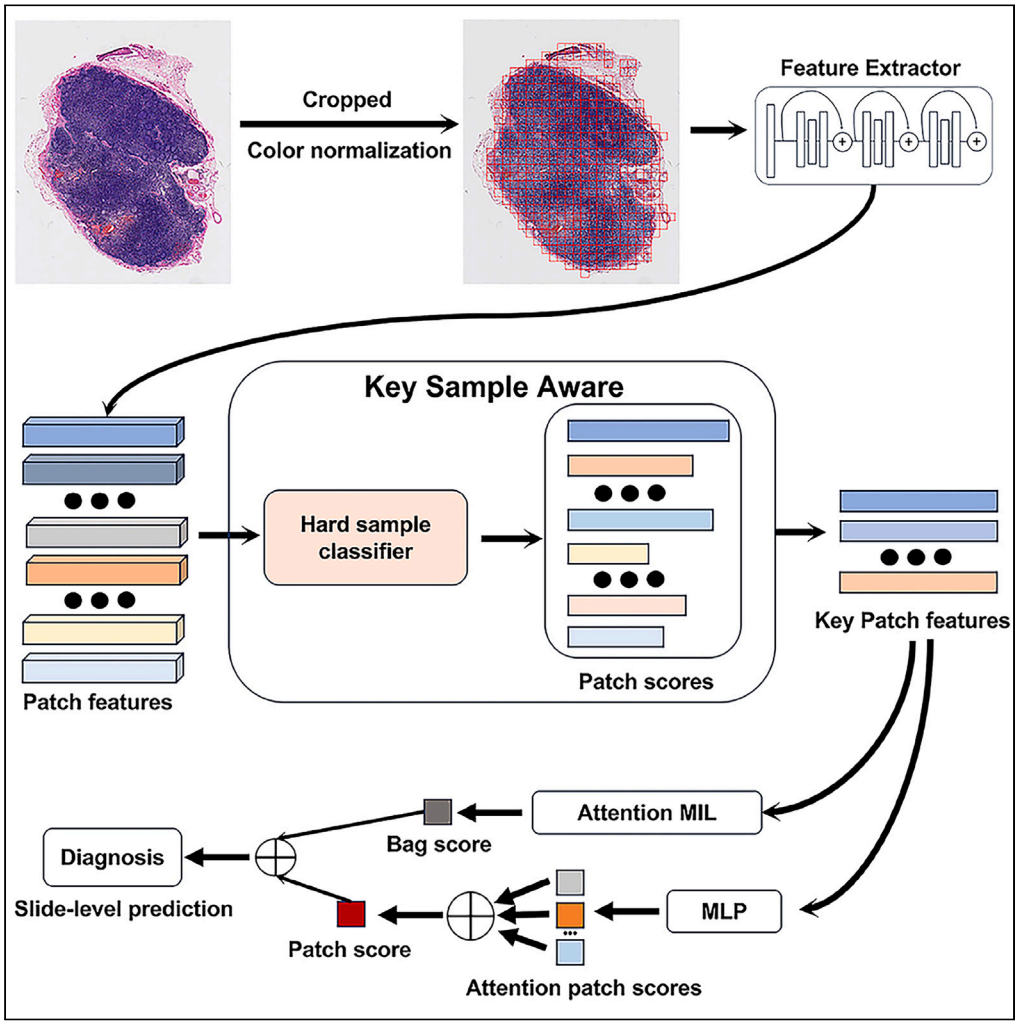
iScience

CellPress
OPEN ACCESS

Article

A deep learning method for predicting the origins of cervical lymph node metastatic cancer on digital pathological images



Runliang Zheng, Xuenian Wang, Lianghui Zhu, ..., Yonghong He, Huijuan Shi, Anjia Han

heyh@sz.tsinghua.edu.cn (Y.H.)
shihj@mail.sysu.edu.cn (H.S.)
hananjia@mail.sysu.edu.cn (A.H.)

Highlights

A large pathological dataset of cervical lymphadenopathy was established

An algorithm predicts cervical lymph node status by targeting key regions

Our method outperforms other methods on both internal and external datasets

Zheng et al., iScience 27, 110645
September 20, 2024 © 2024 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.isci.2024.110645

iScience

CellPress
OPEN ACCESS

Article

A deep learning method for predicting the origins of cervical lymph node metastatic cancer on digital pathological images

Runliang Zheng,[1,5] Xuenian Wang,[1,5] Lianghui Zhu,[1] Renao Yan,[1] Jiawen Li,[1] Yani Wei,[2] Fenfen Zhang,[2] Hong Du,[3] Linlang Guo,[4] Yonghong He,[1,*] Huijuan Shi,[2,*] and Anjia Han[2,6,*]

SUMMARY

The metastatic cancer of cervical lymph nodes presents complex shapes and poses significant challenges for doctors in determining its origin. We established a deep learning framework to predict the status of lymph nodes in patients with cervical lymphadenopathy (CLA) by hematoxylin and eosin (H&E) stained slides. This retrospective study utilized 1,036 cervical lymph node biopsy specimens at the First Affiliated Hospital of Sun Yat-Sen University (FAHSYSU). A multiple-instance learning algorithm designed for key region identification was applied, and cross-validation experiments were conducted in the dataset. Additionally, the model distinguished between primary lymphoma and metastatic cancer with high prediction accuracy. We also validated our model and other models on an external dataset. Our model showed better generalization and achieved the best results on both internal and external datasets. This algorithm offers an approach for evaluating cervical lymph node status before surgery, significantly aiding physicians in preoperative diagnosis and treatment planning.

INTRODUCTION

Cancers still cannot be identified in the clinic and are categorized as cancers of unknown primary (CUP).[1] Diagnosing and treating CUP is notably tricky due to their ambiguous clinical presentations and the absence of reliable biomarkers. A key characteristic of CUP is metastasis without the ability to identify the cancer's primary location through conventional methods, introducing considerable uncertainty in the treatment and prognosis of patients. The diagnosis of CUP typically starts with detecting a metastatic tumor, indicating that the cancer has already spread from its origin to other body parts.[2] Despite significant advancements in imaging and histopathological analysis by modern medicine, standard diagnostic techniques often remain unsuccessful in determining the origin of the cancer. The diagnosis of CUP typically entails a comprehensive approach comprising a detailed patient history, physical examination, blood tests, imageology studies (including computed tomography [CT], MRI, and positron emission tomography [PET scans]), and tissue analysis.[3] Although these diagnostic techniques offer critical information about tumor characteristics, they are frequently time-intensive and expensive.

In response to these diagnostic challenges, computational pathology has emerged as an innovative field blending computer science and pathology, witnessing rapid advancements in recent years.[4] Computational pathology offers a new way to analyze microscopic images of tumor tissue through neural networks to find clues about its site of origin. Deep learning has demonstrated significant potential in handling and interpreting this complex data, offering new ways for diagnosing CUP origin. Through advanced algorithms, computational pathology enables the extraction of significant patterns and features within tumor structures, which might escape direct observation and conventional analysis. Artificial intelligence, empowered by advancements in algorithms, the accumulation of large-scale datasets, and enhanced computational power, exhibits robust capabilities in feature representation learning.[5] Deep learning technology has significantly advanced the medical field, particularly in tumor detection,[6,7] classification,[8] and grading.[6,9] For instance, researchers employed a transfer-based learning approach to construct a comprehensive model for assessing lymph node metastases in colorectal cancer using digitized hematoxylin and eosin (H&E) staining slides.[10] Additionally, a deep learning framework has been developed to determine lymph node status in cervical cancer patients through analysis of H&E staining slides of primary tumors.[11]

Current research mainly aims at detecting the presence or absence of metastatic cancer in lymph nodes.[12–15] However, accurately identifying the primary tumor's location is crucial for guiding the treatment of patients with metastatic tumors, and the research in this area remains

[1]Shenzhen International Graduate School, Tsinghua University, Shenzhen, Guangdong, China
[2]Department of Pathology, the First Affiliated Hospital of Sun Yat-sen University, Shenzhen, Guangdong, China
[3]Department of Pathology, Guangzhou First People's Hospital, South China University of Technology, Guangzhou, China
[4]Department of Pathology, Zhujiang Hospital, Southern Medical University, Guangzhou, China
[5]These authors contributed equally
[6]Lead contact
*Correspondence: heyh@sz.tsinghua.edu.cn (Y.H.), shihj@mail.sysu.edu.cn (H.S.), hananjia@mail.sysu.edu.cn (A.H.)
https://doi.org/10.1016/j.isci.2024.110645

This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

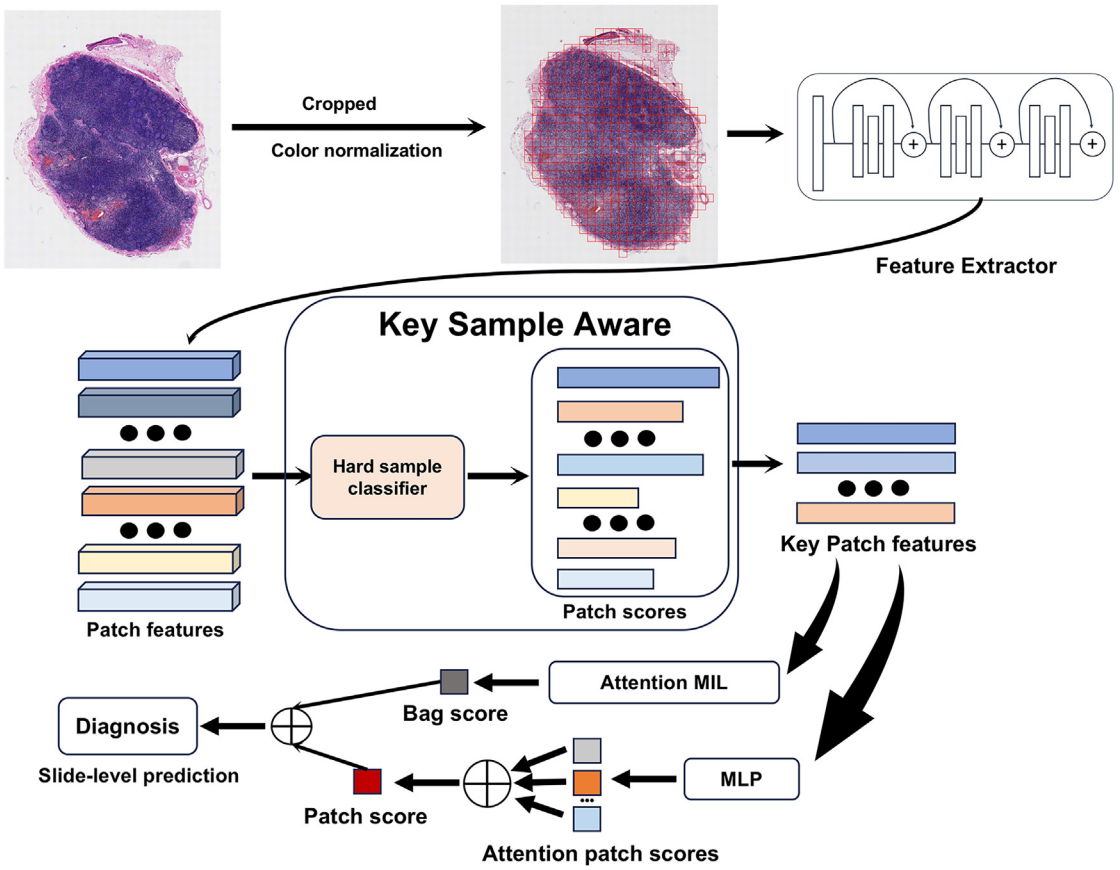


Figure 1. The workflow of KMIL, a multiple-instance learning method for key recognition

lacking.[16] The development of a deep learning algorithm, Tumor Origin Assessment via Deep Learning (TOAD), marks significant progress in this area.[17] TOAD can analyze standard histological sections to predict the primary sites among 18 tumors. This model was trained using entire digital pathological images from known primary tumor sites, enabling it to classify tumors as either primary or metastatic and determine their origins. Similarly, a regional multiple-instance learning algorithm was crafted to determine the origin of bone metastatic cancer based on only H&E stained slides.[18]

Lesions in cervical lymph nodes constitute one-third of all lymph node lesions. Abnormalities in the size or shape of these lymph nodes lead to a diagnosis of cervical lymphadenopathy (CLA).[19] Precisely diagnosing the cause of CLA and identifying the primary tumor site in metastatic CLA cases are essential for determining appropriate patient referral pathways, customizing treatment plans, and conducting tumor staging. Yang et al.[20] employed deep learning techniques to assess the pathological conditions of cervical lymph nodes using PET/CT imaging. Their study specifically focused on differentiating between lymph node metastasis and lymphoma. The results demonstrated robust diagnostic efficacy. However, there are still some gaps in deep learning research on CLA based on pathology.

Multiple instance learning for medical imaging

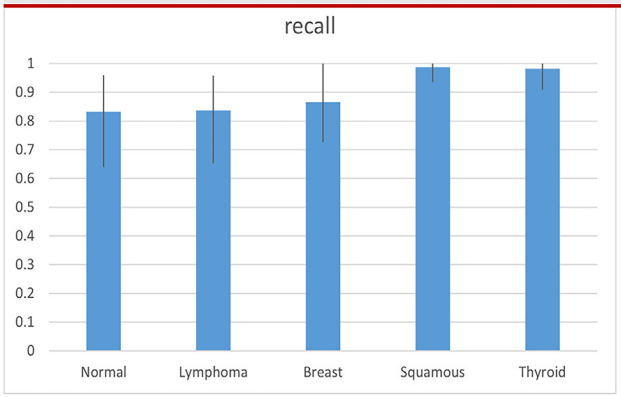
In computational pathology, acquiring pixel-level annotations for medical pathology images is challenging, leading to a need for more usable data. To address this, researchers often divide medical pathology images, which may contain millions of pixels, into smaller patches. These patches, derived from the same whole-slide image (WSI), are collectively treated as a bag labeled with a single identifier. Multiple instance learning (MIL) approaches this by separately processing the extracted image blocks from the WSI and aggregating their features. This aggregation utilizes pooling functions to integrate the features of each image block, allowing subsequent neural network layers to synthesize a comprehensive feature representation of the WSI. This process effectively captures the global information of histological sections. Predominantly, MIL feature aggregation hinges on the MIL pooling operator, consolidating the embedded features derived from WSI patch segmentation into a global representation of the WSI.

Ilse et al.[21] explored the MIL method in pathological studies and proposed a permutation invariant aggregation operator combined with the attention mechanism based MIL (attention based MIL [ABMIL]). The operator can provide information about how each instance affects the judgment of the bag's label. Shao et al.[22] developed a new multi-instance learning framework—transformer-based MIL method (TransMIL).

A

	ABMIL	TransMIL	DTFD	DSMIL	KMIL
ACC	0.7771	0.7801	0.8763	0.8930	**0.9018**
AUC	0.8636	0.9390	0.9695	0.9789	**0.9802**
F1 score	0.6858	0.7574	0.8753	0.8926	**0.9013**

B



C

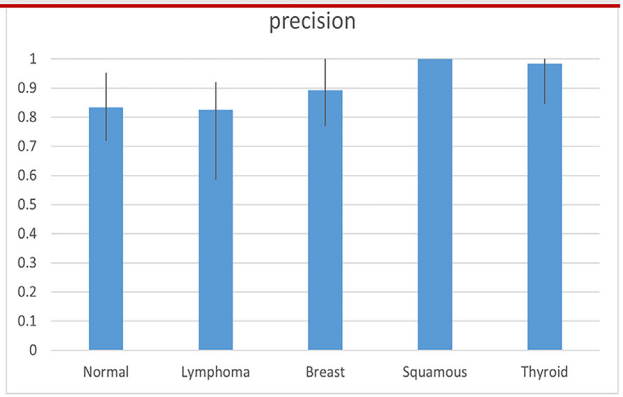


Figure 2. The performance of different models in predicting OCLNMC

(A) AUROC, accuracy, and F1 score of different models; (B) Recall for different categories by KMIL; (C) Precision for different categories by KMIL.

The model mimics the context information and inter-regional correlation pathologists consider when diagnosing. TransMIL uses transformer technology to mine morphological and spatial information in WSI classification and uses location coding and transformer layer to enhance the training efficiency and interpretability of the model. Zhang et al.[23] first proposed the instance probability derivation process under the framework of ABMIL, showing that it is more reliable than the attention score of joint positive region detection. Based on this, they developed a MIL method for double-tier feature distillation (double-tier feature distillation based MIL [DTFD-MIL]). It combines the pseudo-bag concept and double-tier MIL structure to improve the model performance through feature distillation technology. Li's team[24] proposed a MIL method for WSI classification and tumor detection, which solved the problems of WSI's high resolution and lack of localized annotations. Through the innovative introduction of a new MIL aggregator, self-supervised contrast learning, and pyramid fusion mechanism, this method(dual-stream multiple instance learning [DSMIL]) not only optimizes the modeling of the relationship between instances but also effectively alleviates the problem of extensive packet processing and improves the accuracy of classification and location.

Pathological images often have incredibly high dimensions and complex background information, containing many cells, tissue structures, and diseased areas, of which only a few are critical positive samples. Positive samples are critical to the final diagnosis. The existing research methods look for these positive areas from the whole WSI. However, in the actual clinicopathological diagnosis, doctors will only observe some areas and often choose the key areas of interest for observation. In this study, combined with the diagnosis process of pathologists, we proposed a multi-instance learning algorithm to predict the status of cervical lymph nodes by focusing on identifying key regions. Furthermore, this algorithm can predict the origin of metastatic cancer, thereby offering an approach for the personalized treatment of patients with CLA in the future.

RESULTS

Evaluation metrics

The diagnostic performance of our algorithm was assessed by five metrics: accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUROC).

Table 1. The performance of different models

	Recall (%)					Precision (%)				
Methods	Non-tumor	Lymphoma	Breast	Squamous	Thyroid	Non-tumor	Lymphoma	Breast	Squamous	Thyroid
ABMIL	76.30	79.86	58.48	98.74	35.61	71.89	73.94	80.25	84.00	39.09
TransMIL	70.52	72.90	58.64	94.05	79.24	68.02	69.32	75.49	93.99	87.50
DTFD	79.14	77.68	84.85	99.05	97.26	80.09	79.84	84.55	99.41	98.33
DSMIL	80.33	85.34	83.33	98.42	97.27	82.41	80.12	90.94	99.39	99.17
KMIL	81.26	85.44	84.06	99.68	96.36	83.98	81.25	91.17	99.09	99.09

Table 2. The performance of different models on the external dataset

Model	ACC	F1 score	AUROC
ABMIL	0.5494	0.4239	0.7926
TransMIL	0.4871	0.6273	0.8130
DSMIL	0.6918	0.6273	0.9346
DTFD	0.5506	0.4914	0.8442
KMIL	0.7341	0.6766	0.9400

Prediction of the origin of cervical lymph node metastatic

We evaluate five MIL-based deep learning models in the FAHSYSU cohort: ABMIL, TransMIL, DSMIL, DTFD, and KMIL. KMIL is shown in Figure 1. Each model was trained and tested in the FAHSYSU cohort with 10-fold cross-validation experiments. The performance of different models in predicting cervical lymph node (CLN) metastasis status is evaluated by the metrics of the AUROC, accuracy, F1 score in Figure 2A. Further performance breakdowns by our model are provided in Figure 2. Our model achieves an ACC of 90.18%, an AUROC of 98.02%, and an F1 score of 0.9013.

The performance of recall and precision rates of different models in different categories is detailed in Table 1.

Similarly, we compared the generalizations of different models on an external dataset, as shown in Table 2. KMIL demonstrated superior performance across multiple evaluation metrics on an external dataset. Notably, it achieved the highest accuracy with a score of 0.7341. Furthermore, the KMIL model excelled in terms of the F1 score, reaching 0.6766, which suggests a balanced proficiency in precision and recall, critical for models operating under conditions of class imbalance. Additionally, it outperformed other models in the AUROC metric with a score of 0.9400. The generalization of KMIL is better than that of other models, and it is solid and robust. Figure 3 compares the performance of internal and external datasets of KMIL.

As detailed in the Table 3, we evaluated the computational efficiency and model complexity of ABMIL, TransMIL, DSMIL, DTFD, and KMIL. Regarding model size and computational complexity, the proposed model has similar sizes and computational complexities with the models of other works except for TransMIL, while TransMIL is significantly larger.

Ablation study

In the key sample aware scheme, we can calculate the hard score of the sample. The higher the score, the more difficult it is to identify the sample; the lower the score, the more likely it is to be an easy sample. As is shown in Table 4, when the samples with scores less than 0.1 were removed, the model's ACC, F1 score, and AUC all reached the highest. However, as the threshold increased to 0.2, the model's accuracy and F1 score decreased slightly. This suggested that further increases in the threshold might result in more hard samples being incorrectly removed, thereby failing to improve the model's overall performance effectively. KMIL can significantly improve model performance under specific threshold settings. However, it is also essential to select a reasonable threshold to avoid excessive deletion of potentially hard samples and ensure the optimization of model performance.

Table 5 shows the performance changes of the model on the verification set after removing patches whose perceived scores of hard samples are higher than different thresholds. This experiment investigated the specific effects of removing hard samples on model performance. The results show that the performance of the model generally declines when the patches with significant perception scores of hard samples are removed. This suggests that clearing out this hard sample negatively impacts the model's predictive power. Although hard samples are difficult to classify, they contain valuable information for improving model learning and generalization. Removing these samples causes the

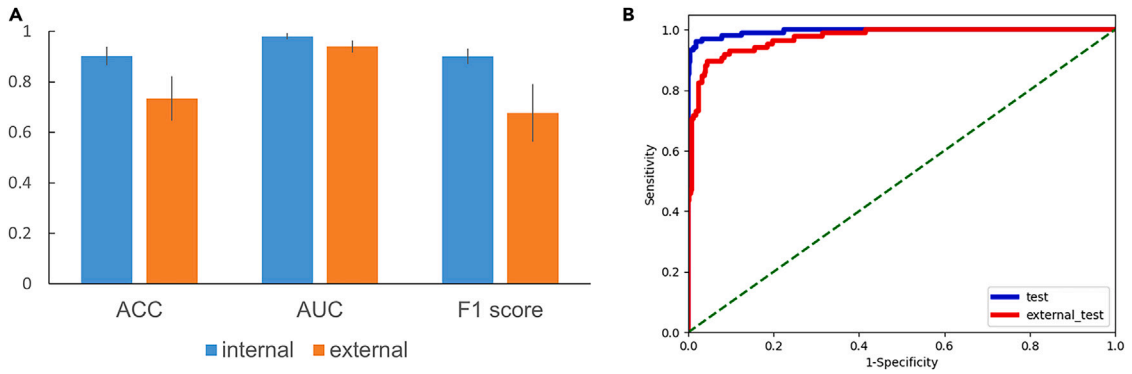


Figure 3. The performance of KMIL on the external dataset

(A) The performance of internal and external datasets of KMIL; (B) Receiver operating characteristic curves of internal and external datasets of KMIL.

Table 3. Comparison of calculation speed and model size of different models

model	Flops	Params	time
ABMIL	61.46M	530.2K	1.064s
TransMIL	602.82M	2673.2K	3.282s
DSMIL	114.38M	641.4K	2.892s
DTFD	83.02M	726.7K	2.301s
KMIL	116.53M	643.6K	2.916s

model to lose necessary training signals, resulting in reduced performance when dealing with complex or boundary cases. The samples with high scores are precisely the key samples that the model needs.

Prediction of cervical lymph node metastasis

We also predicted whether CLA patients had metastatic cancer, called Task A. Our model achieves an ACC of 90.87%, a AUROC of 97.04%, and an F1 score of 0.8834 in Task A.

In clinical diagnosis, distinguishing between non-cancerous conditions, such as inflammation and lymphoma, presents a challenge, particularly for inexperienced pathologists.[25] Our model is also utilized to determine the status of CLN, identifying them as either non-cancerous conditions or lymphoma, which is task B. Our model achieved an accuracy of 83.40% in distinguishing between non-cancerous conditions and lymphoma. Clinically, distinguishing between primary lymphoma and metastatic cancer is crucial. Our model was employed to perform this task, called task C. The KMIL algorithm demonstrated over 90% accuracy in differentiating between the two types of tumors, achieving 93.13% accuracy in identifying lymphoma and 95.38% accuracy in diagnosing metastatic cancer. The performance of our algorithm on these tasks is shown in Table 6.

We verify the generalization of our model for these three tasks on an external dataset, as shown in Figure 4. Figures 4A and 4B illustrate the model's generalization capabilities for task A, while Figures 4C and 4D depict its performance on task B. Figures 4E and 4F demonstrate its generalization for task C. Notably, the model's performance declined on task A and task B, yet it showed improved effectiveness on task C compared to the validation on the internal dataset. Models are more capable of handling simple tasks.

DISCUSSION

Cervical lymph node metastases comprise about 2%–5% of head and neck cancers.[26] Metastases from unknown primary tumors are very rare. They account for roughly 2% of all new cases in this group. However, managing patients with such metastases poses a significant challenge in oncology. Determining the location of primary tumors is crucial for directing the clinical management of patients with metastatic cancer.

A deep learning model utilizing multiple-instance learning was developed to predict OCLNMC on H&E stained slides accurately. The algorithm requires only the WSI level of labeling without more granular or region-specific annotations. It can directly identify metastatic cancer and its origin in routine cases. This simplifies the complex task of determining metastatic cancer sources in clinical settings. As a result, it reduces diagnostic time for patients and saves medical resources.

Due to the sparse distribution of positive regions in pathological images, these areas often constitute a small portion of the WSI. Utilizing traditional MIL methods for feature aggregation from WSI results in a predominance of features from negative patches. This significantly complicates the task of the classifier and impacts its training efficiency. Researchers manually annotate region of interest (ROI) on WSIs and subsequently apply the MIL method for prediction.[11] However, our algorithm is specifically designed for precise, automatic identification and analysis of crucial regions in WSIs. Initially, we segment WSIs into numerous small patches and utilize a pre-trained ResNet50[27] network to extract feature vectors from each patch. Subsequently, a key recognition module evaluates and prioritizes these feature vectors to identify and retain those with critical information. These selected feature vectors, retaining the most vital data in the WSI, are then processed through a MIL framework. In this framework, an attention module assigns scores to each key feature vector, focusing on the most informative features. Then, using the attention pooling mechanism,[28] we condense these key feature vectors to get a comprehensive representation of the whole set. Finally, the score of the bag and the score of each patch are calculated by two MLPs, respectively. The combination of the two scores, which consider the overall information and the local details, forms a comprehensive prediction of the WSI. In short, recognizing and utilizing key features enhances the precision of prediction and facilitates a deeper comprehension of cancer characteristics in WSIs. Selecting hard samples can not only effectively reduce the interference of negative patches on classifier judgment but also significantly improve the learning

Table 4. The perception effect of hard samples with different scores

t	ACC	F1 score	AUC
0	0.8932	0.8928	0.9775
0.1	0.9018	0.9013	0.9802
0.2	0.8930	0.8920	0.9801

Table 5. The relationship between scores and hard samples

t	ACC	F1 score	AUC
0.8	0.8656	0.8577	0.9735
0.85	0.8646	0.8519	0.9755
0.9	0.8685	0.8618	0.9718
1	0.8932	0.8928	0.9775

efficiency and accuracy of the model in pathological image analysis. Based on hard sample aware, this approach optimizes the traditional way of feature aggregation. It ensures the focus on positive samples during model training, thus achieving higher performance and more accurate diagnostic results in pathological image analysis.

The performance metrics demonstrate that our model can precisely predict the status of CLN in CLA patients on H&E stained slides. In most cases of metastatic CLA, the pathological subtypes are predominantly squamous cell carcinoma (SCC) or adenocarcinoma (ADC).[29] For metastatic CLA diagnosed as SCC, primary tumor origins are often identified in areas like the head and neck, lungs, or esophagus. Conversely, for ADC, primary tumors are usually linked to the thyroid, stomach, or lungs. Therefore, accurately identifying the two subtypes in patients with metastatic CLA, particularly those whose primary tumor origins are unclear, is essential for streamlining the diagnostic process and swiftly determining the origins of primary tumors. Our model has now been refined to distinguish between these two pathological subtypes effectively.

In conclusion, our research introduces a highly accurate assistive tool for pathologists to predict origins of cervical lymph node metastatic cancer (OCLNMC) from conventional H&E stained slides. The development of this tool not only simplifies the pathological diagnosis process but also improves the accuracy of prediction and provides strong support for clinical decision-making. By comparing and analyzing different MIL methods, this study demonstrates the significant potential of this technique to improve the accuracy and robustness of model diagnosis. The model outperformed on several key performance indicators, especially in the identification accuracy of specific cancer types. In particular, our model's validation on the external dataset outperforms the internal dataset for identifying primary lymphoma and metastatic cancer. Our key sample aware scheme also provides a new perspective for future pathological research. In pathological research, we eliminate more miscellaneous information, which significantly saves computational cost and efficiency. We can find key features and use them in future studies of vision-language pretraining models for pathology.

Limitations of the study

The current research mainly focuses on the differentiation of various types of tumors and does not subdivide into different subtypes of tumors. For patients with different subtypes, doctors will adopt different treatment methods. Future research can also be carried out on studying tumor subtypes in cervical lymph nodes. Non-Hodgkin lymphoma can be divided into subtypes including B cell lymphoma, NK, and/or T cell lymphoma.

There are specific challenges in metastatic cancer, such as small sample learning and unbalanced data processing. The imbalance of data samples generated by the scarcity of metastatic cancer data far exceeds that of other types of pathological data, and some sample data are often only a few cases for a long time. For example, the number of metastatic cancer, such as lung cancer and gastric cancer in cervical lymph node metastatic cancer are very rare, and the sample number is less than 1% of that of other metastatic cancers.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS

Table 6. The performance of KMIL in different tasks

	ACC of various categories					
Task	non-tumor	lymphoma	metastatic	ACC	F1 score	AUROC
A	0.8242	0.8665	0.9649	0.9087	0.8834	0.9704
B	0.8201	0.8496		0.8340	0.8322	0.8348
C		0.9313	0.9538	0.9471	0.9378	0.9426

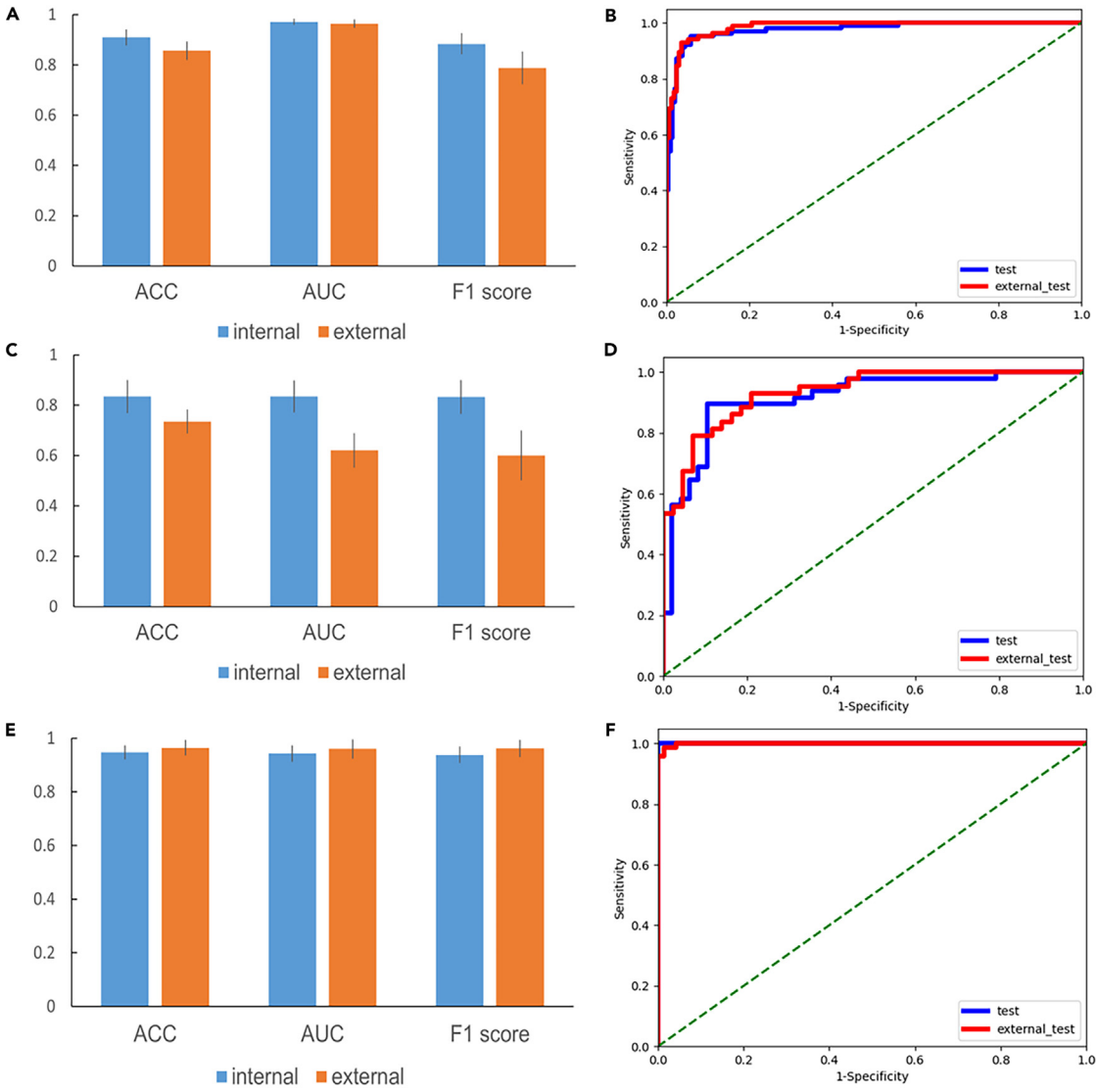


Figure 4. The performance of KMIL for different tasks on the external dataset

(A, C, and E) The performance of internal and external datasets in task (A), (B), and (C); (B, D, and F) Receiver operating characteristic curves of internal and external datasets in task (A), (B), and (C).

- Data acquisition and processing
- Digitization and annotation
- Image preprocessing
- Deep learning method
- Key sample aware scheme

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Implementation details

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110645.

ACKNOWLEDGMENTS

This work was supported by Science and Technology Foundation of Shenzhen City grant number KCXFZ20201221173207022. The work was also supported by the Component Project of Shenzhen Pathology Medical Imaging Intelligent Diagnosis Engineering

Research Center XMHT20230115004 and Jilin Fuyuan Guan Food Group Co., Ltd. Their funding played a crucial role in advancing our research.

AUTHOR CONTRIBUTIONS

A.J.H., R.L.Z., Y.H.H., and H.J.S. conceived the study. R.L.Z., A.J.H., Y.H.H., and H.J.S. designed the experiments. R.L.Z. and X.N.W. completed all codes of data processing, model training and testing. R.L.Z., X.N.W., and L.H.Z. performed the experiments analysis. A.J.H., H.J.S., R.L.Z., J.W.L., Y.N.W., and F.F.Z. curated the FAHSYSU cohort. H.D. and L.L.G. curated the external dataset. R.L.Z. and X.N.W. prepared the manuscript. A.J.H., Y.H.H., and H.J.S. reviewed and edited the manuscript. All authors have read and consented to the publication of the manuscript.

DECLARATION OF INTERESTS

The authors have no conflicts of interest to declare.

Received: March 8, 2024
Revised: June 15, 2024
Accepted: July 30, 2024
Published: August 3, 2024

REFERENCES

1. Rassy, E., and Pavlidis, N. (2020). Progress in refining the clinical management of cancer of unknown primary in the molecular era. Nat. Rev. Clin. Oncol. *17*, 541–554.
2. Massague, J., and Ganesh, K. (2021). Metastasis-initiating cells and ecosystems. Cancer Discov. *11*, 971–994.
3. Alečković, M., McAllister, S.S., and Polyak, K. (2019). Metastasis as a systemic disease: molecular insights and clinical implications. Biochim. Biophys. Acta Rev. Canc *1872*, 89–102.
4. Cui, M., and Zhang, D.Y. (2021). Artificial intelligence and computational pathology. Lab. Invest. *101*, 412–422.
5. Jiang, Y., Yang, M., Wang, S., Li, X., and Sun, Y. (2020). Emerging role of deep learning-based artificial intelligence in tumor pathology. Cancer Commun. *40*, 154–166.
6. Frei, A.L., Oberson, R., Baumann, E., Perren, A., Grobholz, R., Lugli, A., Dawson, H., Abbet, C., Lertxundi, I., Reinhard, S., et al. (2023). Pathologist computer-aided diagnostic scoring of tumor cell fraction: a Swiss national study. Mod. Pathol. *36*, 100335.
7. Dong, H., Yang, G., Liu, F., Mo, Y., and Guo, Y. (2017). Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017 (Springer), pp. 506–517. Proceedings 21.
8. Nojima, S., Kadoi, T., Suzuki, A., Kato, C., Ishida, S., Kido, K., Fujita, K., Okuno, Y., Hirokawa, M., Terayama, K., and Morii, E. (2023). Deep Learning-Based Differential Diagnosis of Follicular Thyroid Tumors Using Histopathological Images. Mod. Pathol. *36*, 100296.
9. Bulten, W., Kartasalo, K., Chen, P.-H.C., Ström, P., Pinckaers, H., Nagpal, K., Cai, Y., Steiner, D.F., Van Boven, H., Vink, R., et al. (2022). Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. Nat. Med. *28*, 154–163.
10. Khan, A., Brouwer, N., Blank, A., Müller, F., Soldini, D., Noske, A., Gaus, E., Brandt, S., Nagtegaal, I., Dawson, H., et al. (2023). Computer-assisted diagnosis of lymph node metastases in colorectal cancers using transfer learning with an ensemble model. Mod. Pathol. *36*, 100118.
11. Guo, Q., Qu, L., Zhu, J., Li, H., Wu, Y., Wang, S., Yu, M., Wu, J., Wen, H., Ju, X., et al. (2023). Predicting Lymph Node Metastasis From Primary Cervical Squamous Cell Carcinoma Based on Deep Learning in Histopathologic Images. Mod. Pathol. *36*, 100316.
12. Challa, B., Tahir, M., Hu, Y., Kellough, D., Lujan, G., Sun, S., Parwani, A.V., and Li, Z. (2023). Artificial Intelligence–Aided Diagnosis of Breast Cancer Lymph Node Metastasis on Histologic Slides in a Digital Workflow. Mod. Pathol. *36*, 100216.
13. Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., Van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J.A.W.M., the CAMELYON16 Consortium, Hermsen, M., Manson, Q.F., et al. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA *318*, 2199–2210.
14. Tang, H., Li, G., Liu, C., Huang, D., Zhang, X., Qiu, Y., and Liu, Y. (2022). Diagnosis of lymph node metastasis in head and neck squamous cell carcinoma using deep learning. Laryngoscope Investig. Otolaryngol. *7*, 161–169.
15. Jin, C., Jiang, Y., Yu, H., Wang, W., Li, B., Chen, C., Yuan, Q., Hu, Y., Xu, Y., Zhou, Z., et al. (2021). Deep learning analysis of the primary tumour and the prediction of lymph node metastases in gastric cancer. Br. J. Surg. *108*, 542–549.
16. Ho, A.S., Kim, S., Tighiouart, M., Gudino, C., Mita, A., Scher, K.S., Laury, A., Prasad, R., Shiao, S.L., Ali, N., et al. (2018). Association of quantitative metastatic lymph node burden with survival in hypopharyngeal and laryngeal cancer. JAMA Oncol. *4*, 985–989.
17. Lu, M.Y., Chen, T.Y., Williamson, D.F.K., Zhao, M., Shady, M., Lipkova, J., and Mahmood, F. (2021). AI-based pathology predicts origins for cancers of unknown primary. Nature *594*, 106–110.
18. Zhu, L., Shi, H., Wei, H., Wang, C., Shi, S., Zhang, F., Yan, R., Liu, Y., He, T., Wang, L., et al. (2023). An accurate prediction of the origin for bone metastatic cancer using deep learning on digital pathological images. EBioMedicine *87*, 104426.
19. Al Kadah, B., Popov, H.H., Schick, B., and Knöbber, D. (2015). Cervical lymphadenopathy: study of 251 patients. Eur. Arch. Oto-Rhino-Laryngol. *272*, 745–752.
20. Yang, Y., Zheng, B., Li, Y., Li, Y., and Ma, X. (2023). Computer-aided diagnostic models to classify lymph node metastasis and lymphoma involvement in enlarged cervical lymph nodes using PET/CT. Med. Phys. *50*, 152–162.
21. Ilse, M., Tomczak, J., and Welling, M. (2018). Attention-based deep multiple instance learning. In International Conference on Machine Learning (PMLR), pp. 2127–2136.
22. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., and Ji, X. (2021). Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Adv. Neural Inf. Process. Syst. *34*, 2136–2147.
23. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., and Zheng, Y. (2022). Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18802–18812.
24. Li, B., Li, Y., and Eliceiri, K.W. (2021). Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14318–14328.
25. Blanc-Durand, P., Jégou, S., Kanoun, S., Berriolo-Riedinger, A., Bodet-Milin, C., Kraeber-Bodéré, F., Carlier, T., Le Gouill, S., Casasnovas, R.-O., Meignan, M., and Itti, E. (2021). Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network. Eur. J. Nucl. Med. Mol. Imag. *48*, 1362–1370.
26. López, F., Rodrigo, J.P., Silver, C.E., Haigentz, M., Jr., Bishop, J.A., Strojan, P., Hartl, D.M.,

Bradley, P.J., Mendenhall, W.M., Suárez, C., et al. (2016). Cervical lymph node metastases from remote primary tumor sites. Head Neck *38*, E2374–E2385.

27. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

28. Er, M.J., Zhang, Y., Wang, N., and Pratama, M. (2016). Attention pooling-based convolutional neural network for sentence modelling. Inf. Sci. *373*, 388–403.

29. Nakanishi, T., Wakai, K., Ishikawa, H., Nawa, A., Suzuki, Y., Nakamura, S., and Kuzuya, K. (2001). A comparison of ovarian metastasis between squamous cell carcinoma and adenocarcinoma of the uterine cervix. Gynecol. Oncol. *82*, 504–509.

30. Zhu, C., Chen, W., Peng, T., Wang, Y., and Jin, M. (2022). Hard sample aware noise robust learning for histopathology image classification. IEEE Trans. Med. Imag. *41*, 881–894.

31. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (Ieee), pp. 248–255.

32. Sudharshan, P., Petitjean, C., Spanhol, F., Oliveira, L.E., Heutte, L., and Honeine, P. (2019). Multiple instance learning for histopathological breast cancer image classification. Expert Syst. Appl. *117*, 103–111.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
ABMIL	Ilse et al.[21]	https://github.com/AMLab-Amsterdam/AttentionDeepMIL
TransMIL	Zhu et al.[18,30]	https://github.com/szc19990412/TransMIL
DSMIL	Li et al.[24]	https://github.com/binli123/dsmil-wsi
DTFD	Zhang et al.[23]	https://github.com/hrzhang1123/DTFD-MIL
KMIL	This paper	https://github.com/wxn22/KMIL
Pytorch	Version 1.13.0	https://pytorch.org/docs/1.13/
Python	Version 3.9.13	https://docs.python.org/release/3.9.13/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Anjia Han (hananjia@mail.sysu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data reported in this paper will be shared by the lead contact upon request.
- All original code has been deposited at GitHub (https://github.com/wxn22/KMIL) and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

We collected cevical lymph node biopsy or surgical specimens which were stained using H&E staining from the First Affiliated Hospital of Sun Yat-sen University and Guangzhou first people's Hospital where key words "neck" and "lymph nodes" in the electronic case systems were searched. These cases in FAHSYSU were diagnosed from 2007 to 2023 and those in the external dataset were diagnosed from 2018 to 2023. We included cervical lymph node metastases cases regardless of histological types and metastatic cancers of known origins, and some biopsies without metastatic carcinomas. We excluded the slides where the tissue was severely squeezed and deformed, affecting the results of diagnoses and the faded ones of which the corresponding paraffin-embedded blocks were missing. Also, we deleted the whole slide images which were not clear. A detail flow of datasets created was available in Figure S1. Study participant details were shown in Table S3 and Table S4. The data were not available and just the image files were used.

This retrospective study received approval from IEC for Clinical Research and Animal Trials at the First Affiliated Hospital of Sun Yat-sen University ([2022]429). All participants provided informed consent before they accepted pathological examinations and operations biopsies.

METHOD DETAILS

Data acquisition and processing

In this study, the FAHSYSU cohort was composed of cervical lymph node biopsy surgical specimens from the First Affiliated Hospital of Sun Yat-sen University in China, and the deep learning method proposed in this study was trained, validated, and tested in the FAHSYSU cohort. We also collected relevant data from Guangzhou First People's Hospital to build an external dataset. The recruitment process for the datasets, including criteria for inclusion and exclusion, is thoroughly documented in the supplemental information (Figure S1). The FAHSYSU cohort that comprises 505 patients is shown in Table S1. The external dataset that comprises 85 WSIs is shown in Table S2. All H&E sections were formalin-fixed and paraffin-embedded tissue sections.

The FAHSYSU cohort included 505 cases diagnosed between January 2007 and March 2023. It was randomly allocated into training (80%), validation (10%), and testing sets (10%) for model development and assessment. In both datasets, the primary sites of cervical lymph node

metastasis originated from common organs, including the breast, thyroid, head, and neck, with the latter primarily consisting of squamous cell carcinomas.

Digitization and annotation

The datasets were scanned by the SQS-1000 scanner (Sqray company, Shenzhen, China) at 40 × magnification. Prior to downloading, all patient-specific private information was removed.

Two pathology experts with two decades of experience were consulted to review the slides to ascertain OCLNMC. They employed a combination of H&E stained slides, immunohistochemical findings, and patient clinical data for diagnostic purposes. In cases of disagreement between the two pathologists, a third expert is called upon to review the slides. If necessary, additional immunohistochemical markers are utilized to verify the metastasis source.

Image preprocessing

For this study, the whole slide images were segmented into 256×256 pixel patches without overlap at 20 × magnification. Before initiating the training process, data augmentation methods such as random flipping and color jittering were applied.

Additionally, the RGB pixel values of each patch were normalized to a range between − 1 and 1, utilizing the mean and standard deviation values from ImageNet.[31]

Deep learning method

A complete digital whole slide image typically encompasses trillions of pixels. The WSI can be segmented into tens or hundreds of thousands of 256×256 patches for training. However, labeling each patch individually would be found to be prohibitively time-consuming. Considering the vast amount of data and the limited number of pathologists available for detailed annotation, such an approach is impractical.[32] Therefore, this study requires labeling only at the WSI level.

We propose the Key Recognition based Multiple-instance learning (KMIL) algorithm, designed for the critical task of data key recognition. This algorithm facilitates the assessment of the status of cervical lymph nodes and determines the origins of metastatic cancer. As shown in Figure 1, for a certain WSI, all patches obtained from segmentation are compiled into a single bag. The ResNet50 network,[27] pre-trained on ImageNet, serves as the primary feature extractor, extracting the features of each patch to aggregate a comprehensive feature set for the bag. Utilizing a specially designed Key Recognition Module, the algorithm identifies significant feature vectors, assigns attention scores to each patch, and selects key feature vectors for further analysis. These key vectors which originally came from the same WSI form a new bag, which is then processed through a MIL architecture. In this framework, an attention module calculates the attention score for each patch, which is then followed by integrating the overall characteristics of the bag through attention pooling. Subsequently, the scores for the bag and individual patches are derived from two separate Multilayer Perceptrons (MLPs). Finally, the combined scores are aggregated to obtain the prediction of the WSI.

Key sample aware scheme

Considering the sparsity of positive samples in pathological images, in most pathological images, the positive regions that can indicate lesions are only a small proportion, which poses a unique challenge to the process of feature aggregation. When dealing with WSIs, most of the patches directly obtained from WSI segmentation tend to reflect the features of negative slides, increasing the classification difficulty of the classifier and slowing down the classifier's training speed. Many negative patches represent easy samples, while the real challenge, the hard samples, are often represented as those patches containing positive labels. In response to the need for pathological image analysis, we pay special attention to emphasizing the importance of positive samples. For a WSI, even if it is divided into thousands to tens of thousands of patches, if only one patch shows positive, then the whole WSI is judged positive. In this process, increasing the Witness Rate, which is the proportion of positive instances in the positive "bag," is the key to significantly improving the performance of the model. Zhu et al.[30] pointed out that the closer the forecast probability of the model output is to the supervision label, the smaller the loss of the sample. However, the normalized probability is more easily analyzed than the loss value. The average prediction probability of an easy sample is usually higher than that of a difficult sample.

As shown in Figure S2, we calculate the attention score sequence of each patch through patch features and then obtain the attention score through the pooling operation. Then, we calculate the deviation degree from the sample center value to the attention score. The difference between this deviation degree and one is defined as the problematic sample score. The larger the score, the more difficult it is to identify the sample. Finally, hard sample classifier selects samples with higher hard sample scores to form new key sample features.

$$X_{att} = \text{attention}(X) \quad \text{(Equation 1)}$$

$$X_{pooling} = \text{pooling}(X_{att}) \quad \text{(Equation 2)}$$

$$hard\_sample\_score = 1 - g(X_{pooling}, central\_value) \quad \text{(Equation 3)}$$

Here, X represents the patch features with dimensions $N \times 1024$, where N is the number of patches obtained from segmenting the WSI. *attention* denotes the attention mechanism, *pooling* represents the pooling operation, *central_value* denotes the central value of the sample, and g represents the function used to calculate the degree of deviation. In this paper, the variance function is used.

QUANTIFICATION AND STATISTICAL ANALYSIS

A ten-fold cross-validation was employed to mitigate the effects of random dataset partitioning. The model was trained using 8-fold data, and 1-fold data served as a validation set to identify the most effective model. Ultimately, the remaining fold data was employed as a test set for evaluating the performance of the model, and we also verified the generalization of the model on the external dataset.

Implementation details

The model was built on the Pytorch framework and trained using an NVIDIA RTX 3090 GPU with 24GB of video memory. Each patch is set to a feature dimension of 1024. The cross-entropy loss function is used in the training process of the model, and the Adam optimizer is selected to optimize the model's performance. As for the adjustment of the learning rate, the CosineAnnealingLR strategy is adopted, the initial learning rate is set as 2e-4, and the weight attenuation coefficient of 1e-5 is set. The batch size of the model training was set to 1, and a total of 400 rounds were performed in the whole training process.