19. G. Moroni and C. Ponticelli, "Rapidly Progressive Crescentic Glomerulonephritis: Early Treatment Is a Must," *Autoimmunity Reviews* 13, no. 7 (2014): 723.
20. R. P. Woroniecki and H. W. Schnaper, In *Seminars in Nephrology*, vol. 29 (Elsevier, 2009), 412–424.
21. N. Messias, "Immunofluorescence Use and Techniques in Glomerular Diseases: A Review," *Glomerular Diseases* 4, no. 1 (2024): 227.
22. V. Agrawal and A. Sharma, "Diagnostic Immunostaining of Renal Biopsies: An Overview of Markers for Glomerular Diseases," *Glomerular Diseases* 5, no. 1 (2025): 176.
23. J. Song, X. Cui, S. Yuan, et al., "Immunofluorescence Staining Profiles of Glomerular Diseases: A Single-Center Retrospective Study," *Available at SSRN 5168817*.
24. G. Bueno, M. M. Fernandez-Carrobles, L. Gonzalez-Lopez, and O. Deniz, "Glomerulosclerosis Identification in Whole Slide Images Using Semantic Segmentation," *Computer Methods and Programs in Biomedicine* 184 (2020): 105273.
25. M. B. Palmer, V. Royal, J. C. Jennette, et al., "Cure Glomerulonephropathy Pathology Classification and Core Scoring Criteria, Reproducibility, and Clinicopathologic Correlations," *Glomerular Diseases* 3, no. 1 (2023): 248.
26. S. A. van Eeghen, L. Pyle, P. Narongkiatikhun, et al., "Unveiling Mechanisms Underlying Kidney Function Changes During Sex Hormone Therapy," *The Journal of Clinical Investigation* 135 (2025): 9.
27. A. Denic, R. J. Glasscock, and A. D. Rule, "Structural and Functional Changes with the Aging Kidney," *Advances in Chronic Kidney Disease* 23, no. 1 (2016): 19.
28. Y. Li, X. Yu, W. Zhang, et al., "Epidemiological Characteristics and Pathological Changes of Primary Glomerular Diseases," *PLoS One* 17, no. 8 (2022): e0272237.
29. A. Denic, J. Mathew, V. V. Nagineni, et al., "Clinical and Pathology Findings Associate Consistently with Larger Glomerular Volume," *Journal of the American Society of Nephrology* 29, no. 7 (2018): 1960.
30. D. L. Hölscher, N. Bouteldja, M. Joodaki, et al., "Next-Generation Morphometry for Pathomics-Data Mining in Histopathology," *Nature Communications* 14, no. 1 (2023): 470.
31. S. P. Border, J. E. Tomaszewski, T. Yoshida, et al., "Investigating Quantitative Histological Characteristics in Renal Pathology Using Histolens," *Scientific Reports* 14, no. 1 (2024): 17528.
32. B. Shickel and A. Bihorac, "The Dawn of Multimodal Artificial Intelligence in Nephrology," *Nature Reviews Nephrology* 20, no. 2 (2024): 79.
33. V. W. Berger and Y. Zhou, "Kolmogorov–Smirnov Test: Overview," *Wiley StatsRef: Statistics Reference Online* (2014).
34. P. E. McKight and J. Najab, "Kruskal-Wallis Test," *The Corsini Encyclopedia of Psychology* (2010): 1–1.
35. S. Kannan, L. A. Morgan, B. Liang, et al., "Segmentation of Glomeruli Within Trichrome Images Using Deep Learning," *Kidney International Reports* 4, no. 7 (2019): 955.
36. J. Silva, L. Souza, P. Chagas, et al., "Boundary-Aware Glomerulus Segmentation: Toward One-to-Many Stain Generalization," *Computerized Medical Imaging and Graphics* 100 (2022): 102104.
37. N. Altini, G. D. Cascarano, A. Brunetti, et al., "Semantic Segmentation Framework for Glomeruli Detection and Classification in Kidney Histological Sections," *Electronics* 9, no. 3 (2020): 503.
38. L. Jiang, W. Chen, B. Dong, et al., "A Deep Learning-Based Approach for Glomeruli Instance Segmentation from Multistained Renal Biopsy Pathologic Images," *The American Journal of Pathology* 191, no. 8 (2021): 1431.
39. F. N. Saikia, Y. Iwahori, T. Suzuki, M. K. Bhuyan, A. Wang, and B. Kijisirikul, "Mlp-Unet: Glomerulus Segmentation," *IEEE Access* 11 (2023): 53034.
40. T. Yao, Y. Lu, J. Long, et al., "Glo-in-One: Holistic Glomerular Detection, Segmentation, and Lesion Characterization with Large-Scale Web Image Mining," *Journal of Medical Imaging* 9, no. 5 (2022): 052408.
41. G. M. Dimitri, P. Andreini, S. Bonechi, et al., "Deep Learning Approaches for the Segmentation of Glomeruli in Kidney Histopathological Images," *Mathematics* 10, no. 11 (2022): 1934.
42. X. Li, R. C. Davis, Y. Xu, et al., "Deep Learning Segmentation of Glomeruli on Kidney Donor Frozen Sections," *Journal of Medical Imaging* 8, no. 6 (2021): 067501.
43. Y. Liu, "A Hybrid CNN-TransXNet Approach for Advanced Glomerular Segmentation in Renal Histology Imaging," *International Journal of Computational Intelligence Systems* 17, no. 1 (2024): 126.
44. P. Andreini, S. Bonechi, and G. M. Dimitri, "Enhancing Glomeruli Segmentation Through Cross-Species Pre-Training," *Neurocomputing* 563 (2024): 126947.
45. N. Bouteldja, B. M. Klinkhammer, R. D. Bülow, et al., "Deep Learning-Based Segmentation and Quantification in Experimental Kidney Histopathology," *Journal of the American Society of Nephrology* 32, no. 1 (2021): 52.
46. Q. He, S. Zeng, S. Ge, et al., "Identifying and Matching 12-Level Multistained Glomeruli via Deep Learning for Diagnosis of Glomerular Diseases," *International Journal of Imaging Systems and Technology* 34, no. 2 (2024): e23032.
47. R. Strudel, R. Garcia, I. Laptev, and C. Schmid, In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), 7262–7272.
48. F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "Nnu-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation," *Nature Methods* 18, no. 2 (2021): 203.
49. C. Messow, K. Gärtner, H. Hackbarth, M. Kangaloo, and L. Lünebrink, "Sex Differences in Kidney Morphology and Glomerular Filtration Rate in Mice.," *Contributions to Nephrology* 19 (1980): 51.
50. J. Nyengaard and T. Bendtsen, "Glomerular Number and Size in Relation to Age, Kidney Weight, and Body Surface in Normal Man," *The Anatomical Record* 232, no. 2 (1992): 194.
51. M. Canney, H. M. Gunning, Y. Zheng, et al., "The Risk of Cardiovascular Events in Individuals with Primary Glomerular Diseases," *American Journal of Kidney Diseases* 80, no. 6 (2022): 740.
52. S. Kishi, H. Nagasu, K. Kidokoro, and N. Kashihara, "Oxidative Stress and the Role of Redox Signalling in Chronic Kidney Disease," *Nature Reviews Nephrology* 20, no. 2 (2024): 101.
53. J. M. J. Valanarasu, H. Xu, N. Usuyama, et al., "Multimodal AI Generates Virtual Population for Tumor Microenvironment Modeling," *Cell* (2025).
54. Z. Li, Y. Li, J. Xiang, et al., "AI-Enabled Virtual Spatial Proteomics from Histopathology for Interpretable Biomarker Discovery in Lung Cancer," *Nature Medicine* (2026): 1–14.

55. L. Yu, M. Yin, R. Deng, et al., “Glo-in-One-V2: Holistic Identification of Glomerular Cells, Tissues, and Lesions in Human and Mouse Histopathology,” *arXiv Preprint arXiv:2411.16961* (2024).
56. K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), 16000–16009.
57. H. Bao, L. Dong, S. Piao, and F. Wei, “Beit: Bert Pre-Training of Image Transformers,” *arXiv Preprint arXiv:2106.08254* (2021).
58. X. Li, Z. Qu, S. Zhao, B. Tang, Z. Lu, and Y. Liu, “Lomar: A Local Defense Against Poisoning Attack on Federated Learning,” *IEEE Transactions on Dependable and Secure Computing* 20, no. 1 (2021): 437.
59. K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), 9729–9738.
60. X. Chen and K. He, In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), 15750–15758.
61. J.-B. Grill, F. Strub, F. Althé, et al., “Bootstrap Your Own Latent—A New Approach to Self-Supervised Learning,” *Advances in Neural Information Processing Systems* 33 (2020): 21271.
62. M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, “Unsupervised Learning of Visual Features by Contrasting Cluster Assignments,” *Advances in Neural Information Processing Systems* 33 (2020): 9912.
63. M. Oquab, T. Darcet, T. Moutakanni, et al., “Dinov2: Learning Robust Visual Features Without Supervision,” *arXiv Preprint arXiv:2304.07193* (2023).
64. N. A. Koohbanani, B. Unnikrishnan, S. A. Khurram, P. Krishnaswamy, and N. Rajpoot, “Self-Path: Self-Supervision for Classification of Pathology Images with Limited Annotations,” *IEEE Transactions on Medical Imaging* 40, no. 10 (2021): 2845.
65. X. Wang, S. Yang, J. Zhang, et al., In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24* (Springer, 2021), 186–195.
66. A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., “An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale,” *arXiv Preprint arXiv:2010.11929* (2020).
67. K. He, X. Zhang, S. Ren, and J. Sun, In *Proceedings of the IEEE International Conference on Computer Vision* (2015), 1026–1034.
68. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009), 248–255.
69. R. J. Chen, T. Ding, M. Y. Lu, et al., “Towards a General-Purpose Foundation Model for Computational Pathology,” *Nature Medicine* 30, no. 3 (2024): 850.
70. Z. Huang, F. Bianchi, M. Yuksekgonul, T. J. Montine, and J. Zou, “A Visual–Language Foundation Model for Pathology Image Analysis Using Medical Twitter,” *Nature Medicine* 29, no. 9 (2023): 2307.
71. M. Y. Lu, B. Chen, D. F. Williamson, et al., “A Visual–Language Foundation Model for Computational Pathology,” *Nature Medicine* 30, no. 3 (2024): 863.
72. A. Radford, J. W. Kim, C. Hallacy, et al., In *International Conference on Machine Learning* (PMLR, 2021), 8748–8763.
73. H. R. Sofaer, J. A. Hoeting, and C. S. Jarnevich, “The Area Under the Precision–Recall Curve as a Performance Metric for Rare Binary Events,” *Methods in Ecology and Evolution* 10, no. 4 (2019): 565.
74. L. McInnes, J. Healy, and J. Melville, “Umap: Uniform Manifold Approximation and Projection for Dimension Reduction,” *arXiv Preprint arXiv:1802.03426* (2018).

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File: advs74818-sup-0001-SuppMat.pdf.