

# Discriminating Chromophobe RCC from Oncocytoma: A Transformer-based Approach Leveraging the Subtleties of Nuclear Structures within Kidney Tumors

Jing Yang<sup>a</sup>, Xingyu Li<sup>b</sup>, Hongjiu Ren<sup>c</sup>, Yanmei Zhu<sup>d,e,f</sup>, Qimin Wang<sup>g</sup>, Ruiqun Qi<sup>h,i</sup>, Xiaoyu Cui<sup>b,\*</sup> and Huamao Jiang<sup>j,\*</sup>

<sup>a</sup>Department of Pathology, First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China

<sup>b</sup>College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China

<sup>c</sup>Department of Pathology, First Affiliated Hospital of China Medical University, Shenyang, China

<sup>d</sup>Department of Pathology, Cancer Hospital of Dalian University of Technology, Dalian, China

<sup>e</sup>Department of Pathology, Liaoning Cancer Hospital & Institute, Shenyang, China

<sup>f</sup>Department of Pathology, Cancer Hospital of China Medical University, Shenyang, China

<sup>g</sup>Department of Pathology, Second Affiliated Hospital of Dalian Medical University, Dalian, China

<sup>h</sup>Department of Dermatology, The First Hospital of China Medical University, Shenyang, China

<sup>i</sup>Key Laboratory of Immunodermatology, Ministry of Education, and National Health Commission, National Joint Engineering Research Center for Theranostics of Immunological Skin Diseases, Shenyang, China

<sup>j</sup>Department of Urology, First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China

## ARTICLE INFO

### Keywords:

Multi-Instance Learning; Chromophobe RCC; Oncocytoma; Deep Learning

## ABSTRACT

**Background:** Renal cell carcinoma (RCC), predominantly seen in adults, includes three common histologic sub-types: clear cell RCC, papillary RCC, and chromophobe RCC (ChRCC), with the latter accounting for 5%–7% of RCC cases. The differentiation of ChRCC from oncocytoma, a benign renal neoplasm, poses a diagnostic challenge because of their histological resemblance, leading to potential overtreatment. To address this diagnostic challenge, the integration of artificial intelligence (AI) algorithms into the diagnostic process represents a promising solution. Leveraging the capabilities of AI, particularly advanced machine learning techniques, can significantly improve the precision and efficiency of differentiating between ChRCC and oncocytoma.

**Methods:** This study focused on developing a Transformer-based multi-instance learning model for the efficient and accurate differentiation of ChRCC and oncocytoma. Data from four medical institutions, comprising 153 whole slide imaging samples, were used. A significant aspect of our research lies in the emphasis on the subtleties of nuclear structures within kidney tumors, a key differentiator between ChRCC and oncocytoma.


**Results:** By integrating these crucial nuclear features into our model, we achieved substantial improvements in diagnostic accuracy. The model demonstrated an impressive average accuracy of 92.50%, area under the curve of 96.42%, precision of 90.32%, and recall of 100%. To further evaluate our model's efficacy, we conducted comparative analyses with diagnostic capabilities of medical professionals with varying levels of expertise. Remarkably, our model's performance ranked second, aligning its proficiency with that of an associate chief physician.

**Conclusions:** The differentiation between chromophobe renal cell carcinoma (ChRCC) and oncocytomas presents a diagnostic challenge due to their histological similarities. The integration of artificial intelligence (AI) algorithms into the diagnostic process has significantly improved the accuracy of distinguishing between these two entities. The high sensitivity and specificity of the AI model underscore the potential of AI in enhancing precision medicine and reducing the risk of overtreatment.

## Introduction

Renal cell carcinoma (RCC) is the predominant form of kidney in adults, accounting for 90%–95% of cases. RCC includes three common histologic subtypes: clear cell renal cell carcinoma (ccRCC), papillary renal cell carcinoma (pRCC) and chromophobe renal cell carcinoma (ChRCC). ChRCC is the third most common form of RCC, constituting 5% to 7% of all RCCs, and originates from the collecting duct epithelium[1–8]. In contrast, oncocytoma, the most common type of benign renal neoplasm, arises from the intercalated cells of renal distal tubules. These two distinct tumor types are increasingly drawing attention from urologists, radiologists and pathologists. ChRCC is characterized by large pale cells with prominent cell membranes and eosinophilic cells; the tumor cells usually have irregular wrinkled nuclei and perinuclear halos. However, some cases lack these characteristic features or show these features in a limited area[9–11]. Moreover,

\*Corresponding author

 cuixy@bmie.neu.edu.cn (X. Cui); Jianghm@jzmu.edu.cn (H. Jiang)

ORCID(s):

difficulties in distinguishing oncocytoma from ChRCC can contribute to instances of overtreatment in some cases. Therefore, a precise diagnosis is crucial to mitigate the risk of overtreatment.

For pathologists, distinguishing ChRCC from oncocytoma only from morphological features from hematoxylin and eosin (H&E) stained histology slides is a diagnostic challenge in daily practice[12]. While the integration of immunohistochemistry and molecular tests has proven highly beneficial for diagnosis, these methods can be both time consuming and costly. A new diagnostic method that can distinguish between ChRCC and oncocytoma from histological images will facilitate efficient and accurate diagnosis.

Several studies have explored the difference between ChRCC and oncocytoma. Jiyeon An et al[13] found that detection of loss of HNF-1 $\beta$  expression by immunostaining was helpful in the diagnosis of ChRCC, especially in cases with similar histological features or equivocal CK7 staining. Khaled et al[14] constructed ChRCC and oncocytoma related gene signature(COGS) to differentiate between ChRCC and oncocytoma using multiple datasets and machine learning. Luis et al[15] observed upregulation of proteins related to the oxidative phosphorylation pathway in oncocytoma compared with ChRCC through quantitative proteomic analysis. Differentiation between these two tumors was also evaluated through deep learning radiomics analysis. Amir et al[16] designed automated PEER measurement and Abeer et al[17] developed ML-base radiomics signatures to distinguish ChRCC and oncocytoma. However, the discrimination between ChRCC and oncocytoma using deep learning techniques applied solely to pathological slides, which exhibits both strong performance and generalizability, has not been extensively explored.

In recent years, notable strides have been taken in the use of deep learning techniques[18–20]. In view of the problem of sample scarcity in pathology image classification, the researchers proposed an innovative method called Dual-Channel Prototype Network(DCPN), which can achieve efficient and accurate pathology image classification with limited data amount[21]. In order to further improve the adaptability and generalization ability of the models on different datasets, the researchers developed a self-supervised learning framework called the Global Contrast-masked Autoencoder(GCMAE). GCMAE aims to capture both local and global features of pathological images through self-supervised learning, thus significantly improving the transfer learning performance of models across datasets[22].at the same time,the Transformer architecture[23] has made remarkable progress in the field of medical image analysis. Transformer architecture was originally designed for natural language processing, demonstrating remarkable ability in handling complex visual data in recent studies. Transformer-based models have become increasingly prevalent in the field of medical image analysis[24]. These models can automatically process histopathology images and learn differentiated features of tumors. In this study, we used a Transformer-based model for differentiation between ChRCC and oncocytoma by analyzing pathological slides.

## Methods

### Clinical Specimens and Data Acquisition

This study included 153 whole slide imaging (WSI) cases from 153 patients who underwent surgery from 2016 to 2023 at the pathology departments of four medical institutions (Liaoning Cancer Hospital, The First Affiliated Hospital of China Medical University, The Second Affiliated Hospital of Dalian Medical University, and The First Affiliated Hospital of Jinzhou Medical University)(Fig.1).To ensure accurate pathological diagnosis, four senior pathologist from these four hospitals reevaluated 153 cases of renal tumor according to WHO 5th diagnosis criteria for renal tumor.[25] Of the 153 WSI cases, 103 cases were diagnosed with ChRCC and 50 cases were diagnosed with oncocytoma(Table 1). WSI was generated using scanning equipment from Hamamatsu Photonics, Aperio GT450, and Motic EasyScan. The inclusion criteria were as follows: (a) Patients were diagnosed with ChRCC or oncocytoma after postoperative pathological examination; (b) All the cases were diagnosed based on both histopathology and immunohistochemical examination by paraffin sections, the result of which strongly supported the diagnosis.(c) WSI was available and sufficiently clear for use for further analysis. This study was approved by the Ethics Committee of The First Affiliated Hospital of Jinzhou Medical University (ethics number: KYLL 2023135) and informed consent was obtained from all patients.

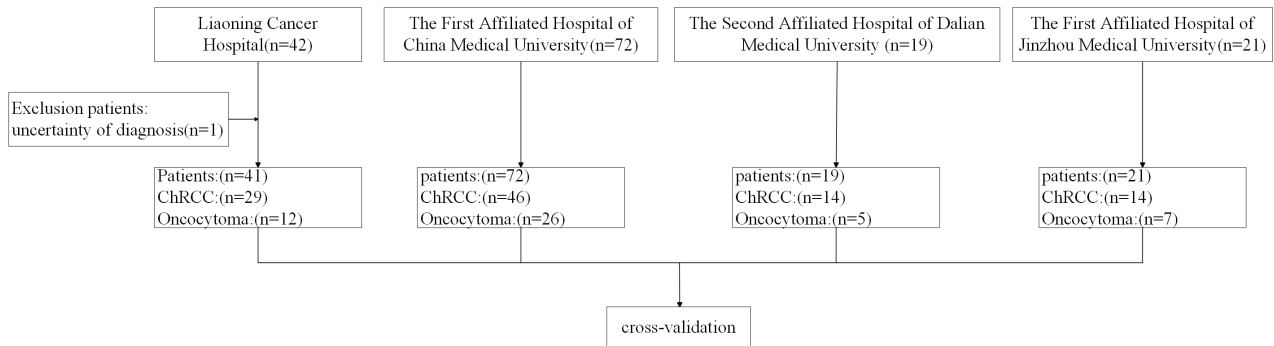
We conducted a comprehensive and detailed statistical analysis of the dataset, with a particular focus on the distribution of image widths and heights at the highest magnification. This extensive examination revealed a large volume of data, underscoring the vastness of our dataset. To ensure compatibility with the prerequisites of model training, we further segmented the original images into smaller patches, each measuring 224x224 pixels. The segmentation facilitated a more manageable and focused analysis, enabling our model to learn from a diverse array of features within the dataset. The distribution is illustrated in Fig.2.

### Deep Learning Method

We used a Transformer-based multi-instance learning architecture, namely TransMIL[26], to address the weakly supervised classification challenge in WSI for pathological diagnosis. The choice of a weakly supervised approach was motivated by the nature of WSIs, where obtaining detailed, pixel-level annotations for training deep learning models is both time-consuming and requires extensive expertise in pathology. Given that these high-resolution images often contain billions

**Table 1**  
Collection of ChRCC and oncocytoma data

| Source of slice (Hospital) | Count of ChRCC | Count of Oncocytoma |
|----------------------------|----------------|---------------------|
| Hospital 1                 | 29             | 12                  |
| Hospital 2                 | 46             | 26                  |
| Hospital 3                 | 14             | 5                   |
| Hospital 4                 | 14             | 7                   |



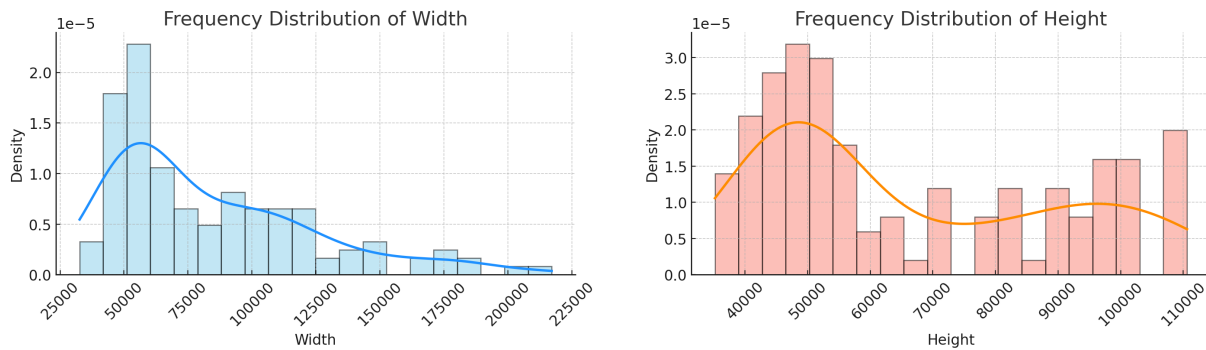
**Figure 1:** Flowchart of the study population

of pixels, it is impractical to expect detailed annotations across the entire slide. Instead, weakly supervised learning leverages coarser or less detailed labels, such as the overall diagnosis for the entire slide, to train models that can still achieve remarkable accuracy at the task level. By integrating the TransMIL architecture within this framework, we capitalize on the transformer’s ability to handle complex, long-range dependencies across different regions of the slides. This approach allows the model to effectively learn from large-scale, high-dimensional WSI data using only minimal, slide-level annotations.

To begin the preprocessing phase of our study, we applied a systematic approach to the gigapixel images obtained through WSI. Our method involved automatically segmenting the tissue regions within each image, ensuring the delineation of non-overlapping patches measuring  $224 \times 224$  pixels Fig 3a. This process was meticulously designed to exclude background regions demonstrating a saturation level below 15%[27], thereby focusing our analysis on the most relevant and information-rich areas of the tissue samples. Given the vast size of our dataset, we subsequently employed a pre-trained convolutional neural network (CNN), such as ResNet50[28], for feature extraction. This step transforms all tissue patches into low-dimensional feature embeddings, effectively reducing the dimensionality of the feature space and thereby decreasing the computational requirements for the subsequent deep learning model. Moreover, this process extracts highly concise and relevant features necessary for accurate diagnosis from the images. A Transformer-based Patch Transformer (TPT) module was designed 3b. This module comprises two Transformer encoder layers and a location coding layer, utilizing Positional Encoding with Grid (PPEG) for location coding. PPEG allows for the encoding of various granularities of location information. Therefore, TransMIL not only handles large volumes of WSI data, but also provides deeper insights to help improve the accuracy of pathology diagnosis.

To further advance our methodological precision, we incorporated an additional, critical step involving the extraction of nuclear features from each tissue patch utilizing HoverNet[29] Fig 3c. This specialized neural architecture is expressly designed for the meticulous segmentation and classification of cell nuclei within histopathological images, an essential process given the paramount importance of nuclear morphology in accurate pathological diagnosis and classification. Following the segmentation with HoverNet, the extracted nuclear patches undergo a secondary feature extraction process via ResNet50. The utilization of HoverNet for the precise segmentation of nuclear features, followed by the application of ResNet50 to these nuclear patches, ensures a comprehensive extraction of morphological details. This sophisticated, two-tiered feature extraction process leverages the forefront of advancements in deep learning and computer vision technologies, markedly improving the diagnostic accuracy of our analysis. By meticulously extracting and analyzing nuanced nuclear features, this approach enables the model to more effectively differentiate between ChRCC and oncocytoma.

In the training phase, we employed the cross-entropy loss[30], utilizing the Lookahead[31] optimizer with a learning rate of  $2e-4$  and a weight decay of  $1e-5$ . A small batch size of 1 was used in the training process. The features of each patch were captured by the ResNet50 model, which was pre-trained on ImageNet[32], resulting in each patch being embedded into a 1024-dimensional vector. Moreover, nuclear features were fused at a ratio of 10:1. Throughout the training, the dimension of each feature embedding was reduced from 1024 to 512 through a fully connected layer. Consequently, the feature embedding of each patch can be denoted as  $H_i \in R^n \times 512$ . All experiments were conducted using RTX A6000.



**Figure 2:** Histograms of slide height and width at maximum magnification provide a visual representation of the distribution of the dimensions (height and width) of slides when viewed at their highest magnification level.

## Evaluation Metrics

We evaluated the performance of the model using the following evaluation metrics: accuracy (ACC), area under the receiver operating characteristic curve (AUC), precision (P), and recall(R)[33–35].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

Where TP is the number of samples in the positive category accurately predicted by the model; TN is the number of samples in the negative category correctly identified by the model; FP is the number of false positives (when the model incorrectly predicts the negative category as positive); and FN is the number of false negatives (when the model incorrectly predicts the positive category as negative).

The AUC specifically refers to the area under the ROC curve. This curve plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. The AUC value, which ranges from 0 to 1, serves as a measure of the model's ability to discriminate between positive and negative classes. A value of 0.5 suggests no discriminative power, equivalent to random guessing, while a value of 1 indicates perfect discrimination, where the model correctly classifies all positive and negative examples. An AUC closer to 1 reflects a model with higher accuracy and better performance in distinguishing between the two categories.

## Pathologists Diagnostic Test

To rigorously assess our model's diagnostic capabilities for ChRCC and oncocytoma, it was essential to benchmark its performance against the traditional gold standard—expert evaluations by pathologists. We engaged 13 pathologists from the First Affiliated Hospital of Jinzhou Medical University and Lingyuan Central Hospital, representing a broad spectrum of experience levels. This approach was intended to encompass a comprehensive array of diagnostic perspectives and methodologies that are intrinsic to current clinical settings. This comparative analysis was pivotal for substantiating the model's precision and reliability, as well as for pinpointing potential areas where the model excels or may need enhancements compared with conventional diagnostic techniques.

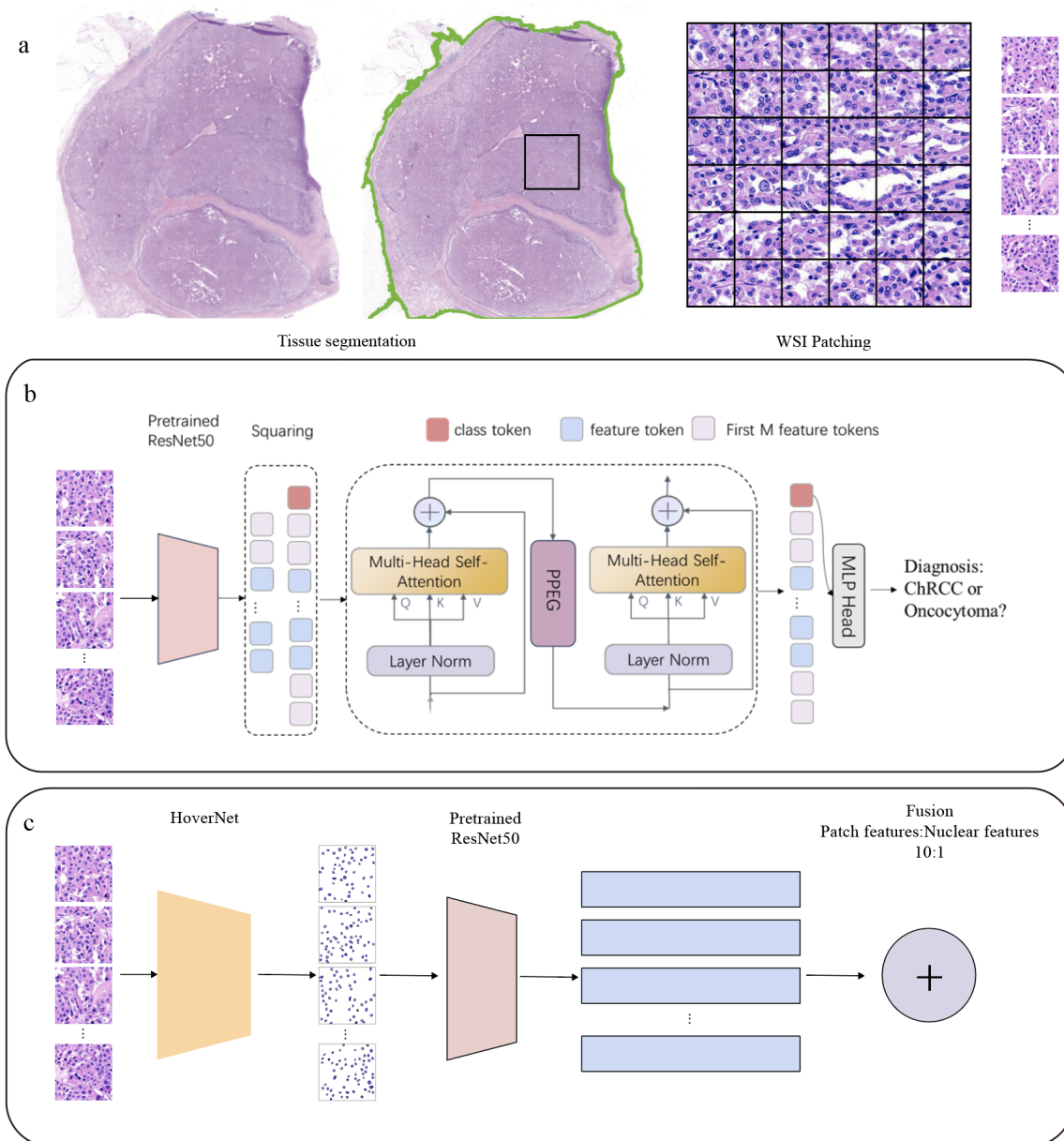
## Statistics

We performed a comprehensive and detailed statistical analysis of four different hospital case classifications and datasets, with a special focus on the distribution of image widths and heights at the highest magnification, to compare the performance of different machine learning models.

## Results

### Different Magnification Experiment

Variations in magnification reveal different levels of pathological detail. To investigate the critical role of magnification levels in the accurate diagnosis of pathological conditions, we performed a series of comprehensive experiments. Through meticulous analysis, we sought to understand how features, observed across different levels of detail from low to high magnifications, contribute to the model's ability to accurately distinguish between pathological conditions. We averaged



**Figure 3: Overview of the TransMIL framework and architecture.** **a** Following segmentation, image patches were extracted from the tissue regions of the WSI. **b** Patches were embedded in feature vectors by a pre-trained ResNet50. The sequence was then processed with the TPT module: 1) squaring of sequence; 2) correlation modeling of the sequence; 3) conditional position encoding and local information fusion; 4) deep feature aggregation; and 5) diagnosis of ChRCC or oncocytoma. **c** The feature extraction is carried out using a pre-trained ResNet50 model after performing nuclear detection using HoverNet.

cross-center outcomes at various magnifications, summarized in 2, to elucidate the classification accuracy of our model using images of varying magnifications. The incorporation of high-magnification image features significantly enhanced the model's performance, thereby highlighting their indispensable role in refining diagnostic precision.

### Cross-center Experiment

The variability in medical imaging data from differences in equipment, staining protocols, patient populations, and even subtle preferences in image acquisition techniques across medical institutions can have a marked impact on the performance of deep learning models. To rigorously assess the model's generalizability and performance across different datasets, we adopted a novel approach: training the model on the dataset from one center and testing it on the datasets from the remaining three centers. This procedure was systematically rotated, ensuring that each center's dataset served as the training set once, with the others used for testing, thereby allowing us to evaluate the model's robustness and adaptability across diverse data sources. The final results are the average outcomes of these four experimental runs, offering a comprehensive overview of

**Table 2**

Performance of deep learning methods in differentiating chromophobe renal cell carcinoma and oncocytoma at various magnifications

|     | Method      | ACC          | AUC          | Precision    | Recall       |
|-----|-------------|--------------|--------------|--------------|--------------|
| 10X | Meanpooling | 58.20        | 61.34        | 79.29        | 53.63        |
|     | Maxpooling  | 56.30        | 60.79        | 85.34        | 48.98        |
|     | Transmil    | 70.89        | 66.59        | 79.85        | 76.57        |
| 20X | Transmil    | <b>85.61</b> | <b>90.10</b> | <b>84.59</b> | <b>96.48</b> |

**Table 3**

Performance of deep learning methods in differentiating chromophobe renal cell carcinoma and oncocytoma at various hospitals

|            | ACC   | AUC   | Precision | Recall       |
|------------|-------|-------|-----------|--------------|
| Hospital 1 | 84.82 | 93.49 | 82.75     | <b>97.29</b> |
| Hospital 2 | 87.65 | 89.40 | 87.30     | <b>96.49</b> |
| Hospital 3 | 79.85 | 85.46 | 78.18     | <b>96.62</b> |
| Hospital 4 | 90.15 | 92.05 | 90.12     | <b>95.50</b> |

**Table 4**

Difference in performance between the model at 20x magnification and the model at 20x magnification after fusing nuclear features.

| TransMIL            | ACC          | AUC          | Precision    | Recall     |
|---------------------|--------------|--------------|--------------|------------|
| w/o Nuclear feature | 85.61        | 90.10        | 84.59        | 96.48      |
| w/ Nuclear feature  | 92.50(+6.89) | 96.42(+6.32) | 90.32(+5.73) | 100(+3.52) |

the model's performance in a cross-center context. Our findings, detailed in Table 3, demonstrate the deep learning models' proficiency in maintaining high recall across datasets from different centers. This achievement is particularly important as it highlights the models' ability to consistently identify true positive cases across a variety of settings, which is crucial in medical diagnostics where the cost of missing a positive case can be extremely high.

### Fusion of Nuclear Features

In the challenging task of diagnosing kidney tumors, particularly when differentiating between ChRCC and oncocytoma, the inherent histological similarities between lesions present a challenging obstacle. Traditional diagnostic methods hinge on the detailed observation of nuclear structures within the tissues, where only subtle differences may indicate one disease over the other. Recognizing the limitations of current techniques, our study focused on harnessing these critical nuclear features through deep learning, aiming to enhance diagnostic accuracy.

By incorporating key nuclear features into our deep learning model, we observed a substantial improvement in diagnostic performance. The precision of our approach is reflected in the model's average ACC, AUC, precision, and recall metrics, the details of which are presented in Table 4.

### AI vs. Pathologists: Diagnostic Accuracy

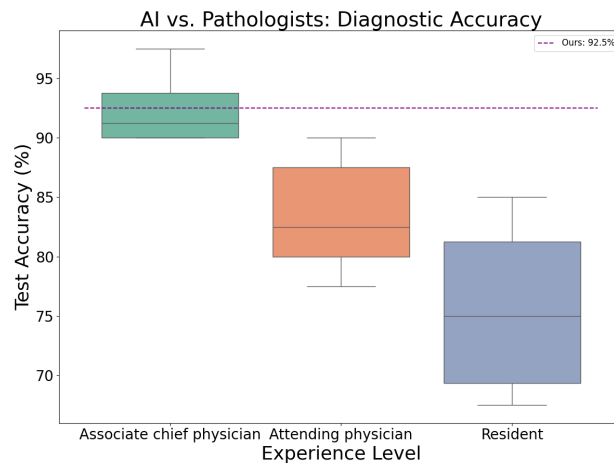
Table 5 presents a comprehensive view of the test accuracies across various levels of medical experience, ranging from residents to associate chief physicians. This allowed us to draw significant insights into the differences in diagnostic accuracy among doctors at different stages of their careers. Associate chief physicians exhibited the highest level of diagnostic accuracy, with the top score reaching 97.5%. Their accuracies were consistently high, with multiple entries at or above 90%. This indicates that doctors with the most experience and presumably the highest level of expertise demonstrate superior diagnostic capabilities. Their high performance underscores the value of extensive clinical experience and specialized knowledge in achieving diagnostic excellence. Attending physicians showed a broader range of accuracies, from 77.5% to 90%. This variability suggests that while attending physicians possess considerable experience and skill, there is more variance in their diagnostic accuracy compared with the uniformly high performance of associate chief physicians. This group likely represents a mix of specializations and individual competencies, which might contribute to the wider spread of accuracies. Residents, who are at the beginning of their medical careers, showed the lowest range of accuracies, from 67.5% to 85%. This is expected, as residents are still in the process of gaining the experience and knowledge necessary to reach higher levels of diagnostic proficiency. The range of accuracies within this group reflects the learning curve and growth potential at this

**Table 5**

The results of physicians with different levels of expertise in the classification of pathologic diagnosis of ChRCC and tumor cell tumors

| Test Accuracy | ExperienceLevel           | Doctor Name |
|---------------|---------------------------|-------------|
| 97.50         | Associate chief physician | Dr.1        |
| <b>92.50</b>  | -                         | <b>Ours</b> |
| 92.50         | Associate chief physician | Dr.2        |
| 90.00         | Associate chief physician | Dr.3        |
| 90.00         | Associate chief physician | Dr.4        |
| 90.00         | Attending physician       | Dr.5        |
| 87.50         | Attending physician       | Dr.6        |
| 85.00         | Resident                  | Dr.7        |
| 82.50         | Attending physician       | Dr.8        |
| 80.00         | Attending physician       | Dr.9        |
| 80.00         | Resident                  | Dr.10       |
| 77.50         | Attending physician       | Dr.11       |
| 70.00         | Resident                  | Dr.12       |
| 67.50         | Resident                  | Dr.13       |

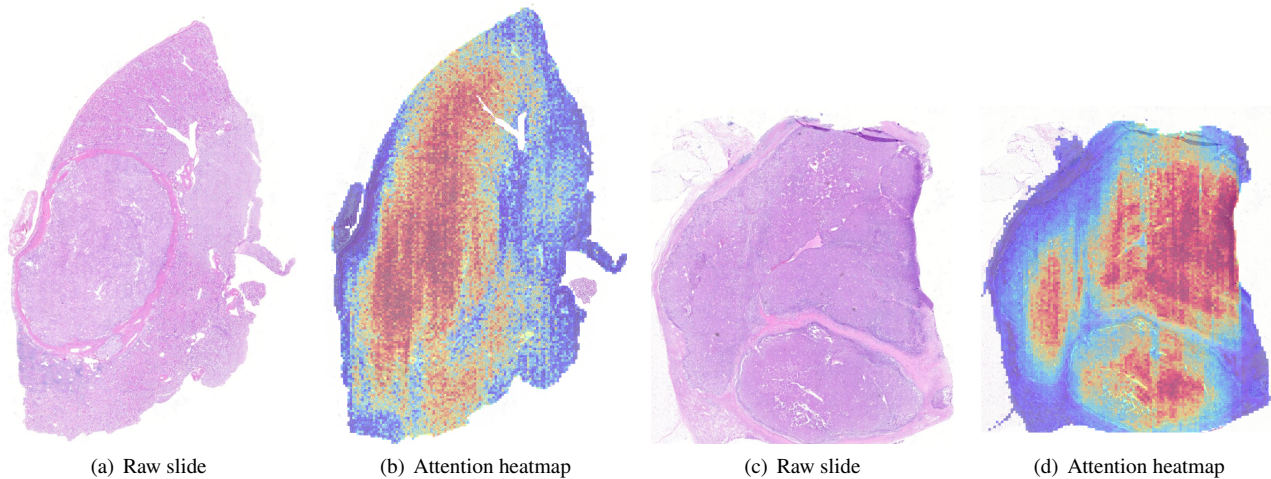
stage of a medical career. We aimed to demonstrate the model’s practical utility and its potential to complement or assist in the diagnostic process. The evaluation, as outlined in Fig. 4, placed our model’s test accuracy in the second rank, indicating a level of diagnostic proficiency that aligns with associate chief physicians. This positioning not only validates the model’s effectiveness in distinguishing between ChRCC and oncocytoma but also underscores its potential as a supportive tool for pathologists.



**Figure 4:** Comparison between artificial intelligence (AI) and doctors with various experience levels in terms of test accuracy. The differently colored boxes represent the distribution of test accuracies across experience levels. This visual representation allows us to see the variance and spread in test accuracies among doctors with different levels of experience. The purple dashed line, annotated as “Ours: 92.5%,” denotes the performance of our model in test accuracy.

### Interpretability and Attention Visualization

By visually representing the attention of a model, particularly in the context of medical imaging, clinicians gain insight into the features that the model deems important for its predictions. This enhances trust in the model’s decisions and aids clinicians in understanding how the model arrives at its conclusions, ultimately improving the integration of machine learning models into clinical practice. Figure 5(a) and 5(c) shows slides collected from The First Hospital of Jinzhou Medical University. As shown in Figure 5(b) and 5(d), attention scores from TransMIL were visualized as a heatmap to determine the ROI and interpret the important morphology used for diagnosis. There was a high consistency between the cancerous regions on the slides and heatmaps, illustrating great interpretability and attention visualization of the TransMIL.



**Figure 5:** Visualization of attention scores as a heatmap generated by TransMIL on medical slides obtained from The First Hospital of Jinzhou Medical University. The heatmap highlights regions of interest indicative of significant morphological features relevant for diagnosis, showcasing the interpretability and attention visualization capabilities of TransMIL.

## Discussion

Our research introduces a comprehensive approach towards enhancing the diagnostic accuracy for ChRCC and oncocytoma through the deployment of a transformer-based multi-instance learning framework, TransMIL. By conducting experiments across varying magnifications, we uncovered the significant influence of high-magnification image features on the model’s classification capability. Our cross-center experiments further validated the model’s generalizability and robustness, demonstrating its ability to maintain high diagnostic accuracy across diverse datasets.

A key innovation in our study was the integration of nuclear features into the deep learning model, which marked a significant advancement over traditional diagnostic methods. This integration not only mitigated the challenge posed by the histological similarity between ChRCC and oncocytoma to a certain extent, but also significantly enhanced the model’s performance metrics, as demonstrated by the improved accuracy (92.50%), AUC (96.42%), precision (90.32%), and recall scores (100%).

The comparison between our AI model and pathologists revealed the model’s commendable diagnostic proficiency, positioning it as a valuable tool to assist pathologists, especially in challenging cases. This comparison not only highlights the model’s potential in enhancing diagnostic accuracy but also its role in mitigating the risks associated with diagnostic ambiguities.

Moreover, our study emphasized the importance of interpretability and attention visualization in medical diagnostics. By providing visual insights into the model’s decision-making process, we facilitate a deeper understanding and trust in AI-driven diagnostics among clinicians. This approach underscores the potential of AI to complement traditional diagnostic methodologies, paving the way for more accurate and efficient diagnosis and treatment strategies in renal pathology.

Despite the notable advancements achieved by our model, it was not infallible, with accuracy not reaching 100%. The primary reason for the misclassification of the methods studied in this paper lies in the limitations of the model in understanding atypical pathological slides. Pathological images contain a wealth of details and subtle differences, such as cellular morphology and abnormal changes in tissue structure, and this information often needs to be interpreted by pathologists with deep expertise and extensive experience. While the current model can learn certain patterns and features with the support of a large amount of training data, it may not be able to accurately capture key information when facing rare diseases, atypical lesions, or complex background interference, which may lead to misjudgment[36].

Reflecting on these insights, it becomes evident that the journey towards perfecting AI-driven diagnostic tools in pathology is ongoing. The challenges posed by atypical pathological slides—rich in detail yet subtle in their differences—underscore the necessity for continued innovation and integration of deep learning technologies with clinical expertise. Our research highlights the significant strides made in enhancing diagnostic accuracy through AI and also illuminates the path forward. As we aim to further refine these models, the collaborative effort between technologists and medical professionals will be paramount. Together, we can advance towards a future where AI not only complements but also elevates the precision and reliability of medical diagnostics, bridging the gap between current limitations and the potential for truly transformative healthcare solutions.



## Conclusions

In this study, the Transformer-based Multi-Instance Learning Model (TransMIL) was used to successfully achieve an effective differentiation between ChRCC and oncocytoma by digging deeper into the subtle features of the cellular nuclear structure in renal tumours. Compared with pathologists with different medical experiences, the diagnostic ability of the model is close to the level of associate chief physician, demonstrating its great potential as an auxiliary diagnostic tool.

## Declarations

### Ethics approval and consent to participate

This study was approved by the Ethics Committee of The First Affiliated Hospital of Jinzhou Medical University (ethics number: KYLL 2023135) on August 9, 2023, and informed consent was obtained from all patients.

### Consent for publication

All the authors have read the manuscript and agree to publish.

### Availability of data and materials

The raw data of this article will be available from the corresponding author upon reasonable request.

### Competing interests

All the authors declare that they have no competing interests.

## Funding

This research received a grant from horizontal project of Jinzhou Medical University (STC-2S21094) and Liaoning Provincial Medical and Industrial Cross Joint Fund (2022-YGJC-06)

## Authors contributions

literature search:XL,YZ.figures:XL.study design:JY,XL,XC, HJ. data collection:JY,HR,YZ,QW,HJ.data analysis:RQ,XC, HJ. data interpretation:XL,RQ,XC.All authors contributed to the revision and approved the final manuscript.

## Acknowledgements

We would like to thank the four medical institutions that participated in this study: the Liaoning Provincial Cancer Hospital, the First Affiliated Hospital of China Medical University, the Second Affiliated Hospital of Dalian Medical University, and the First Affiliated Hospital of Jinzhou Medical University, which provided us with rich clinical samples and data. Meanwhile, we express our deepest gratitude to the 153 patients who participated in the study. In addition, this study was supported by grants from Jinzhou Medical University (STC-2S21094) and Liaoning Provincial Medical and Industrial Cross Joint Fund(2022-YGJC-06).

## References

- [1] Walter B, Hartmann A, Hofstädter F, Junker K, Moch H, Bertz S, et al. Immunohistochemical marker panel differentiates between the three most common subtypes of renal cell carcinoma independent from histomorphologic criteria. *Virchows Archiv*. 2012;460:343-352.
- [2] Motzer RJ, Bander NH, Nanus DM. Renal-cell carcinoma. *New England Journal of Medicine*. 1996;335(12):865-875.
- [3] Srigley JR, Delahunt B, Eble JN, Egevad L, Epstein JI, Grignon D, et al. The International Society of Urological Pathology (ISUP) vancouver classification of renal neoplasia. *The American journal of surgical pathology*. 2013;37(10):1469-1489.
- [4] Chandrasekaran D, Sundaram S, Kadhiresan N, Padmavathi R, dsc , sdjh , et al. Programmed death ligand 1; An immunotarget for renal cell carcinoma. *Asian Pacific Journal of Cancer Prevention: APJCP*. 2019;20(10):2951.
- [5] Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, et al. The cancer genome atlas comprehensive molecular characterization of renal cell carcinoma. *Cell reports*. 2018;23(1):313-326.
- [6] VM GO, LM GT, Hoz N. Immunohistochemical profile of renal cell tumours. *Revista espanola de patologia: publicacion oficial de la Sociedad Espanola de Anatomia Patologica y de la Sociedad Espanola de Citologia*. 2019;52(4):214-221.
- [7] Truong LD, Shen SS. Immunohistochemical diagnosis of renal neoplasms. *Archives of pathology & laboratory medicine*. 2011;135(1):92-109.
- [8] Kuroda N, Tanaka A, Ohe C, Nagashima Y. Recent advances of immunohistochemistry for diagnosis of renal tumors. *Pathology International*. 2013;63(8):381-390.
- [9] Ng KL, Rajandram R, Morais C, Yap NY, Samaratunga H, Gobe GC, et al. Differentiation of oncocytoma from chromophobe renal cell carcinoma (RCC): can novel molecular biomarkers help solve an old problem?. *Journal of Clinical Pathology*. 2013.
- [10] Zhang L, Henske EP. Chromophobe renal cell carcinoma: New genetic and metabolic insights. in *Urologic Oncology: Seminars and Original Investigations*;38:678-681Elsevier 2020.
- [11] Mazal PR, Exner M, Haitel A, Krieger S, Thomson RB, Aronson PS, et al. Expression of kidney-specific cadherin distinguishes chromophobe renal cell carcinoma from renal oncocytoma. *Human pathology*. 2005;36(1):22-28.
- [12] Al-Aynati M, Chen V, Salama S, Shuhaibar H, Treleaven D, Vincic L, et al. Interobserver and intraobserver variability using the Fuhrman grading system for renal cell carcinoma. *Archives of pathology & laboratory medicine*. 2003;127(5):593-596.

- [13] An J, Park CK, Kim M, Joo JW, Cho NH. HNF-1 $\beta$  as an immunohistochemical marker for distinguishing chromophobe renal cell carcinoma and hybrid oncocytic tumors from renal oncocytoma. *Virchows Archiv*. 2021;478:459–470.
- [14] Satter KB, Tran PMH, Tran LKH, Ramsey Z, Pinkerton K, Bai S, et al. Oncocytoma-related gene signature to differentiate chromophobe renal cancer and oncocytoma using machine learning. *Cells*. 2022;11(2):287.
- [15] Carvalho LB, Jorge S, López-Fernández H, Lodeiro C, Dhir R, Campos Pinheiro L, et al. Proteomic analysis of chromophobe renal cell carcinoma and benign renal oncocytoma biopsies reveals shared metabolic dysregulation. *Clinical Proteomics*. 2023;20(1):54.
- [16] Baghdadi A, Aldhaam NA, Elsayed AS, Hussein AA, Cavuoto LA, Kauffman E, et al. Automated differentiation of benign renal oncocytoma and chromophobe renal cell carcinoma on computed tomography using deep learning. *BJU international*. 2020;125(4):553–560.
- [17] Alhussaini AJ, Steele JD, Nabi G. Comparative Analysis for the Distinction of Chromophobe Renal Cell Carcinoma from Renal Oncocytoma in Computed Tomography Imaging Using Machine Learning Radiomics Analysis. *Cancers*. 2022;14(15):3609.
- [18] Tabibu S, Vinod P, Jawahar C. Pan-Renal Cell Carcinoma classification and survival prediction from histopathology images using deep learning. *Scientific reports*. 2019;9(1):10509.
- [19] Tomita , Naofumi , Cheung , Yvonne Y, Hassanpour , Saeed , et al. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans.
- [20] Yu-Jen YJC, Kai-Lung H, Che-Hao H, Wen-Huang C, Chusnul HS, dsc , et al. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *Oncotargets Therapy*. 2015;8:2015-2022.
- [21] Quan H, Li X, Hu D, Nan T, Cui X. Dual-Channel Prototype Network for Few-Shot Pathology Image Classification. *IEEE Journal of Biomedical and Health Informatics*. 2024;28(7):4132-4144.
- [22] Quan H, Li X, Chen W, Bai Q, Zou M, Yang R, et al. Global contrast-masked autoencoders are powerful pathological representation learners. *Pattern Recognition*. 2024;156:110745.
- [23] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. *arXiv*. 2017.
- [24] Guo D, Terzopoulos D. A Transformer-Based Network for Anisotropic 3D Medical Image Segmentation. in *2020 25th International Conference on Pattern Recognition (ICPR) 2021*.
- [25] WHO. Classification of tumours of the urinary system and male genital organs.ed.5. Lyon, France: International Agency for Research on Cancer; 2022.
- [26] Shao Z, Bian H, Chen Y, Wang Y, Zhang J, Ji X, et al. TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification. 2021.
- [27] Duran-Lopez L, Dominguez-Morales JP, Conde-Martin AF, Vicente-Diaz S, Linares-Barranco A, dsc , et al. PROMETEO: A CNN-Based Computer-Aided Diagnosis System for WSI Prostate Cancer Detection. *IEEE Access*. 2020;PP(99):1-1.
- [28] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. in *IEEE Conference on Computer Vision and Pattern Recognition 2016*.
- [29] Graham S, Vu QD, Raza SEA, Azam A, Tsang YW, Kwak JT, et al. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis*. 2019;58:101563.
- [30] Lin TY, Goyal P, Girshick R, He K, Dollár P, dsc , et al. Focal loss for dense object detection. in *Proceedings of the IEEE international conference on computer vision*:2980–2988 2017.
- [31] Zhang MR, Lucas J, Hinton G, Ba J. Lookahead Optimizer: k steps forward, 1 step back. 2019.
- [32] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*. 2015;115(3):211-252.
- [33] Powers DMW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. 2020.
- [34] Hanley JA, Mcneil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve.. *Radiology*. 1982;143(1):29.
- [35] Soleymani R, Granger E, Fumera G. F-measure curves: A tool to visualize classifier performance under imbalance. *Pattern Recognition*. 2020;100:107146.
- [36] Morales S, Engan K, Naranjo V. Artificial intelligence in computational pathology - challenges and future directions. *Digital Signal Processing*. 2021.