Supplementary Information

Supplementary Information 1:

Correlation analysis of pathologists' slide-reviewing behaviors and fatigue testing. The various behavioral characteristics were assessed with regard to their correlations with and potential impact on pathologists' diagnostic accuracy, as depicted in Supplementary Figure 1.a. This matrix shows the covariance between variables. Among the numerous factors affecting pathologists' diagnostic accuracy, the one with the greatest impact was the proportion of tissue in the WSI (correlation coefficient 0.38). Among the slide-reviewing behaviors of the pathologists, the one with the strongest correlation with accuracy was the initial fixation scale (-0.25), followed by the number of searches (0.15). Among them, diagnostic accuracy is defined as whether pathologists can make a correct diagnosis without knowing the real diagnosis during the process of reviewing WSI with the cooperation of an eye tracker. The proportion of tissue is defined as the ratio of the foreground size to the WSI. The first fixation scale is defined as the image size observed by the pathologist's first fixation behavior. This is represented by the size within the cluster calculated by DBSCAN [50] using the observation points captured by the eye tracker as the original data. The number of searches is also calculated by DBSCAN and described as the number of times the pathologist's observation moves from one cluster to another. These results suggest that experienced pathologists tend to focus on the specific areas with a smaller scope during the initial fixation process and engage in more searching behavior to gain deeper insights into the WSI. Additionally, there were correlations among most slide- reviewing behavior characteristics; however, most of these indicators had relatively weak correlations with pathologists' diagnostic accuracy.

Based on our correlation analysis of the pathologists' slide-review behavior, we compared the two feature variables, "Initial Fixation Scale" and "Number of Searches", over the continuous slide-reading time during the pathologists' review process. As shown in Supplementary Figure 1.b, both of these indicators exhibited substantial changes around the 50-minute mark. Therefore, the pathologists were considered in a "fatigue state" after working continuously for 50 minutes and were encouraged to take breaks to ensure that the collected data were suitably robust.

Supplementary Information 2:

We further demonstrate that PEAN does not rely on extensive data collection from pathologists. By utilizing a limited amount of gaze-tracking data provided by pathologists, PEAN can effectively learn and capture essential pathology expertise. This capability circumvents the time and resource demands associated with traditional methods that require large-scale data collection. In the initial training phase, we used only 30% of the slide-reviewing data, enabling the PEAN model to successfully decode core pathological knowledge. Subsequently, the model was further trained with the remaining 70% of the data, which included only WSIs and pathological diagnoses. CONCH served as the image encoder. On the internal testing set, PEAN achieved an ACC of 93.1% and an AUC of 0.981; on the external testing set, PEAN achieved an ACC of 86.5% and an AUC of 0.935. Compared to the weakly supervised learning methods in Table 1, PEAN's performance, even with only 30% data, remains competitive. This ability allows the model to continuously integrate new information as medical data grows, progressively improving diagnostic accuracy and generalization capability. This feature not only reduces the cost of subsequent data collection but also enables the PEAN model to rapidly adapt to different clinical settings and datasets, addressing varying clinical needs. Providing PEAN with new data on different diseases in the future will allow the model to continuously accumulate and update its knowledge base, facilitating seamless integration into an evolving medical environment. Particularly in the context of rare disease identification, where relevant datasets are scarce, PEAN—already incorporating pathologists' prior knowledge and extensive pathological image features—holds promise for maintaining long-term competitiveness.

Supplementary Information 3:

To more comprehensively evaluate the generalization ability of PEAN and the feasibility of practical clinical application, we also collected 100 WSIs from each of two other medical institutions (20 for each of the five involved categories), forming a new external testing dataset. Since these represent different patient populations, medical equipment, and operating procedures, it provides relatively high diversity and complexity. CONCH is used as the image encoder in PEAN-C model training. The rest of the experimental settings are consistent with those in Section 2.3. The performance of PEAN-C model and other 8 DL models applying to this new independent testing dataset is shown in Supplementary Tables 3-4. As shown, PEAN-C model achieves the best performance in all cases. These results indicate that PEAN can effectively adopted to different testing environments and avoid overfitting due to a single source dataset.

Supplementary Table 1. Recall of Each Disease with ResNet50

Methods	Internal Testing Set					External Testing Set				
	Nevus	BCC	Melanoma	SCC	SK	Nevus	BCC	Melanoma	SCC	SK
DLCCP	91.8%	54.1%	57.1%	64.9%	96.3%	87.5%	89.4%	58.4%	69.6%	12.9%
SLC	82.0%	78.0%	51.0%	68.9%	97.0%	93.0%	87.9%	36.4%	54.6%	19.2%
HSL	88.2%	87.4%	40.8%	72.9%	96.0%	86.8%	87.9%	54.5%	63.9%	39.9%
CLAM	77.7%	46.5%	75.5%	43.6%	97.3%	83.2%	65.9%	63.6%	39.7%	47.2%
AB-MIL	88.5%	37.7%	44.9%	15.6%	95.7%	89.4%	51.1%	46.8%	13.9%	9.1%
TransMIL	91.1%	76.7%	44.9%	51.1%	95.8%	82.0%	86.7%	66.2%	68.6%	32.5%
DS-MIL	82.6%	80.5%	67.4%	88.0%	91.6%	79.5%	94.0%	88.3%	76.8%	18.5%
IB-MIL	83.0%	79.3%	65.3%	84.9%	93.4%	44.3%	71.1%	63.3%	28.0%	86.0%
PEAN-C	86.2%	74.2%	57.1%	81.8%	97.3%	94.10%	89.7%	40.3%	45.4%	86.7%

Supplementary Table 2. Recall of Each Disease with Conch

Methods	Internal Testing Set					External Testing Set				
	Nevus	BCC	Melanoma	SCC	SK	Nevus	BCC	Melanoma	SCC	SK
DLCCP	96.1%	81.1%	69.4%	66.2%	96.0%	95.0%	93.7%	89.6%	82.5%	28.3%
SLC	91.2%	84.3%	67.4%	78.7%	95.5%	97.6%	93.7%	83.1%	59.3%	59.1%
HSL	91.2%	86.2%	65.3%	84.0%	96.1%	96.4%	93.1%	80.5%	89.2%	47.6%
CLAM	79.0%	78.6%	61.2%	85.8%	92.7%	73.5%	88.5%	87.0%	91.8%	70.3%
AB-MIL	95.4%	84.3%	26.5%	46.7%	93.2%	86.2%	94.3%	88.3%	76.8%	21.3%
TransMIL	96.1%	86.8%	28.6%	68.9%	93.7%	80.3%	84.0%	81.8%	79.4%	61.5%
DS-MIL	82.6%	80.5%	67.4%	88.0%	93.7%	84.1%	85.8%	81.8%	85.1%	57.0%
IB-MIL	94.4%	82.4%	63.3%	70.2%	96.2%	87.5%	87.9%	83.1%	84.0%	63.3%
PEAN-C	96.4%	86.2%	83.7%	88.4%	98.6%	96.8%	95.8%	88.3%	82.5%	83.9%

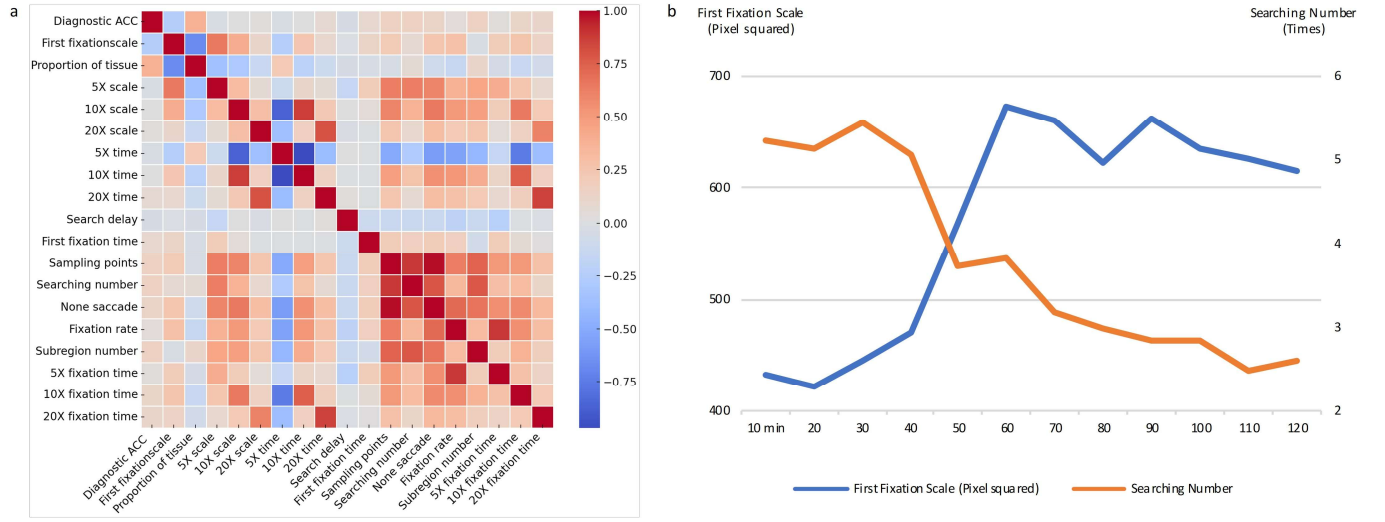
Supplementary Table 3. Comparison of PEAN-C with baselines on Added External Testing 1

Methods	ACC	AUC	Recall of Each Disease				
			Nevus	BCC	Melanoma	SCC	SK
DLCCP	82%	0.944	95%	100%	50%	75%	90%
SLC	85%	0.959	95%	95%	60%	85%	90%
HSL	90%	0.973	95%	95%	95%	70%	95%
CLAM-SB	78%	0.943	95%	95%	90%	15%	95%
ABMIL	75%	0.892	80%	80%	95%	60%	70%
TransMIL	81%	0.949	95%	95%	70%	45%	100%
DS-MIL	81%	0.95	95%	95%	60%	60%	95%
IB-MIL	87%	0.971	95%	95%	65%	80%	100%
PEAN-C	94%	0.987	100%	100%	95%	75%	100%

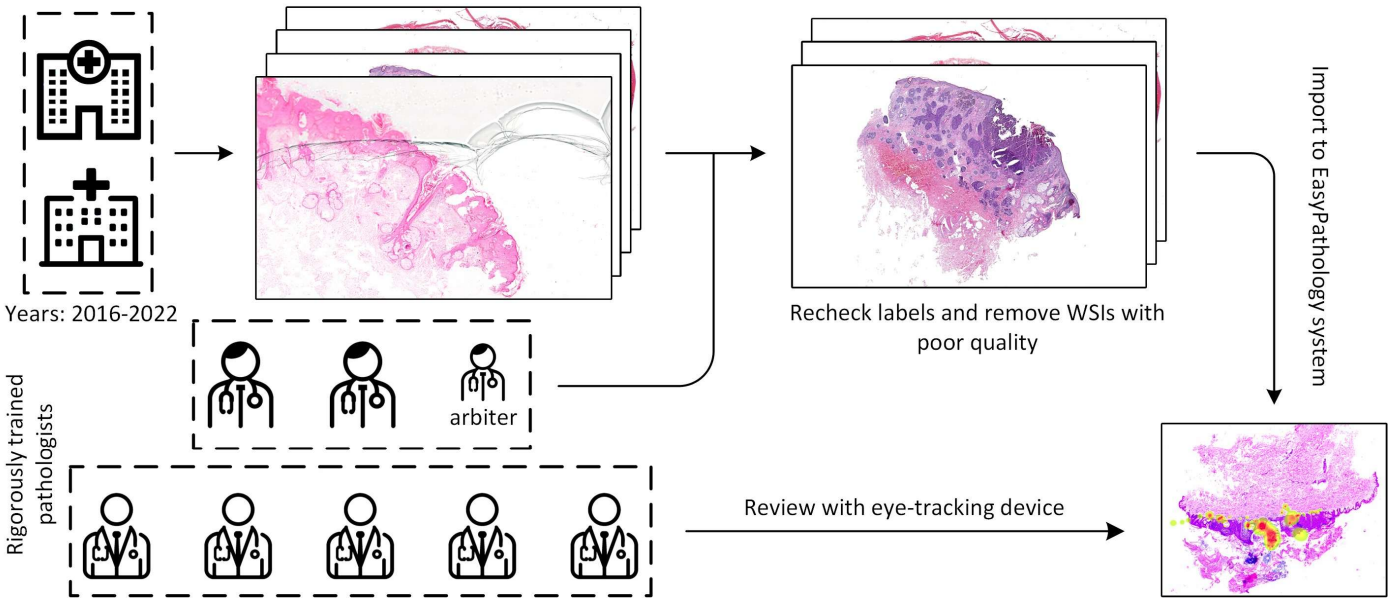
Supplementary Table 4. Comparison of PEAN-C with baselines on Added External Testing 2

Methods	ACC	AUC	Recall of Each Disease				
			Nevus	BCC	Melanoma	SCC	SK
DLCCP	86%	0.953	80%	100%	60%	0.95%	95%
SLC	89%	0.968	75%	95%	75%	100%	100%
HSL	92%	0.969	80%	100%	80%	100%	100%

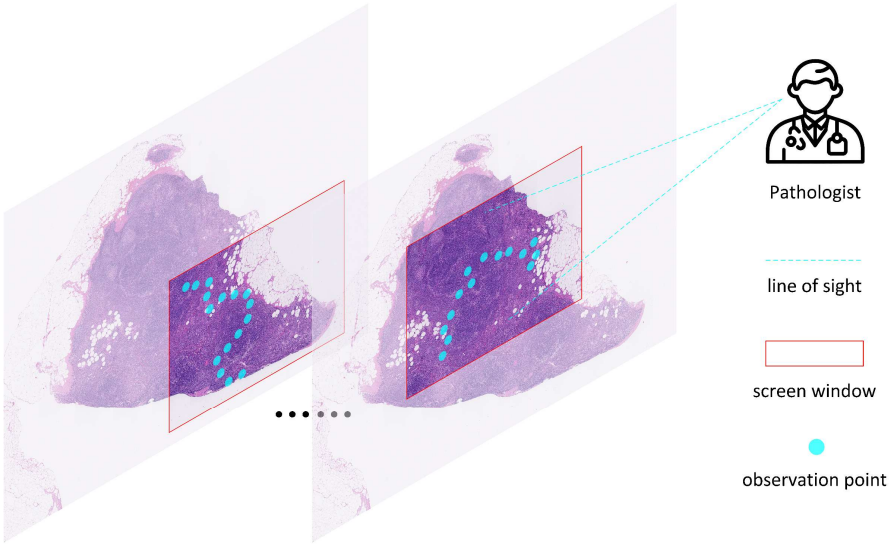
CLAM-SB	82%	0.947	40%	95%	80%	100%	95%
ABMIL	56%	0.872	85%	0%	95%	100%	0%
TransMIL	88%	0.947	75%	90%	80%	95%	100%
DS-MIL	79%	0.918	70%	90%	50%	90%	95%
IB-MIL	87%	0.96	45%	100%	95%	100%	95%
PEAN-C	96%	0.988	85%	100%	95%	100%	100%



Supplementary Figure 1. a. Correlation analysis of pathologists' slide-reviewing behaviors. **b.** The variation curves of Plots of two key pathologist slide-reviewing features of pathologist over continuous working time.



Supplementary Figure 2. The process for dataset collection.



Supplementary Figure 3. Schematic Figure of sliding window operation and observation points