npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-024-01301-7

Clinically applicable optimized periprosthetic joint infection diagnosis via AI based pathology

Ye Tao[1,5], Yazhi Luo[2,5], Hanwen Hu[1,5], Wei Wang[3], Ying Zhao[3], Shuhao Wang[3], Qingyuan Zheng[1], Tianwei Zhang[1], Guoqiang Zhang[1], Jie Li[4] ✉ & Ming Ni[1] ✉

Periprosthetic joint infection (PJI) is a severe complication after joint replacement surgery that demands precise diagnosis for effective treatment. We enhanced PJI diagnostic accuracy through three steps: (1) developing a self-supervised PJI model with DINO v2 to create a large dataset; (2) comparing multiple intelligent models to identify the best one; and (3) using the optimal model for visual analysis to refine diagnostic practices. The self-supervised model generated 27,724 training samples and achieved a perfect AUC of 1, indicating flawless case differentiation. EfficientNet v2-S outperformed CAMEL2 at the image level, while CAMEL2 was superior at the patient level. By using the weakly supervised PJI model to adjust diagnostic criteria, we reduced the required high-power field diagnoses per slide from five to three. These findings demonstrate AI's potential to improve the accuracy and standardization of PJI pathology and have significant implications for infectious disease diagnostics.

Periprosthetic joint infection (PJI) is a serious complication post joint replacement, with an incidence rate between 1 and 2%[1–5]. This complication necessitates complex, multi-stage joint revision or joint fusion, leading to poor prognosis and posing substantial burdens on individuals and society at large[6]. Accurate diagnosis is pivotal for effective treatment, and early diagnosis can help alleviate symptoms and improve prognosis. The 2018 International Consensus Meeting (ICM) guidelines provide a standardized approach by combining clinical, serological, joint fluid, and imaging assessments[7].

In the diagnosis of other infectious diseases, pathology is often considered the gold standard for confirming the pathogen and the characteristic lesions. However, with the pathological diagnosis of PJI, doctors have difficulty directly observing pathogens under a microscope, and specific tissue and cellular changes, such as vascular proliferation, tissue necrosis, and neutrophil aggregation are definitive evidence of localized infection[8–10]. This complicated identification process tends to entail intensive analysis by experienced pathology experts and is less cost-effective, especially when medical resources are limited[11–13].

In the standardization of pathological identification of infectious tissues, research efforts have been directed toward the aggregation of neutrophils. According to the 2018 ICM guidelines and related research, at least one pathological slide from a patient should cover five high-power fields (at 400× magnification), with each field containing five or more neutrophils[7,14]. However, in practice, this method depends on the pathologist's expertise in identifying neutrophils. Less experienced doctors may misdiagnose or underdiagnose PJI, affecting the clinical utility of pathological diagnosis[15].

The introduction of artificial intelligence (AI) into medical image processing renders it feasible to standardize pathological diagnosis and enhance its accuracy[16]. However, developing an AI model for pathological diagnosis to identify PJI-infected tissues substantially differs from well-established models used for cancer diagnosis[17]. The intelligent cancer pathology diagnosis model achieves expert-level accuracy for diagnosing superficial tissues like skin, cervical, and gastric cancers by precisely delineating boundaries[18–21]. Nonetheless, PJI pathology lacks well-defined boundaries for infection-positive indicators, making it difficult to employ existing model training strategies directly.

Current infection diagnostic models require special staining and manual annotation of various pathogens. Moreover, this approach calls for labor-intensive annotation during model training[22] and is applicable to pathogens whose forms are directly observable under high magnifications, such as malaria parasites or fungi[19,23]. For less clear pathogens, AI may produce false positives and overlook features like necrosis and inflammation, reducing sensitivity. Simultaneously, research on intelligent PJI diagnosis, such as Kuo et al.'s meta-classifier model, uses multiple heterogeneous

[1]Orthopedics Department, Fourth Medical Center, Chinese PLA General Hospital, Beijing, China. [2]Department of computation, information and technology, Technical University of Munich, Munich, Germany. [3]Thorough Lab, Thorough Future, Beijing, China. [4]Department of Pathology, First Medical Center, Chinese PLA General Hospital, Beijing, China. [5]These authors contributed equally: Ye Tao, Yazhi Luo, Hanwen Hu. ✉e-mail: lijiek812@foxmail.com; niming301@163.com

inputs and improves diagnostic accuracy from an AUC of 0.958–0.988 by applying IF-THEN rules and decision trees[24]. Studies like Yeo et al.'s use AI to predict infection likelihood post-surgery[25]. However, these models often resemble black-box approaches, complicating result interpretation. Therefore, focusing on intelligent single-index PJI diagnostics holds greater value as it can more intuitively enhance the accuracy of existing diagnoses.

This study aimed to enhance the pathological diagnostic accuracy of PJI and improve its clinical utility by standardizing the intelligent diagnostic process and providing guidelines for diagnosing infectious diseases. We compared three learning frameworks, i.e., PJI supervised learning models, weakly supervised learning models, and self-supervised learning models, using pathological images of periprosthetic joint tissues from patients with confirmed PJI and non-PJI cases. By evaluating different models and architectures, we developed a broadly applicable PJI diagnostic model that optimizes existing standards and improves diagnostic precision through AI quantification analysis.

Results

Patient population

This study collected data from 150 patients admitted between December 2017 and May 2023 at the Chinese PLA General Hospital, Beijing, China. Among these patients, 94 were confirmed to have had PJI through bacterial culture or genetic sequencing, while 56 patients were diagnosed with aseptic revision. Baseline data, including age, sex, body mass index (BMI), American Society of Anesthesiologists (ASA) classification for anesthesia (Table 1), and types of infecting bacteria (including gram-positive bacteria, gram-negative bacteria, fungi, mycobacteria, and polymicrobial infections) were collected from these patients (Table 2).

Table 1 | Patient baseline data

	PJI	Non-PJI
Patient count	94	56
Male	50	18
Female	44	38
Age (mean ± SD)	59.5 ± 13.58	59.98 ± 10.94
BMI (mean ± SD)	25.53 ± 3.725	25.72 ± 3.601
ASA (mean ± SD)	1.660 ± 0.4764	1.661 ± 0.4778

The inter-group differences among patients were statistically insignificant ($P < 0.05$).
BMI body mass index, *ASA* American Society of Anesthesiologists, *SD* standard deviation.

Table 2 | Bacterial strains found in PJI patients

Strains of bacteria	Case number (*n*)
Gram-positive cocci	
Staphylococcus aureus + MRSE	18 + 2
Staphylococcus epidermidis	29
Other *Staphylococcus* (*S. saprophyticus* + *S. capitis* + *S. lugdunensis*)	3 + 2 + 3
Streptococcus	9
Mycobacterium tuberculosis	1
Gram-negative rods	
Escherichia coli	2
Brucella	2
Xylose-oxidizing *Acinetobacter*	1
Ochrobactrum anthropi	1
Pseudomonas putida	1
Klebsiella pneumoniae	1
Gram-positive rods	
Mycobacterium (Mtb + RGM)	5 + 2
Burkholderia pseudomallei	1
Propionibacterium acnes	1
Corynebacterium striatum	1
Clostridium difficile	1
Bacillus anthracis	1
Fungus	
Candida (*C. albicans* + *C. parapsilosis* + *C. albicans*)	1 + 1 + 1
Pseudohyphae yeast	1
Multiple bacteria	3

Detection of all bacteria included both bacterial cultures and second-generation gene sequencing.

Comparison of multi-model AUC

In comparing the diagnostic results of various models, we tested five intelligent models: DINO v2, EfficientNet v2-S, ResNet-50, CNN, MobileNet v3, and CAMEL2, which include self-supervised, weakly supervised, and supervised types. Although DINO v2, as a self-supervised model for PJI, has an AUC of 1, indicating complete separation in this dataset, it cannot be used as the main result for comparison with other models. The test dataset sensitivity of DINO v2 is 96.1%, and its specificity is 71.15%, suggesting that the model has lower diagnostic efficiency when faced with external datasets it has not been trained (Fig. 1).

Similarly, the sensitivities of ResNet-50, CNN, and MobileNet v3 are 86.10%, 86.51%, and 86.51%, while their specificities are 86.62%, 88.82%, and 91.23%, respectively. The AUCs for these models are as follows: 0.51 for ResNet-50, 0.94 for CNN, and 0.96 for MobileNet v3 (Fig. 2). The summary results of these models were inferior to those of CAMEL2 and EfficientNet v2-S.

Comparison of PJI supervised and weakly supervised models

We further tested the best-performing PJI supervised (EfficientNet v2-S) and weakly supervised (CAMEL2) models, with results presented in two parts for comparison. At the image level, the PJI supervised learning model achieved a sensitivity of 95.96% and a specificity of 89.90%. In contrast, the PJI weakly supervised learning model had a sensitivity of 90.91% but a lower specificity of 82.58% (Fig. 3a), with these sensitivity and specificity values, the thresholds for the PJI supervised and weakly supervised models are 0.1050 and 0.1701, respectively. At the patient level, the PJI supervised learning model adhered strictly to the 2018 ICM diagnostic guidelines, setting more than five high-power fields (30 patches) per slide, achieving a sensitivity of 80.00% and a specificity of 90.38%. In comparison, the weakly supervised learning model used the ROC curve annotation areas from each slide to determine an optimal threshold (20 patches), resulting in a sensitivity of 88.42% and a specificity of 92.31% (Fig. 3b).

Similarly, further analysis of additional evaluation metrics (including accuracy, recall, and F1 score) for both models on positive images indicated that the supervised learning model performs better in image recognition under the 2018 ICM diagnostic standards compared to its weakly supervised counterpart (Fig. 3c). Patient-level analysis based on accuracy, recall, and F1 score (Fig. 3d) confirmed that, under the diagnostic standard of more than 20 patches per slide, the weakly supervised learning model outperforms the supervised learning model based on the 2018 ICM standards.

In the ROC curve results, the image-level supervised learning model achieved an AUC of 0.9652, outperforming the weakly supervised learning model with an AUC of 0.9397 (Fig. 3e). At the patient level, the weakly supervised learning model had an AUC of 0.9460, while the supervised learning model had an AUC of 0.9078 (Fig. 3f). These results indicate that the weakly supervised learning model and the new standard demonstrate excellent diagnostic performance.

From the above data, the PJI supervised model outperforms the PJI weakly supervised model at the image level. However, after further statistical analysis of the prediction values of positive and negative sample images for

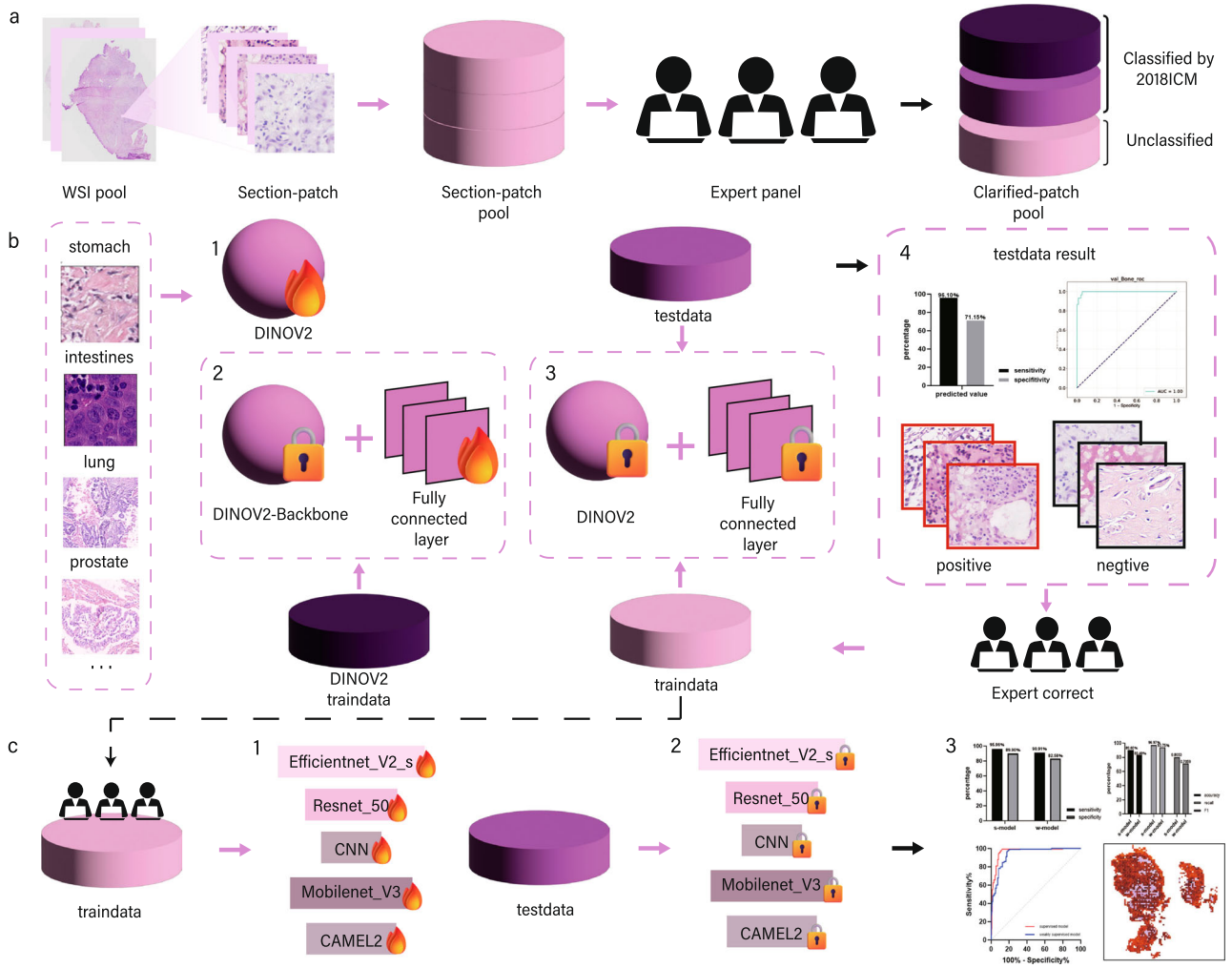

Fig. 1 | The flowchart illustrating the study design. Purple arrows indicate input, black arrows indicate output, flames represent trainable components, and locks denote testing-only components. **a** Data processing: WSI datasets were segmented into 600 × 600-pixel patches and divided for DINO v2 training, testing, and additional training. **b** Self-supervised model and augmentation: **b1** pathological images trained the DINO v2 model. **b2** The DINO v2 backbone extracted features, with the fully connected layer trained. **b3** Test data were reserved for testing, with additional data used for self-supervised tasks. **b4** Self-supervised model testing results. **c** Multi-model training: **c1** expert-reviewed data trained various models. **c2** and **c3** Each model was optimized, tested, and compared.

both the PJI supervised and weakly supervised models, we found that the prediction values of the weakly supervised model are more concentrated compared to the supervised model. Observing only the prediction values of positive samples for both models, the weakly supervised model shows more consistent values, better capturing the common features of positive samples. On the other hand, the supervised model shows a greater difference between the mean values of positive and negative predictions, indicating that the supervised model performs better in distinguishing between positive and negative samples. Both models have their own advantages (Fig. 3g). Meanwhile, the supervised model for PJI, EfficientNet V2-S, was trained for a total of 200 epochs, reaching the lowest loss of 0.00064 at epoch 115 (Fig. 3h). The weakly supervised model for PJI, CAMEL2, reached the lowest loss of 0.07279 at epoch 49 (Fig. 3i).

Human–machine testing of PJI supervised and weakly supervised models

We evaluated the diagnostic results of each doctor against those of the intelligent models using confusion matrices. The results showed that in these 142 images, the dark overlap areas between the intelligent models and the experts had higher values, while the light areas had lower values, indicating that the diagnostic results of the intelligent models were very close to those of the experts. In some cases, the light areas even had a value of zero, suggesting that for these samples, the diagnostic levels of the two models were close to that of the experts (Fig. 4). Although this method does not fully prove that the models can achieve clinical diagnostic standards, it does indicate that our models' image interpretation abilities are close to those of clinical experts.

Visual comparison of PJI supervised and weakly supervised models

Next, we evaluated both the supervised and weakly supervised models for PJI from a clinical perspective. The evaluation focused on three dimensions: accuracy, completeness, and reliability. Reliability assessed whether the visualized images covered all positive areas specified by the 2018 ICM standards; accuracy evaluated whether the model excessively covered areas beyond the positive regions; and integrity referred to whether the visualized images allowed for easy extraction of relevant results. A score of 0 indicates complete disagreement (0–10%); 1 indicates mostly disagreement (10–30%); 2 indicates partial agreement (30–50%); 3 indicates moderate agreement (50–70%); 4 indicates substantial agreement (70–90%); and 5 indicates full agreement (90–100%). In the figure, points closer to the origin indicate dissatisfaction with the visualization results, while points further from the origin indicate satisfaction (Fig. 5).

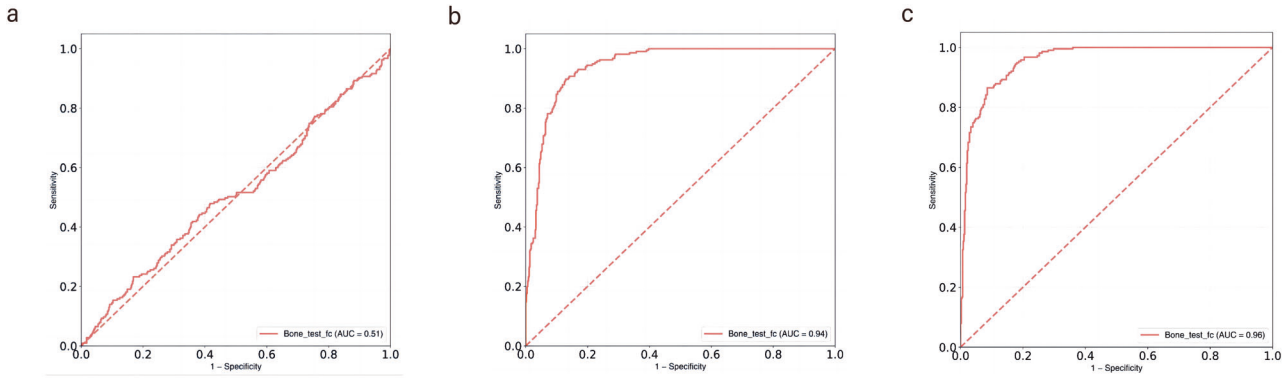

Fig. 2 | The ROC curves for ResNet-50, CNN, and MobileNet v3. This figure presents confusion matrices resulting from pairwise comparisons among three physicians and two models. **a**, **b**, and **c** represent ResNet-50, CNN, and MobileNet v3, respectively. The *x*-axis represents 1-specificity and the *y*-axis represents sensitivity. As 1-specificity increases, sensitivity rises. The AUC, the area under the curve, is close to 1, indicating high diagnostic performance.

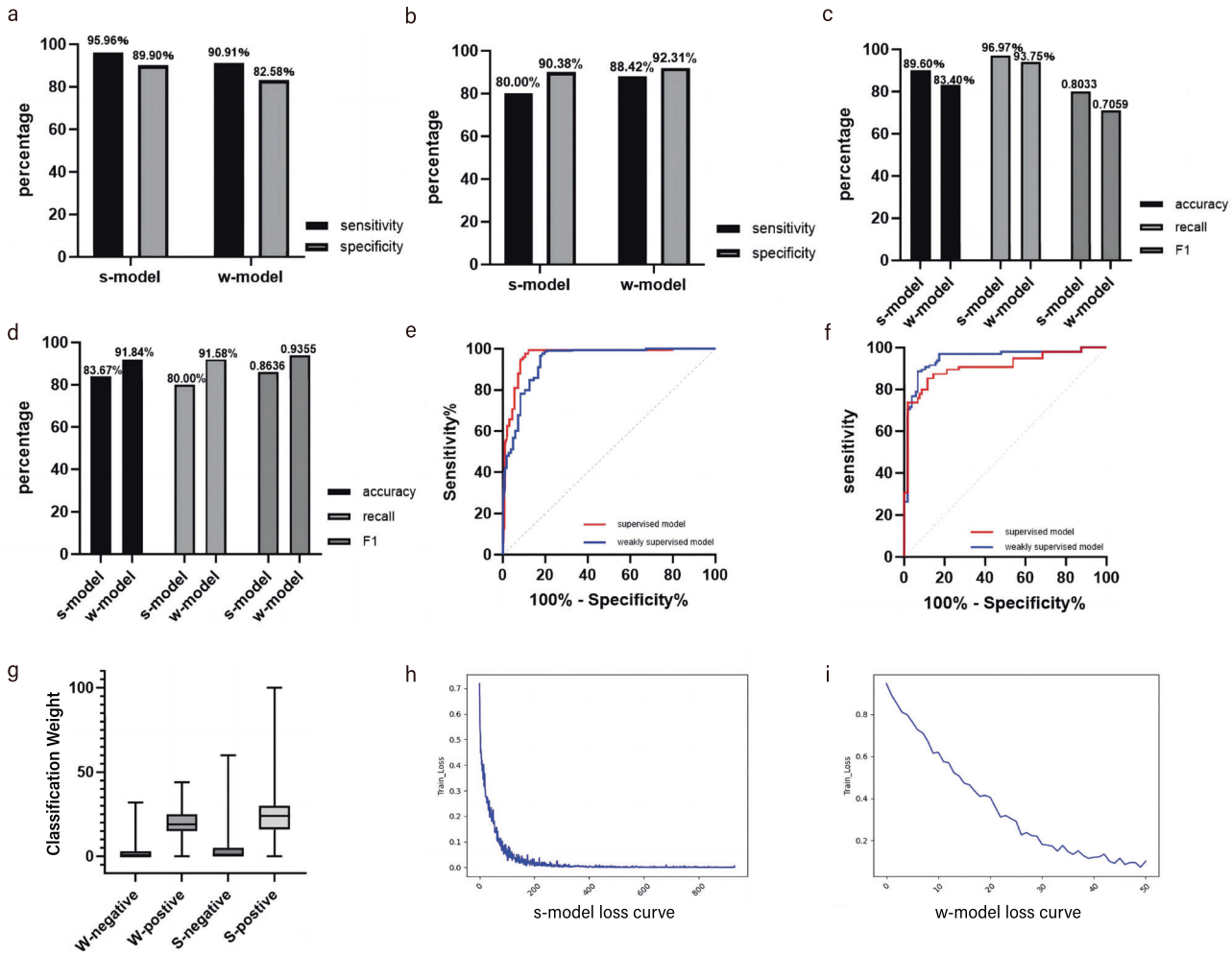

Fig. 3 | The performance of the PJI supervised and weakly supervised learning models. s- refers to the corresponding test results of the PJI supervised learning model, and w- refers to the corresponding test results of the PJI weakly supervised learning model. The red line represents the ROC curve of the PJI supervised learning model, and the blue line represents the ROC curve of the PJI weakly supervised learning model. **a** Image-level comparison of sensitivity and specificity. **b** Patient-level comparison of sensitivity and specificity. **c** Image-level accuracy, recall, and F1 score of the models. **d** Patient-level accuracy, recall, and F1 score of the models. **e** Image-level ROC curves for the two models. **f** Patient-level ROC curves for the two models. **g** The degree of data dispersion at the image level. The weakly supervised model has a mean ± standard deviation of 0.03433 ± 0.02211 for the negative set and 0.2059 ± 0.05993 for the positive set; the supervised model has 0.03780 ± 0.02328 and 0.2614 ± 0.1009, respectively. **h** Loss curves for the PJI supervised learning model. **i** Loss curves for the PJI weakly supervised learning model.

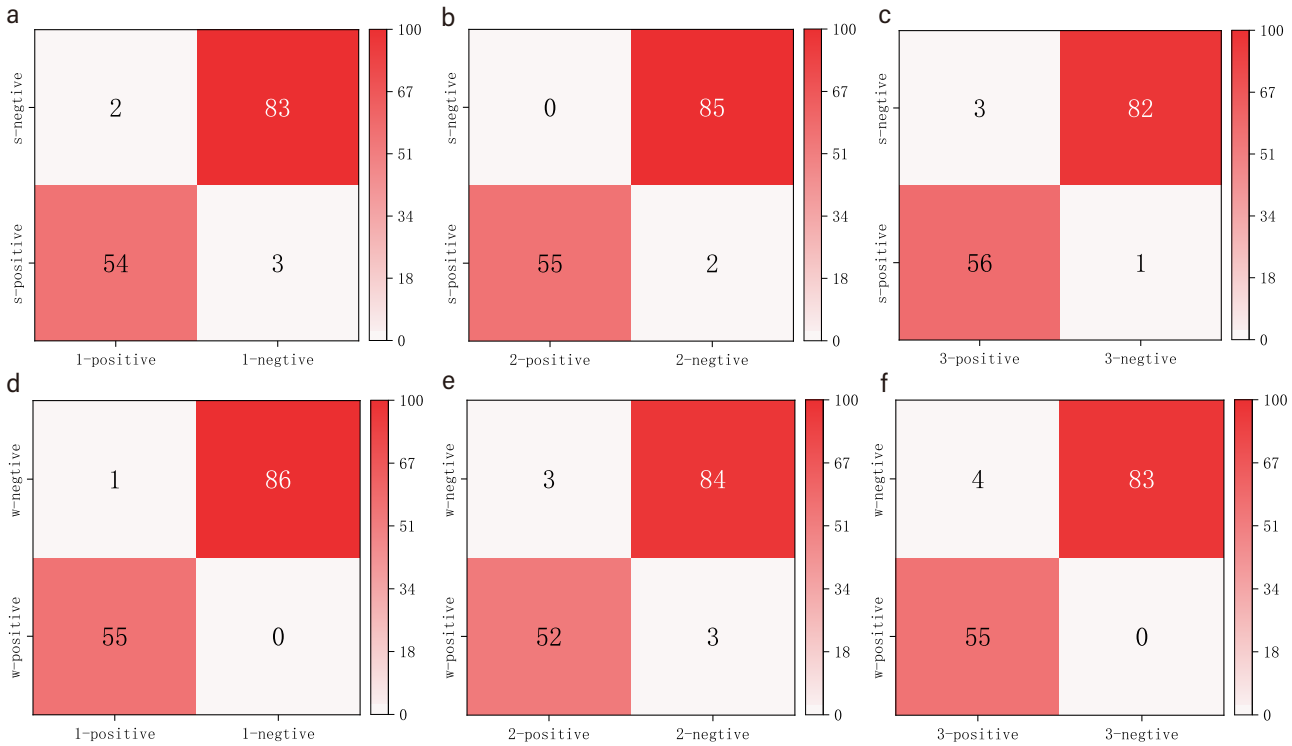

Fig. 4 | The human–machine comparison test result. 1/2/3 correspond to the diagnostic results of Experts 1/2/3, indicated on the horizontal axis; the symbols s/w represent the diagnostic results of the PJI supervised and weakly supervised models, indicated on the vertical axis. **a**, **b**, and **c** show the confusion matrix results comparing PJI supervised models with Experts 1/2/3, while **d**, **e**, and **f** present the confusion matrix results comparing PJI weakly supervised models with Experts 1/2/3. The darker the red, the larger the number. The top-left and bottom-right squares represent areas where the experts' diagnoses and the model's diagnoses are the same, while the other squares represent areas where the diagnoses differ.

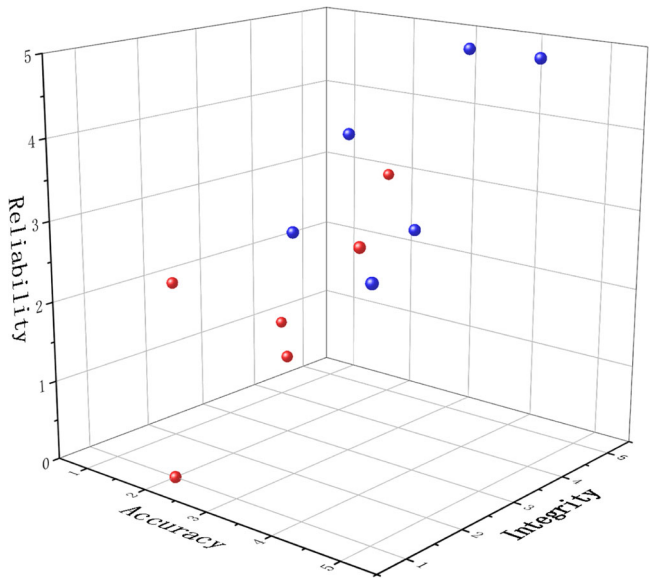

Fig. 5 | The visual differences between the supervised learning (s-model) and weakly supervised learning (w-model) models. The three-dimensional data formed by the w-model is notably distant from the coordinate point, whereas the s-model is closer to the coordinate origin. This indicates that, on average, the w-model outperforms the s-model in terms of accuracy, completeness, and reliability, thereby displaying superior visualization effects.

The results indicated that the weakly supervised learning model outperforms its supervised counterpart. It was not only more comprehensive and accurate in identification but also yielded reliable diagnostic outcomes solely from the regions identified by the model (Fig. 5). This suggests that, from a clinical perspective, the weakly supervised learning model excels in segmentation. Based on patient-level results, we hypothesize that this method might involve additional infection indicators beyond neutrophil aggregation features, which were not previously observed. Furthermore, medical interpretation of the annotated areas aids in further optimizing the pathological diagnosis of PJI.

Analysis of visual results

The visual outcomes of the PJI intelligent pathological diagnosis model reveal that the weakly supervised learning model provides notably finer and more detailed regions. Specifically, tissue images show not only neutrophil aggregation but also a loss of the tissue's original structure, resulting in a more porous appearance (Fig. 6a). Additionally, there are observable differences in the cytoplasm and nuclear morphology of neutrophils relative to their proximity to blood vessels (indicated by the red arrows). Neutrophils closer to the blood vessels have less cytoplasm and a rod-shaped nucleus, while those farther away exhibit more cytoplasm and a lobular-shaped nucleus (Fig. 6b).

Discussion

The rapid and accurate diagnosis of PJI has always been a challenge in the field of arthroplasty. The 2018 ICM criteria have limitations[7]: serological markers and nuclear imaging tests are highly sensitive, but their specificity is relatively low, thus making them poor indicators for definitive diagnosis[26–29]. Pathogen culture of joint fluid and tissues is the gold standard for infection confirmation but it relies heavily on the experience of the physician[30,31]. Moreover, some pathogens (e.g., fungi and mycobacteria) require stringent culture conditions, resulting in prolonged diagnostic times and delayed treatment[32,33]. Genetic sequencing is faster but limited by contamination, equipment availability, and cost[34–36]. Pathological examination has relatively high specificity but its sensitivity is lower, requiring the pathologist's experience. Moreover, during surgery, PJI pathological examination must

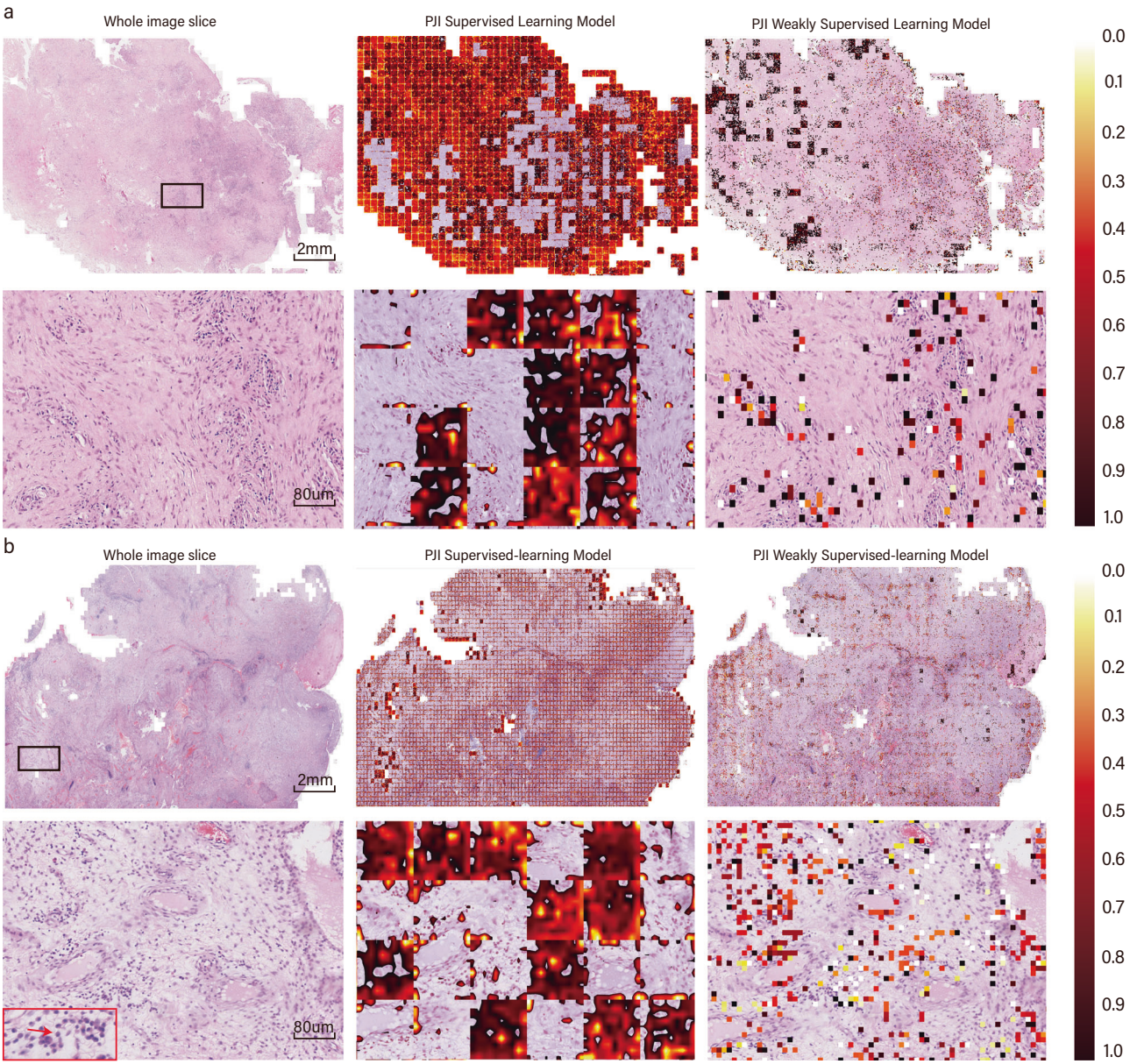

Fig. 6 | The visual outcomes of the PJI intelligent pathological diagnosis model. From left to right, the images represent a whole image slide, a visualization heatmap of the PJI supervised learning model, and a visualization heatmap of the PJI weakly supervised learning model. The color gradient from light to dark indicates diagnostic weight from low to high. **a** The tissue shows not only an aggregation of neutrophils but also a loss of its original structure, becoming more porous. **b** Differences in the cytoplasm and nuclear morphology of neutrophils are observed, depending on their proximity to blood vessels (indicated by the direction of the red arrow).

be completed quickly to allow clinicians to make an accurate diagnosis promptly, aiding the surgical procedure[8,9,14,15]. We are leveraging AI to optimize pathological diagnosis methods for PJI, leveraging AI's high throughput, accuracy, and reproducibility to overcome the limitations of manual detection.

AI in image recognition has enhanced clinical diagnosis by improving efficiency and accuracy in radiological and pathological exams[18–21]. While neural networks excel at identifying tumors and specific tissues[37], recognizing infected tissues remains challenging due to the dispersed and varied nature of cells in infection. Infected tissues often have non-specific changes and unclear boundaries, making identification challenging. Annotating neutrophils across entire slides is labor-intensive due to their small size and diversity. Direct machine learning recognition on whole slides is difficult, so relying solely on neutrophil segmentation for PJI diagnosis can lead to misidentification and reduced accuracy.

The direct application of neural-network segmentation models for neutrophil identification has yielded unsatisfactory results. Hence, we used a classification model based on the 2018 ICM guidelines. By defining *at least 5 neutrophils in high-power fields* as the criterion, we trained a ResNet-34 based supervised learning model for PJI image-level diagnosis[38]. While the ResNet model has shown high accuracy in diagnosing cancers[39–43], it achieved only 93.22% accuracy and 96.49% recall in PJI image-level diagnosis, and an AUC of 0.81 at the patient level[38]. Consequently, it is not yet a reliable pathological recognition model.

Because the classification model addresses the entire area in identifying images in positive patches, it might only cover a small fraction of neutrophils and their surroundings within high-power field images (neutrophil pixels ~6 × 6 within a high-power field image of ~1800 × 1200). This renders it susceptible to interference from other nonfeature areas. To prevent an excess of normal tissues or noninfectious inflammatory tissues in the

training images, we downscaled the classification recognition units from an approximately high-power field (1800 × 1200 pixels) to the commonly used 600 × 600 pixels.

EfficientNet balances network depth, width, and image resolution better than ResNet, improving speed and accuracy while reducing parameters[44,45]. In this study, we leveraged EfficientNet v2-S for training post-classification patch images resulted in a PJI intelligent pathological classification model with improved accuracy in recognizing segmented pathological images. The model showed strong performance in both internal and external image-level tests and acceptable diagnostic capability in patient-level validation compared to manual pathology results.

By visualizing diagnostic weights on test set images using the PJI supervised learning model, we obtained a *heatmap* showing the model's diagnostic performance. Previous studies have shown that infection-related features extend beyond just neutrophil aggregation[13]. Solely relying on neutrophils can reduce diagnostic sensitivity, similar to identifying apples as the key feature of an apple tree, even though not all apple trees have apples. Additionally, the existing PJI pathological diagnostic standards lack explicit quantification of other infection-related features within the infected area and do not have a clear definition of infection boundaries, limiting the utilization of other infection-related characteristics. Our analysis of the model's *heatmap* revealed that identified regions did not fully match necrotic or exudative areas around neutrophils (Fig. 6). This discrepancy may be due to the classification model's limitation in pinpointing exact locations, a task better suited for neural-network segmentation models.

Therefore, this study employed a weakly supervised learning CAMEL2 model to construct an approximate segmentation model. This model identifies diagnostic regions for PJI from classification-labeled patches, effectively handling fuzzy boundaries in clinical images[46]. It converts a classification model into a segmentation model by fine-tuning with labeled data, dividing patches into grid segments, and creating multi-instance learning labels with diagnostic potential but limited precision. Through self-supervised learning of these tendency diagnosis labels, specific diagnostic information for each instance can be obtained. Pixel-level diagnosis can be achieved through iterative learning loops, thereby facilitating the training of approximate segmentation models[47].

The weakly supervised learning model matched the performance of the supervised model in image-level tests, achieving similar accuracies, recall rates, and ROC curves. However, in patient-level testing, the weakly supervised learning model outperformed its supervised-learning counterpart. By adjusting the area threshold for the recognition regions, we substantially enhanced the sensitivity of PJI pathology diagnosis without compromising specificity. Our proposed criterion for diagnosing PJI requires more than 20 units of 600 × 600 area containing over two neutrophils (excluding those within vessels) on a single pathological slide. This reduces the diagnostic fields from five high-power fields to three. Moreover, the *heatmap* generated by the weakly supervised model closely aligns with neutrophils and necrotic areas. Diagnosis was based on the annotated images from this heatmap, and visualization also revealed structural changes and tissue looseness in addition to neutrophil aggregation (Fig. 6a). Neutrophil morphology distribution might also provide insights for PJI diagnosis and treatment (Fig. 6b). This research will help us look into how the infection process affects tissues and advance pathological studies.

Methods

Dataset establishment

This retrospective study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the Ethics Committee of Chinese PLA General Hospital (Date: 29/02/2024, No. S2024-032-01). The need for informed consent was waived because the study utilized medical record data obtained from previous clinical diagnosis and treatment. With this approval, we were authorized to access clinical data, including pathological images, from patients undergoing revision surgery after joint replacement at the Chinese PLA General Hospital.

According to the 2018 ICM guidelines for pathological tissue collection and processing, we collected 462 frozen pathological slides from 150 patients at the Chinese PLA General Hospital, ensuring that each patient had at least three samples of periprosthetic soft tissues obtained during revision surgery (including soft tissues on the femoral side, tibial side, and synovial tissues)[7]. For all frozen slides, the layers for cutting were selected by the pathology department of PLAGH, and the slides were routinely hematoxylin and eosin (H&E) stained. Subsequently, the Unic-PRCICE-610 digital scanner (40×: 0.25 μm/pixel) was used to convert the frozen pathological slides into whole slide images (WSIs). The generated images were manually inspected individually to ensure the image dataset preprocessing.

Our prior research on intelligent pathological diagnosis of PJI showed that excessively large training patches could lead to imprecise identification of minute neutrophils[38]. To achieve better diagnostic results with deep learning networks, we reduced five or more neutrophils, as specified in the current criterion, to two or more neutrophils, with pixel area correspondingly lowered from 1200 × 1800 to 600 × 600 pixels.

This adjustment ensured that image annotations were in line with the existing diagnostic requirements while enhancing the training efficacy of intelligent models.

We used the OpenSlide tool to segment the WSI, generating 600 × 600-pixel patches. Then, we applied Otsu's method to filter the foreground tissue images and recorded their coordinates to screen all the segmented patches. Finally, we manually selected the effective patches for confirmation. Careful attention was paid to maintain image quality no lower than 0.25 μm/pixel. Ultimately, we obtained a total of 1.6 million (1,588,787) patches.

We first regrouped the segmented and selected effective dataset into DINO v2 training data, test data, and additional training data, ensuring that each group was independent and non-overlapping (Fig. 1a). Since the additional training data were the largest dataset, we found it challenging to annotate it within a short period. Therefore, we employed a self-supervised model to assist with the annotation. This model relies on a large amount of unrelated data and uses minimal training data to meet the requirements for preliminary image classification.

Establishment of the PJI self-supervised learning model (DINO v2 model)

The DINO v2 model was initially trained on multi-organ tumor images. In this study, we used it as a feature extraction tool and provided the previously obtained small amount of training data as input to the network. To prevent the small dataset from contaminating the model and leading to suboptimal results, we set the DINO v2 model as a non-trainable backbone, added a fully connected (FC) layer, and trained only the FC layer. Due to the limited number of trainable parameters and the small number of positive samples, there was no significant improvement in training outcomes as the data volume increased during the model testing phase. Since both DINO v2 and the FC layer were locked, the data were used solely to generate training sets for other models without parameter adjustments, which did not accurately reflect the performance of the DINO v2 model. Therefore, we present the average results from multiple trainings, with a sensitivity of 96.1% and a specificity of 71.15% (Fig. 1b). On the other hand, by adjusting the parameters of the FC layer, we achieved complete separation for a specific test set, as shown in the figure, with an AUC of 1. However, this only indicates a difference between negative and positive samples and does not prove high model accuracy, as different parameters are required to achieve complete separation for another test set. In other words, the optimal parameters for the model are not fixed and cannot be used as a routine diagnostic model. This approach generated 22,457 negative patches and 4596 positive patches for training other models, significantly reducing experimental time and labor costs.

The labels generated during inference by the self-supervised model were used as the training set for EfficientNet v2-S. Due to the presence of incorrect annotations in the labels, we introduced an active learning paradigm. After review by the expert panel, they were used as input for various experimental models. Supported by the large datasets generated in this

research step, each model was able to achieve optimal performance through parameter adjustment and structural optimization. This also provided a sufficient data reserve for our subsequent experiments (Fig. 1c).

Manual annotation

After ensuring sufficient data preparation, we also made adequate personnel preparations. The pathology expert panel consisted of three pathologists with over 15 years of experience. Two experts performed annotations independently, and in case of disagreement, a third expert was consulted to make the final decision. For patches that still caused disagreement, consensus was reached by comparing notes within the panel. The diagnostic criterion for positive patches was based on the presence of at least two neutrophils within a 600 × 600-pixel area (excluding neutrophils within vessels or clustered blood clots). These patches were exclusively obtained from patients with clinically diagnosed PJI.

Dataset overview

Next, we clarified the different uses of the datasets in the study. To ensure comparability between different PJI intelligent pathological diagnostic models, all models utilized the same training set data. The training set, classified by DINO v2 and reviewed by an expert panel, comprised 22,457 negative patches and 4596 positive patches. This dataset was used as the traindata for training the models. Simultaneously, the test set (testdata) was randomly selected from 147 slices not used in training, including 456 negative patches and 215 positive patches, for model testing. Additionally, 142 images, not involved in the test and validation sets, were randomly selected to form a human–machine comparison test set. The data groups were independent of each other at both the patient and image levels.

PJI supervised learning model

To tailor the existing PJI pathological diagnostic standards to the requirements of intelligent diagnosis, we set the learning objective as the presence of a sufficient number of neutrophils within a unit area. We built a supervised learning model using EfficientNet v2-S as the backbone. The network begins with a convolutional layer (conv 3 × 3) with a stride of 2 and progresses through a series of Fused-MBConv and MBConv blocks, each defined by specific kernel sizes (k 3 × 3) and channels. Strides vary between 1 and 2, depending on the layer. Some blocks are marked with Squeeze-and-Excitation (SE) ratios. The network culminates in a conv 1 × 1 layer followed by pooling and a FC layer, ultimately leading to an output with 1280 channels. Each layer or block specifies the number of channels it outputs and the total layers it comprises. The model was implemented in TensorFlow using Adam optimizer (Fig. 7). The model performance was compared in terms of the weights of every 100 steps on the validation set until the model accurately identified a sufficient number of neutrophils per unit area.

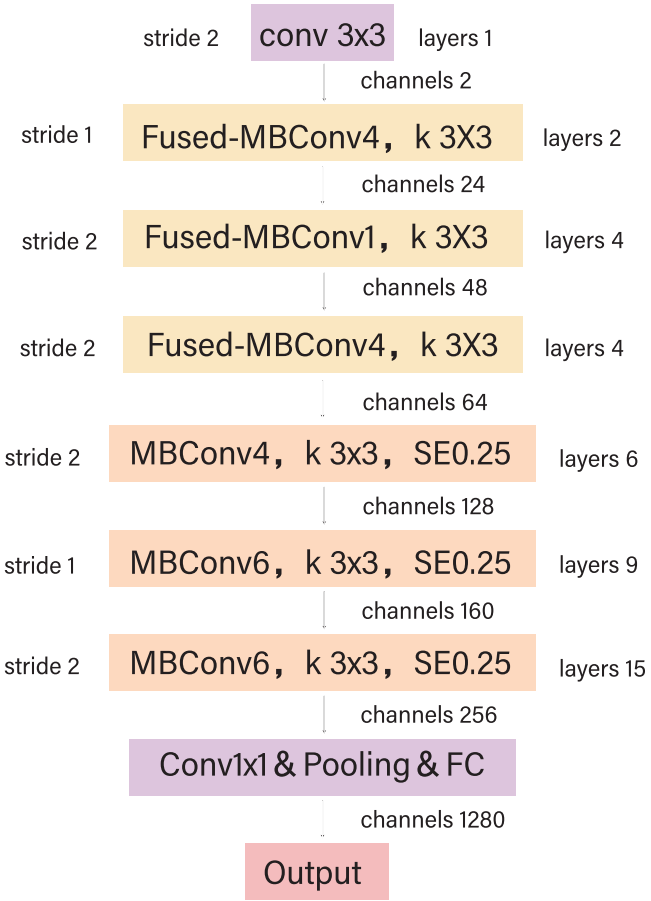

Fig. 7 | Architecture of PJI supervised learning model. Using EfficientNet v2-S as the backbone, the model begins with a convolutional layer (conv 3 × 3) with a stride of 2, followed by a series of Fused-MBConv and MBConv blocks, where strides vary between 1 and 2. Some blocks include SE (Squeeze-and-Excitation) ratios. The network concludes with a conv 1 × 1 layer, followed by pooling and a fully connected layer, resulting in an output of 1280 channels. The model is implemented in TensorFlow using the Adam optimizer, and its weights are compared on the validation set every 100 steps.

PJI weakly supervised learning model

To address this labeling challenge, we transform the problem of coarse-grained labeling classification into a fully supervised fine-grained image classification task using CAMEL2. Additionally, we extend supervision information by generating pseudo-labels for each image. In this study, we adopt the concept of Multi-Instance Learning to expand annotation information and construct a high-quality instance-level dataset from the original image-level dataset using instance-level labels. Terahertz images are partitioned into grids of varying sizes, with each image becoming an independent Bag. Each small grid, as an instance, belongs to its corresponding Bag and shares label information. The instance corresponding to an image is represented as $X = \{x_1, x_2, x_3 \ldots\ldots x_n\}$, and the corresponding instance label $\{y_1, y_2, y_3 \ldots\ldots y_n\}$ belongs to the Bag-level label Y, satisfying Formula 1.

$$Y = \begin{cases} 1 & if\ \exists y_i = 1 \\ 0 & else \end{cases} \quad (1)$$

Each time a positive sample and a negative sample are selected to form a patch-level input image pair, the segmented patches of different magnifications have unique label information, which is completely derived from image-level coarse-grained labels. In order to retain more pathological information, in CAMEL2, we choose to expand the image size as much as possible, and finally, the image size of each patch level is 2048 × 2048. The images of different magnifications are segmented into $N \times N$ grid instances with equal size ($N = 256$). Process each instance through the model and apply a softmax operation to obtain the predicted probability values for each instance. In negative samples, every instance inherits the WSI level label, and for these instances, we assign a label of 0. For instances from positive samples, we hypothesize that at least K% are directly related to the disease. Among the positive instances, we select the top K% with the highest confidence as positive instances by sorting. During backpropagation, we use cross-entropy loss to update the parameters, which is represented as follows:

$$Loss = -\sum_j (y_j \log p_j + (1 - y_j) \log(1 - p_j)) \quad (2)$$

where y_j represents the instance-level label and $p_j = Soft\max(\text{model}(x_{100_j}, x_{400_j}))$ represents the prediction of model as the relevant probability value (Fig. 8).

PJI self-supervised learning model

To achieve rapid and accurate patch annotation and to differentiate between infected and non-infected images, we employed a self-supervised learning

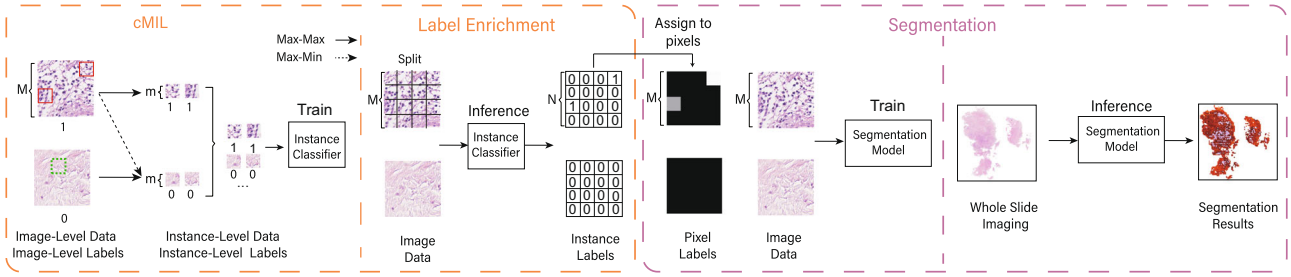

Fig. 8 | Architecture of the PJI weakly supervised learning model. The model has three components: cMIL, Label Enrichment, and Segmentation. cMIL performs fine-grained segmentation, Label Enrichment extends image data, and Segmentation re-segments the image. Using CAMEL2, we transform coarse-grained labeling into a fine-grained classification task, generating pseudo-labels and applying Multi-Instance Learning (MIL) to create an instance-level dataset. Terahertz images are divided into grids that share label information with the entire image. Positive and negative samples form patch-level pairs, with images expanded to 2048 × 2048 pixels, segmented into 256 × 256 grid instances, and processed with softmax to obtain probabilities. In negative samples, instances inherit a label of 0, while in positive samples, the top K% of confident instances are selected as positive. Cross-entropy loss updates the model during backpropagation.

architecture using DINO v2 with Vision Transformer (ViT) (Fig. 9). For the pretraining protocol, we utilized the state-of-the-art self-supervised training paradigm DINO v2 on the constructed large-scale pan-cancer pathology dataset, with ViT-L/16 as the chosen network architecture. For pretraining on the constructed large-scale pan-cancer pathology dataset, we employed the state-of-the-art self-supervised training paradigm DINO v2 with ViT-L/16 as the chosen network architecture.

Framework of DINO v2 model[48]. At the image processing level, a teacher–student model structure was employed, with two networks using different data augmentation methods for data input. The teacher model is computed using the exponential moving average (EMA) from the student model, which differs from the concept of a distillation model.

The architecture of student g_{θ_s} and teacher g_{θ_t} consists of a main network backbone f (ViT) and a projection head $h : g = h \circ f$. The projection head includes a three-layer multi-layer perceptron (MLP) with 2048 hidden dimensions, followed by layer normalization and a FC layer with K dimensions. For softmax, both use temperature softmax, which controls the sharpness of the output distribution.

$$P_i(x)^{(j)} = \exp\left(\frac{g_{\theta_i}(x)^{(j)}}{\tau_i}\right) \Big/ \sum_{k=1}^{K} \exp\left(\frac{g_{\theta_i}(x)^{(k)}}{\tau_i}\right), \tau_i > 0, i = s, t \quad (3)$$

Loss:

$$Min_{\theta_s} \sum_{x \in \{x_1^g, x_2^g\}} \sum_{x' \in V, x' \neq x} H\left(P_t(x), P_s(x')\right), H(a, b) = -alogb \quad (4)$$

Average EMA:

$$\theta_t \leftarrow \lambda\theta_t + (1 - \lambda)\theta_s, \lambda \quad \text{cosine from 0.996 to 1} \quad (5)$$

The biggest change in DINO v2 compared to DINO is the utilization of Patch tokens. First, we apply patch masking to the two augmented views u and v, obtaining their masked views $\hat{u}$ and $\hat{v}$. Taking $\hat{u}$ as an example for simplification, the student network outputs the projection of its patch tokens for the masked view $\hat{u}$ as $\hat{u}_s^{patch} = P_\theta^{patch}(\hat{u})$, while the teacher network's projection of patch tokens for the unmasked view u is $u_t^{patch} = P_{\theta'}^{Patch}(u)$. We define the training objective of iBOT here as

$$\mathscr{L}_{MIM} = -\sum_{i=1}^{N} m_i \cdot P_{\theta'}^{patch}(u_i)^T \log P_\theta^{patch}(\hat{u}_i) \quad (6)$$

The DINO v2 model learns representations of unlabeled pathological sections through a self-supervised learning loss function.

Model testing

We selected 147 slices not used in training, which were randomly sampled to form the model test set, including 456 negative patches and 215 positive patches. Using the same test set, we evaluated all models, including DINO v2, EfficientNet v2-S, ResNet-50, a self-constructed CNN (with five convolutional layers and max pooling), MobileNet v3, and CAMEL2. We first tested all models, using the area under the ROC curve (AUC) as the primary reference indicator to compare the diagnostic performance of each model at the PJI image level. For models with poor diagnostic performance at the PJI image level, such as ResNet-50, CNN, and MobileNet v3, we found that adjusting parameters could not improve their diagnostic accuracy, so we promptly excluded them. For the top-performing models, such as EfficientNet v2-S and CAMEL2, we conducted detailed data analysis and statistical evaluation, including sensitivity, specificity, accuracy, recall, F1 score, and ROC curve. After thoroughly analyzing this data, we found through human–machine comparison that the performance of these two models was very close to that of experts.

We then conducted comprehensive patient-level testing. In patient-level (entire slice) diagnostic testing, the pathological diagnosis was typically made by a junior pathologist (with <15 years of experience) and subsequently cross-validated by two senior pathologists with over 15 years of experience. This validation, combined with other validation methods, such as bacterial culture and second-generation gene sequencing, is deemed, to some degree, the gold standard. The supervised learning model was tested against the criterion of five high-power fields (30 patches) per slice, as stipulated by the 2018 ICM. Conversely, in the weakly supervised learning model, patient-level diagnosis was established by combining the annotated diagnostic regions per slice and plotting ROC curves. Subsequently, we compared the supervised-learning PJI intelligent pathological diagnostic model and its weakly supervised-learning counterpart, collected parameters such as sensitivity, specificity, recall, accuracy, F1 score, and ROC curves, and performed patient-level diagnostic testing.

To further analyze whether the models could accurately detect PJI regions, we conducted model visualization studies. To assess the visual effectiveness of supervised learning and weakly supervised learning models in identifying infected regions, we formed a panel of pathologists to evaluate six visualized outcomes generated by each of these models. The evaluation criteria primarily involved the accuracy, completeness, and reliability of the annotated regions. Reliability is used to assess whether the expert panel has covered all the positive areas specified by the 2018 ICM Standards in the visualized images. Accuracy checks if the model excessively covers areas beyond the positive regions. Integrity refers to whether the visualized images allow for easy and accurate interpretation of the results (Fig. 5). The expert panel used a 5-point scale to statistically assess each criterion and eventually compared the visual effectiveness between the supervised learning and weakly supervised learning models. Through subjective quantification and

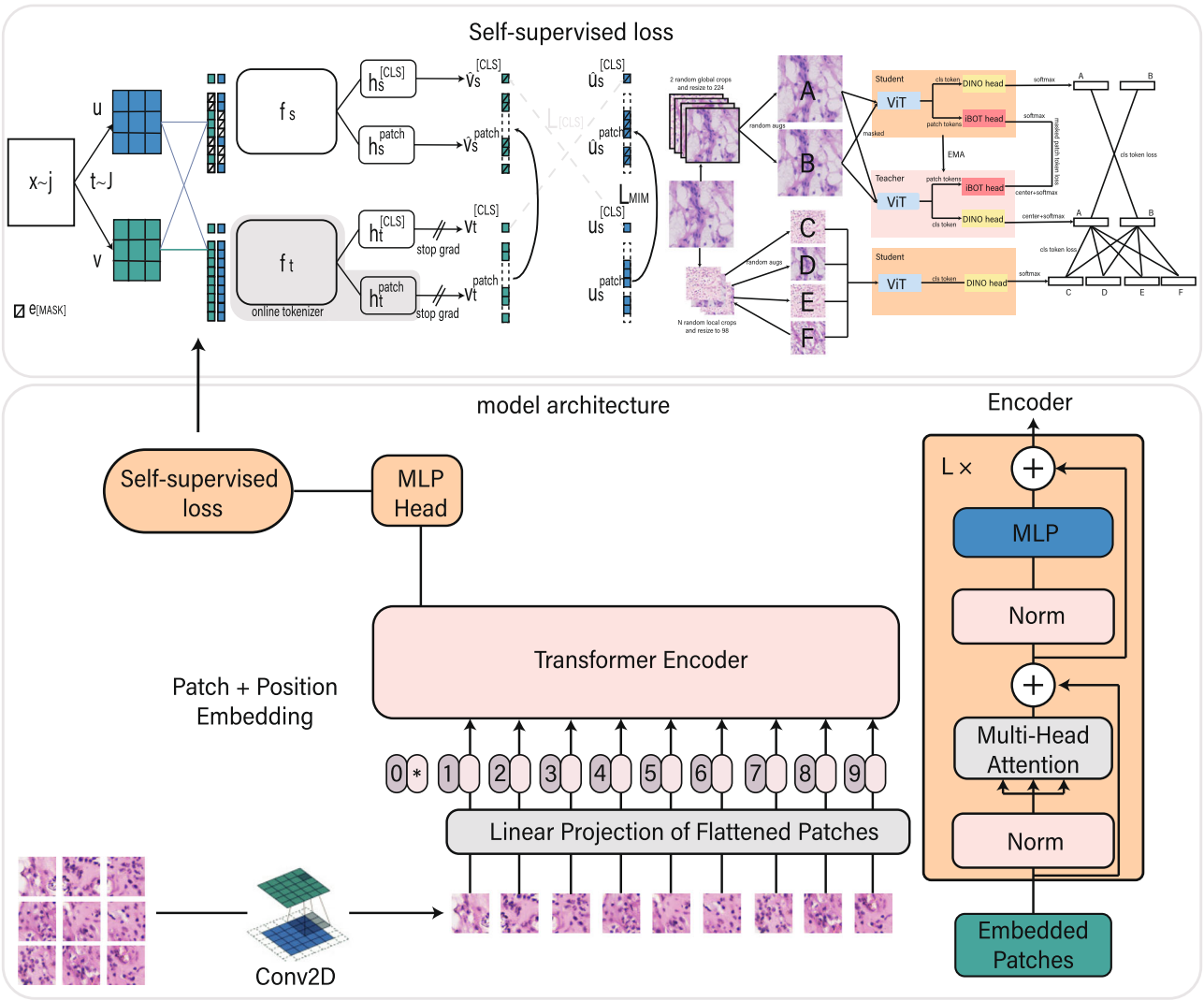

Fig. 9 | Architecture of PJI self-supervised learning model. A teacher–student model structure with different data augmentations is used, where the teacher model is updated using the student model's exponential moving average (EMA). Both networks feature a ViT backbone, a projection head, and use temperature softmax. DINO v2 introduces patch tokens and masking, with the student network projecting masked views and the teacher network projecting unmasked views. The training objective of iBOT is defined based on this setup. DINO v2 model learns representations of unlabeled pathological sections through a self-supervised learning loss function.

case analysis by experts, we finally conducted diagnostic threshold research with the help of weakly supervised models.

The baseline data were subjected to Chi-square statistical analysis using SPSS 26.0, where the definition of good reliability was considered >0.9. This analysis aimed to assess potential differences in age, sex, BMI, and ASA scores between patients with PJI and their non-PJI counterparts to validate the consistency between the two groups. All analyses related to sensitivity, specificity, accuracy, recall, F1 score, ROC curve, and other data for the intelligent models were conducted using GraphPad Prism 10.

Data availability

The datasets generated and analyzed during the current study are not publicly available due to the need for confidentiality of patient data in Chinese PLA General Hospital, but are available from the corresponding author on reasonable request. The imaging materials were acquired from the Imaging Department of the Chinese PLA General Hospital, and they have not given their permission for researchers to share their data. The imaging materials requests can be made to the Imaging Department of the Chinese PLA General Hospital via this email: jfjzyygw@163.com. The statistical results on imaging data, please contact the corresponding authors Ming Ni and Jie Li, via these emails niming301@163.com and lijiek812@foxmail.com.

Code availability

The code for weakly supervised learning was open-sourced at https://github.com/ThoroughFuture/CAMEL2. We also open-sourced the core components of the deep learning system at https://github.com/ThoroughFuture/PathFrame.

Received: 19 January 2024; Accepted: 16 October 2024;
Published online: 26 October 2024

References

1. Evans, J. T. et al. How long does a knee replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. *Lancet* **393**, 655–663 (2019).
2. Kapadia, B. H. et al. Periprosthetic joint infection. *Lancet* **387**, 386–394 (2016).

3. Tande, A. J. & Patel, R. Prosthetic joint infection. *Clin. Microbiol. Rev.* **27**, 302–345 (2014).
4. Lenguerrand, E. et al. Risk factors associated with revision for prosthetic joint infection after hip replacement: a prospective observational cohort study. *Lancet Infect. Dis.* **18**, 1004–1014 (2018).
5. Lenguerrand, E. et al. Risk factors associated with revision for prosthetic joint infection following knee replacement: an observational cohort study from England and Wales. *Lancet Infect. Dis.* **19**, 589–600 (2019).
6. Chang, C. H. et al. Increased periprosthetic hip and knee infection projected from 2014 to 2035 in Taiwan. *J. Infect. Public Health* **13**, 1768–1773 (2020).
7. Parvizi, J. et al. The 2018 definition of periprosthetic hip and knee infection: an evidence-based and validated criteria. *J. Arthroplasty* **33**, 1309–1314.e2 (2018).
8. Lee, P., Chandel, N. S. & Simon, M. C. Cellular adaptation to hypoxia through hypoxia inducible factors and beyond. *Nat. Rev. Mol. Cell Biol.* **21**, 268–283 (2020).
9. Moffarah, A. S., Al Mohajer, M. & Hurwitz, B. L. Skin and soft tissue infections. *Microbiol. Spectr*. **4** https://doi.org/10.1128/microbiolspec.DMIH2-0014-2015 (2016).
10. Jabbour, J. F. & Kanj, S. S. Gram-negative skin and soft tissue infections. *Infect. Dis. Clin. N. Am.* **35**, 157–167 (2021).
11. Sigmund, I. K. et al. Diagnostic accuracy of neutrophil counts in histopathological tissue analysis in periprosthetic joint infection using the ICM, IDSA, and EBJIS criteria. *Bone Joint Res.* **10**, 536–547 (2021).
12. George, J. et al. Are frozen sections and MSIS criteria reliable at the time of reimplantation of two-stage revision arthroplasty? *Clin. Orthop. Relat. Res.* **474**, 1619–1626 (2016).
13. AbdullGaffar, B. et al. Should we abandon routine microscopic examination in bariatric sleeve gastrectomy specimens? *Obes. Surg.* **26**, 105–110 (2016).
14. Sigmund, I. K., Luger, M. & Windhager, R. Diagnosing periprosthetic joint infections: a comparison of infection definitions: EBJIS 2021, ICM 2018, and IDSA 2013. *Bone Joint Res.* **11**, 608–618 (2022).
15. Tohtz, S. W. et al. Validity of frozen sections for analysis of periprosthetic loosening membranes. *Clin. Orthop. Relat. Res.* **468**, 762–768 (2010).
16. Goldenberg, S. L., Nir, G. & Salcudean, S. E. A new era: artificial intelligence and machine learning in prostate cancer. *Nat. Rev. Urol.* **16**, 391–403 (2019).
17. Wall, J. & Krummel, T. The digital surgeon: how big data, automation, and artificial intelligence will change surgical practice. *J. Pediatr. Surg.* **55**, 47–50 (2020).
18. Yuan, C. et al. The application of deep learning based diagnostic system to cervical squamous intraepithelial lesions recognition in colposcopy images. *Sci. Rep.* **10**, 11639 (2020).
19. Han, S. S. et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PLoS ONE* **13**, e0191493 (2018).
20. Hsiao, Y. J. et al. Application of artificial intelligence-driven endoscopic screening and diagnosis of gastric cancer. *World J. Gastroenterol.* **27**, 2979–2993 (2021).
21. Sharma, P. & Hassan, C. Artificial intelligence and deep learning for upper gastrointestinal neoplasia. *Gastroenterology* **162**, 1056–1066 (2022).
22. Chen, W. F. et al. Recognition rate advancement and data error improvement of pathology cutting with H-DenseUNet for hepatocellular carcinoma image. *Diagnostics* **11**, 1599 (2021).
23. Delgado-Ortet, M. et al. A deep learning approach for segmentation of red blood cell images and malaria detection. *Entropy* **22**, 657 (2020).
24. Kuo, F. C., Hu, W. H. & Hu, Y. J. Periprosthetic joint infection prediction via machine learning: comprehensible personalized decision support for diagnosis. *J. Arthroplasty* **37**, 132–141 (2022).
25. Yeo, I. et al. The use of artificial neural networks for the prediction of surgical site infection following TKA. *J. Knee Surg.* **36**, 637–643 (2023).
26. Rohe, S. et al. C-reactive protein during the first 6 postoperative days after total hip arthroplasty cannot predict early periprosthetic infection. *Arch. Orthop. Trauma Surg.* **143**, 3495–3503 (2023).
27. Plate, A. et al. Synovial C-reactive protein features high negative predictive value but is not useful as a single diagnostic parameter in suspected periprosthetic joint infection (PJI). *J. Infect.* **78**, 439–444 (2019).
28. Lee, S. H. et al. Do serum C-reactive protein trends predict treatment outcome in patients with knee periprosthetic joint infection undergoing two-stage exchange arthroplasty? *Diagnostics* **12**, 1030 (2022).
29. Gemmel, F. et al. Prosthetic joint infections: radionuclide state-of-the-art imaging. *Eur. J. Nucl. Med. Mol. Imaging* **39**, 892–909 (2012).
30. Li, C. et al. The role of biopsy in diagnosing infection after hip and knee arthroplasty: a meta-analysis. *Arch. Orthop. Trauma Surg.* **143**, 1779–1792 (2023).
31. Fink, B. et al. The diagnostic value of routine preliminary biopsy in diagnosing late prosthetic joint infection after hip and knee arthroplasty. *Bone Joint J.* **102-B**, 329–335 (2020).
32. Marculescu, C. E. et al. Fungi, mycobacteria, zoonotic and other organisms in prosthetic joint infection. *Clin. Orthop. Relat. Res.* **451**, 64–72 (2006).
33. Uhel, F. et al. Mycobacterium tuberculosis prosthetic joint infections: a case series and literature review. *J. Infect.* **78**, 27–34 (2019).
34. Azad, M. A. et al. Comparison of the BioFire joint infection panel to 16S ribosomal RNA gene-based targeted metagenomic sequencing for testing synovial fluid from patients with knee arthroplasty failure. *J. Clin. Microbiol.* **60**, e0112622 (2022).
35. Hong, H. L. et al. Targeted versus shotgun metagenomic sequencing-based detection of microorganisms in sonicate fluid for periprosthetic joint infection diagnosis. *Clin. Infect. Dis.* **76**, e1456–e1462 (2023).
36. Flurin, L. et al. Clinical use of a 16S ribosomal RNA gene-based Sanger and/or next generation sequencing assay to test preoperative synovial fluid for periprosthetic joint infection diagnosis. *mBio* **13**, e0132222 (2022).
37. Hermsen, M. et al. Deep learning-based histopathologic assessment of kidney tissue. *J. Am. Soc. Nephrol.* **30**, 1968–1979 (2019).
38. Tao, Y. et al. A preliminary study on the application of deep learning methods based on convolutional network to the pathological diagnosis of PJI. *Arthroplasty* **4**, 49 (2022).
39. Xu, W., Fu, Y. L. & Zhu, D. ResNet and its application to medical image processing: research progress and challenges. *Comput. Methods Programs Biomed.* **240**, 107660 (2023).
40. Saied, M. et al. Efficient pulmonary nodules classification using radiomics and different artificial intelligence strategies. *Insights Imaging* **14**, 91 (2023).
41. Park, Y. J. et al. Performance evaluation of a deep learning system for differential diagnosis of lung cancer with conventional CT and FDG PET/CT using transfer learning and metadata. *Clin. Nucl. Med.* **46**, 635–640 (2021).
42. Li, J. et al. Predicting gastric cancer tumor mutational burden from histopathological images using multimodal deep learning. *Brief. Funct. Genom.* **23**, 228–238 (2024).
43. Khosravi, P. et al. Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. *EBioMedicine* **27**, 317–328 (2018).
44. Cheung, C. Y. et al. A deep learning model for detection of Alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study. *Lancet Digit. Health* **4**, e806–e815 (2022).
45. Kurt, Z. et al. Evaluation of EfficientNet models for COVID-19 detection using lung parenchyma. *Neural Comput. Appl.* **35**, 12121–12132 (2023).

46. Lee, J. et al. CaMeL-Net: centroid-aware metric learning for efficient multi-class cancer classification in pathology images. *Comput. Methods Programs Biomed.* **241**, 107749 (2023).
47. Shi, X. et al. Graph temporal ensembling based semi-supervised convolutional neural network with noisy labels for histopathology image analysis. *Med. Image Anal.* **60**, 101624 (2020).
48. Oquab, M. et al. DINOv2: learning robust visual features without supervision. Preprint at https://doi.org/10.48550/arXiv.2304.07193 (2023).

Acknowledgements

The authors would like to thank Moxuan Yang for helpful discussions. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Author contributions

Here are the names denoted by their initials: Ye Tao (Y.T.), Yazhi Luo (Y.Z.L.), Hanwen Hu (H.W.H.), Wei Wang (W.W.), Ying Zhao (Y.Z.), Shuhao Wang (S.H.W.), Qingyuan Zheng (Q.Y.Z.), Tianwei Zhang (T.W.Z.), Guoqiang Zhang (G.Q.Z.), Jie Li (J.L.), Ming Ni (M.N.). All authors have read and approved the manuscript. M.N., J.L., and S.H.W. conducted research design. H.W.H., Y.T., and Y.Z.L. conducted data collection and calculation. J.L., Y.T., and H.W.H. identified pathological data and constructed a training set. Y.L., W.W., Y.Z., and S.H.W. designed and trained an artificial intelligence model. Q.Y.Z., G.Q.Z., and T.W.Z. conducted data proofreading and statistical analysis. Y.T., H.W.H., and Y.Z.L. conducted article writing. M.N. and J.L. conducted article proofreading, and all authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Jie Li or Ming Ni.