Proceedings of ICBioMed 2026 Symposium: Biomedical Engineering and Smart Medicine
DOI: 10.54254/2753-8818/2026.HZ36589

AI-Based Histopathological Diagnosis of Ameloblastoma: Research Progress, Limitations and Prospects

Han Dai

School of Stomatology, Southern Medical University, Guangzhou, China
13684929320@163.com

Abstract. Locally aggressive and prone to recurrence, ameloblastoma (AME) is among the most common odontogenic epithelial tumors of the jaws. Histopathological diagnosis is often difficult because AME morphologically overlaps with other jaw lesions; where specialist expertise is scarce, misdiagnosis can lead to suboptimal surgery and poorer outcomes. This narrative review examines current evidence on AI-assisted histopathological diagnosis of AME. We describe the deep learning workflow and evaluate three tasks—pixel-level tumor epithelium segmentation, benign-malignant stratification, and multi-class differential diagnosis—with attention to intra-cohort comparisons. Reported results are encouraging: three-class classification reaches AUC 0.91, and tumor epithelium segmentation attains mean Intersection over Union (mIoU) 0.846, albeit in single-center retrospective settings. Clinical translation is hindered by cross-dataset domain shift, absent benchmarks, limited interpretability, and privacy constraints. We also outline an "AI screening plus pathologist review" arrangement that could improve efficiency. AI offers genuine promise as an adjunct, but progress will require standardized multi-center datasets, more interpretable algorithms, and prospective validation.

Keywords: artificial intelligence, deep learning, U-Net, ameloblastoma, histopathological diagnosis

1. Introduction

Ameloblastoma (AME) is among the most prevalent odontogenic epithelial tumors of the jaws, accounting for approximately 10% of mandibular and maxillary neoplasms [1]. Although histologically benign, it exhibits locally aggressive behavior and a notable risk of postoperative recurrence [1], making accurate diagnosis critical for patient prognosis. Histopathological examination is the gold standard for diagnosing jaw cysts and odontogenic tumors [2]. However, a substantial proportion of AME cases are initially misdiagnosed as lesions such as odontogenic keratocyst (OKC) due to overlapping clinical and radiological features, particularly in resource-limited settings with few specialized oral pathologists [1]. While clinical and radiological findings support provisional diagnosis, definitive classification depends on histopathological assessment, which is subject to appreciable inter-observer variability [3].

The widespread adoption of whole slide imaging (WSI) has lowered technical barriers to AI-driven computational pathology. In breast, lung, and gastrointestinal pathology, deep learning

DOI: 10.54254/2753-8818/2026.HZ36589

systems have achieved board-certified pathologist-level performance in tumor detection and benign-malignant stratification, with multiple clinical-grade products approved for routine use [4]. By contrast, AI in oral and maxillofacial pathology (OMFP) remains underdeveloped, with most studies relying on conventional machine learning [5]. As a niche subspecialty with rare disease entities, uneven pathologist distribution, and a lack of standardized multi-center datasets, OMFP has produced only small-scale exploratory deep learning studies on AME, mostly single-center retrospective works with limited samples [5]. In this review, we detail the standard AI workflow for digital histopathology, summarize current deep learning applications in AME identification and differential diagnosis, discuss key barriers to clinical deployment, and propose future directions.

2. AI workflow for digital histopathological analysis

2.1. Mainstream deep learning models

Computational histopathology research builds on two categories of deep learning architectures, complemented by training strategies for insufficient annotated samples [6]. The first comprises CNN classification backbones (VGG16, ResNet, DenseNet), which output slide-level diagnostic labels without pixel-level localization, primarily for lesion classification and benign-malignant stratification.

The second category is task-specific segmentation networks, represented by U-Net. First proposed at MICCAI 2015, U-Net adopts a symmetric encoder-decoder architecture with skip connections fusing low-level epithelial contour and high-level tumor semantic features, mitigating boundary blurring from downsampling. It has become the de facto baseline for medical image segmentation, including histopathology [7], with reliable performance on AME and periapical cyst WSIs validated in single-center cohorts [8, 9]. Given the scarcity of annotated histology samples, transfer learning leverages ImageNet pre-trained representations to alleviate overfitting from random initialization [10]; beyond natural image pre-training, pathology pre-trained models (e.g., PathNet) can further capture histology-specific patterns [10]. Ablation studies show most CNN architectures yield improved AUC and Dice after pre-training, and pre-trained backbones can be embedded into U-Net encoders to enhance segmentation [10]. In summary, CNN backbones suit slide-level classification and subtyping, U-Net specializes in pixel-level segmentation, and transfer learning is a critical performance booster in data-limited oral pathology.

2.2. Dataset construction and preprocessing

Formalin-fixed paraffin-embedded tissues are sectioned at 4 μm, stained with hematoxylin and eosin (H&E), and scanned at 20×; faded, damaged, and tissue-deficient slides are excluded at quality control [8, 10]. Senior pathologists delineate tumor-positive and stromal regions of interest (ROIs) following WHO criteria, with all ROIs independently labeled by at least two senior oral pathologists and inter-observer consistency quantified by intraclass correlation coefficient (ICC) or Cohen's Kappa. Datasets are partitioned 8:1:1 at the case level—all patches from the same patient assigned to one subset—to prevent data leakage. Preprocessing comprises: (1) tissue masking to remove background, (2) color normalization (e.g., Macenko) to offset staining batch variation, (3) patch extraction to tile WSIs for network input, and (4) augmentation (rotation, flipping, color jittering) to mitigate overfitting.

DOI: 10.54254/2753-8818/2026.HZ36589

2.3. Training hyperparameters and quantitative metrics

Common training configurations include the Adam optimizer (learning rate 1×10^{-4} to 1×10^{-5}), cross-entropy loss for classification and Dice loss for segmentation, with batch size adjusted to GPU memory. Classification metrics include Area Under the Receiver Operating Characteristic Curve (AUC), F1-score, and balanced accuracy. AUC >0.85 indicates good discriminative performance [8, 9, 11]; F1-score balances false-negative and false-positive predictions [5, 9]; balanced accuracy mitigates class imbalance bias. Segmentation metrics include mIoU and Dice coefficient (Dice = 2×IoU/(1+IoU)), with values closer to 1 indicating greater overlap with ground truth. AME positive tumor region mIoU reaches 0.846 [8], and periapical cyst lining epithelium Dice reaches 0.685 [9].

Unified formulas theoretically enable cross-model comparison; however, multi-center WSIs suffer intrinsic domain shift from scanner differences, staining variation, inconsistent annotation, and uneven cohort distribution [6, 11]. No universal benchmark exists, and direct cross-study extraction of AUC or Dice may yield biased conclusions. Intra-dataset ablation under unified preprocessing, hyperparameters, and metrics provides the most reliable basis for horizontal comparison [11]. Figure 1 illustrates the overall AI workflow.

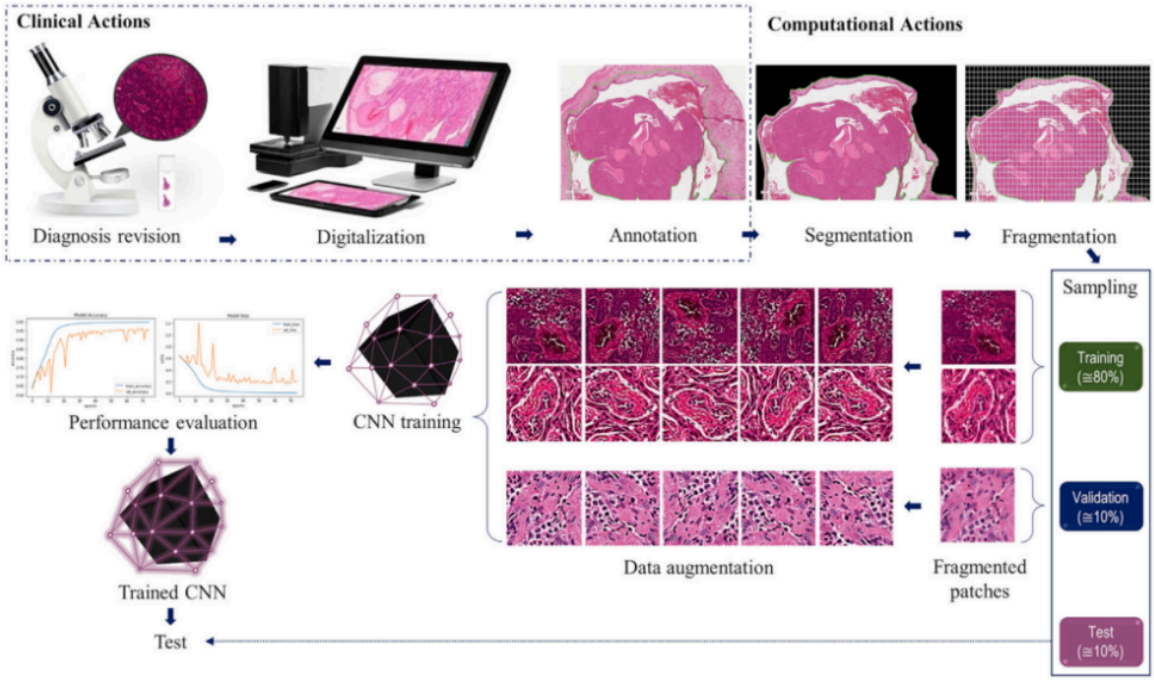

Figure 1. AI workflow for digital histopathological analysis

3. Practical application of AI in AME histopathological diagnosis

3.1. Clinical scenarios and fair cross-model evaluation

CNN architectures, often combined with transfer learning, have produced several clinically promising pipelines for AME histopathological analysis. These efforts cluster around three directions.

Of the three directions, pixel-wise segmentation of tumor epithelium has received the most attention. Both vanilla U-Net and encoder-modified variants with pre-trained CNN backbones are commonly used as baselines [5, 8]; one single-center study of 90 AME WSIs reported mIoU 0.846

DOI: 10.54254/2753-8818/2026.HZ36589

for positive tumor regions, with the model capturing delicate follicular contours [8]. Binary benign-malignant classification is another focus, where ResNet, VGG16, and DenseNet variants distinguish conventional AME from ameloblastic carcinoma (AC)—Giraldo-Roldán et al. [12] reported ResNet50 accuracy of 75% and F1-score 0.77. Histological subtyping (follicular versus plexiform) trails behind, supported only by small pilot studies. Transfer learning has proven helpful here: EfficientNetV2B0 reached AUC 0.91 on the three-way adenomatoid odontogenic tumor (AOT)/AME/AC task when transfer learning was applied [10].

Caution is needed when comparing metrics across studies, since differences in scanners, staining protocols, annotation criteria, and patient populations all introduce domain shift. With no unified public AME WSI benchmark available, AUC or Dice values taken from separate publications cannot reliably rank algorithms [11]. The strongest evidence instead comes from intra-dataset ablation, where preprocessing, hyperparameters, and metrics are held fixed while architectures are tested head-to-head. Table 1 lists representative results.

Table 1. Deep learning model performance for AME histopathological WSI analysis (H&E)

Ref.	Architecture	Task	TL	AUC	BA	ACC	SE	SP
[8]	U-Net	AME segmentation	No	0.92	NR	NR	NR	NR
[12]	ResNet50	AME vs AC	NR	NR	NR	75%	0.84	0.65
	DenseNet	AME vs AC	NR	NR	NR	61%	0.44	0.78
	VGG16	AME vs AC	NR	NR	NR	65%	0.61	0.68
[10]	EfficientNetV2B0	AOT/AME/AC	Yes	0.91	0.81	0.79	0.74	0.88
	DenseNet121	AOT/AME/AC	Yes	0.85	0.74	0.72	0.64	0.84
	EfficientNetV2B0	AOT/AME/AC	No	NR	0.68	0.67	0.54	NR
	DenseNet121	AOT/AME/AC	No	0.78	0.74	0.73	0.65	0.84

Note: TL=Transfer Learning, BA=Balanced Accuracy, ACC=Accuracy, SE=Sensitivity, SP=Specificity, NR=Not Reported. Data are derived from intra-dataset ablation experiments within each independent cohort; direct cross-study comparison is not recommended due to inherent domain shift across datasets.

3.2. AI-mediated differential diagnosis and clinical implications

Two lesions in particular—OKC and AC—overlap morphologically with AME on H&E-stained slides under the 5th WHO Classification [3]. OKC displays 5–8 layers of stratified squamous lining epithelium with palisaded basal cells, while AC preserves ameloblastic architecture but shows conspicuous atypia and mitotic figures [3]. Even experienced pathologists achieve only moderate agreement on manual review (Kappa 0.52–0.71) [1, 3]. The treatment implications are substantial: conventional AME calls for wide local resection, OKC for enucleation or marsupialization, and AC for radical jaw resection with neck dissection when indicated [1, 2]. An incorrect diagnosis thus has direct therapeutic consequences.

A hierarchical pipeline can be built around the sequence "feature recognition–lesion stratification–treatment reference." At the feature level, transfer-learning VGG16 recognizes OKC palisaded basal epithelium (AUC=0.997) [13]; ResNet50 separates benign AME from malignant AC [12]; and U-Net outlines AME epithelial boundaries [8]. Preoperative radiological AI adds further reference information [14, 15]. The resulting "AI screening plus pathologist review" model applies AI to straightforward, high-confidence cases and refers ambiguous ones to senior pathologists,

DOI: 10.54254/2753-8818/2026.HZ36589

potentially raising throughput and reducing bias—an advantage especially welcome in under-resourced settings. Whether these models generalize beyond their training cohorts remains an open question, however, as nearly all are trained on single-center retrospective data and large prospective trials are still lacking.

4. Challenges and future directions

4.1. Dataset limitations

Perhaps the most immediate barrier to generalization is the lack of standardized quality control in AME datasets. WSIs from different centers vary noticeably in color, contrast, and resolution, and without an agreed threshold for resolving epithelial boundaries, even follicular tumor segmentation can become unreliable [10]. Color normalization offers only partial relief. Then there is the sample problem: most cohorts enroll under 100 patients from a single institution, leaving gaps in ethnic diversity and subtype coverage [8], while AME's rarity and the annotation burden make scaling difficult. Retrospective test sets introduce further bias by over-representing typical cases, whereas real practice includes more atypical, borderline, or inflammation-obscured specimens on which performance is largely unknown [10].

4.2. Algorithm limitations and insufficient clinical validation

A second set of problems concerns the algorithms themselves. There is a noticeable gap between accuracy and deployability. Lightweight models (MobileNet, VGG16) reach only ~0.65 and ~0.63 balanced accuracy on three-class tasks and drift across centers; EfficientNetV2B0 performs better (AUC 0.91) but needs hardware few primary hospitals have [10]. Opacity is a further concern—most end-to-end models cannot indicate which regions drove a prediction, leaving clinicians unable to verify outputs. Gradient-weighted Class Activation Mapping (Grad-CAM) can highlight attention regions but remains an interim solution [11]. Most importantly, rigorous clinical evidence is lacking: nearly all models are evaluated retrospectively, multi-center prospective trials are absent [5, 12], and standardized validation protocols and regulatory pathways are not yet in place.

4.3. Data storage and privacy risks

Storage and sharing practices add another layer of difficulty. A single WSI can occupy several gigabytes, and without consistent archiving or a common metadata schema, multi-modal training opportunities are limited [8, 11]. Privacy regulations make cross-institutional collaboration harder still: histopathological images are sensitive health data, and compliance rules often prevent their transfer, leaving nearly all models trained on single-center data—DenseNet121's AUC dropped >10% on external images [10]. Federated learning circumvents this by training across institutions without moving raw files. Data governance nevertheless lags behind: de-identification is not standardized, and re-identification risk from distinctive morphology remains inadequately addressed [11].

4.4. Future development directions

How, then, should the field move forward? Multi-center standardized datasets are the obvious priority—harmonized tissue processing and scanning, data spanning all subtypes and stages, and de-identified repositories along the lines of TCGA or CAMELYON17 [10, 11]. Cross-center

DOI: 10.54254/2753-8818/2026.HZ36589

generalization must also be tackled directly through heterogeneous training data, domain adaptation, and federated learning; a domain-adapted mIoU >0.8 for AME segmentation would be a meaningful target [8]. On the model side, the challenge is combining lightweight deployment with accuracy—cascaded U-Net/ResNet designs and attention-based heat maps are promising, and pathological prior knowledge may boost robustness [5, 8]. Validation needs to mature in parallel: benchmark test sets with mixed cases, multi-center prospective trials, and regulatory-compliant procedures are overdue [10]. Ultimately, adoption depends on whether tools fit real workflows—embedded in digital slide platforms and Laboratory Information System (LIS) in ways that support pathologists rather than disrupting their routines.

5. Conclusion

Deep learning shows promising auxiliary diagnostic value for AME histopathology, with competitive performance in segmentation, benign-malignant stratification, and differential diagnosis in single-center retrospective cohorts. However, current research is constrained by single-center designs limiting generalizability, absent standardized benchmarks impeding fair comparison, and insufficient interpretability limiting clinical trust. The present review is likewise limited by the small number of eligible primary studies, most of which are small-sample and retrospective. Future priorities include multi-center dataset standardization, interpretable algorithm optimization, and large-scale prospective validation. With advances in federated learning and explainable AI, these technologies may gradually integrate into routine practice, improving diagnostic accuracy and outcomes for AME patients.

References

[1] Ghai, S. (2022). Ameloblastoma: An updated narrative review of an enigmatic tumor. *Cureus, 14*(8), e27734. https://doi.org/10.7759/cureus.27734

[2] Mendes, R. A., Carvalho, J. F., & van der Waal, I. (2010). Characterization and management of the keratocystic odontogenic tumor in relation to its histopathological and biological features. *Oral Oncology, 46*(4), 219–225. https://doi.org/10.1016/j.oraloncology.2010.01.012

[3] Soluk-Tekkesin, M., & Wright, J. M. (2022). The World Health Organization classification of odontogenic lesions: A summary of the changes of the 2022 (5th) edition. *Turk Patoloji Dergisi, 38*(2), 168–184. https://doi.org/10.5146/tjpath.2022.01573

[4] Niazi, M. K., et al. (2019). Digital pathology and artificial intelligence. *Lancet Oncology, 20*(5), e253–e261. https://doi.org/10.1016/S1470-2045(19)30154-8

[5] Khanagar, S. B., et al. (2021). Developments, application, and performance of artificial intelligence in dentistry-a systematic review. *Journal of Dental Sciences, 16*(1), 508–522. https://doi.org/10.1016/j.jds.2020.06.019

[6] Litjens, G., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

[7] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. arXiv:1505.04597.

[8] Qiao, X., Li, M., Shen, Z., et al. (2025). Preliminary application of artificial intelligence in the pathological diagnosis of ameloblastoma. *Chinese Journal of Oral and Maxillofacial Surgery, 23*(2), 122–128. https://doi.org/10.19438/j.cjoms.2025.02.004

[9] Hao, Y., Huang, M., Li, M., et al. (2023). Preliminary application of artificial intelligence in pathological diagnosis of periapical cysts. *Prevention and Treatment of Oral Diseases, 31*(9), 641–646.

[10] Giraldo-Roldán, D., et al. (2026). Impact of transfer learning on CNNs for odontogenic tumor diagnosis. *Head and Neck Pathology, 20*(1), 24. https://doi.org/10.1007/s12105-025-01875-y

[11] McGenity, C., et al. (2024). Artificial intelligence in digital pathology: A systematic review and meta-analysis of diagnostic test accuracy. *NPJ Digital Medicine, 7*(1), 114. https://doi.org/10.1038/s41746-024-01106-8

DOI: 10.54254/2753-8818/2026.HZ36589

[12] Giraldo-Roldan, D., et al. (2023). Deep learning applied to the histopathological diagnosis of ameloblastomas and ameloblastic carcinomas. *Journal of Oral Pathology & Medicine, 52*(10), 988–995. https://doi.org/10.1111/jop.13481
[13] Sakamoto, K., Morita, K., Ikeda, T., & Kayamori, K. (2019). Deep-learning-based identification of odontogenic keratocysts in H&E-stained jaw cyst specimens.
[14] Yang, H., et al. (2020). Deep learning for automated detection of cyst and tumors of the jaw in panoramic radiographs. *Journal of Clinical Medicine, 9*(6), 1839. https://doi.org/10.3390/jcm9061839
[15] Chai, Z. K., et al. (2022). Improved diagnostic accuracy of ameloblastoma and odontogenic keratocyst on CBCT by artificial intelligence. *Frontiers in Oncology, 11*, 793417. https://doi.org/10.3389/fonc.2021.793417