Appendix 1

CT scanner and imaging parameters

At Luhe Hospital, the CT images were acquired in an axial orientation using a Philips Ingenuity Core 128 CT scanner at 120 kV, 300 mAs, and 1 mm thickness. The acquired images were reconstructed using a Philips Portal Workstation in sagittal view (3 mm thickness) in the digital imaging and communications in medicine (DICOM) format and downloaded to Hina MIIS-RIS PACS system. At Xuanwu Hospital, the CT images were acquired in axial orientation using a GE Revolution CT scanner at 120 kV, 200-500 mAs, and 0.625 mm thickness. The acquired images were reconstructed using a GE AW4.7 Workstation to sagittal view (2 mm thickness) in DICOM format and downloaded to a UniWeb Viewer PACS system.

We further used a Hina MIIS-RIS at Luhe Hospital and a UniWeb Viewer system at the Xuanwu Hospital, and we converted the CT slices from the DICOM format to the JPEG format. The format conversion was achieved by linear mapping of the values in the DICOM image to a minimum of 0 and a maximum of 255. Converting the DICOM image to a JPEG image led to a perceptible loss in image quality, but this quality loss had little effect on the ability of the surgeons to make a correct diagnosis.

Annotation of dataset

Surgeons implemented annotations on JPEG image of each CT slice using the LabelMe tool (version 5.0.1). The injured vertebrae annotation process consisted of three steps: independent annotation, consistency checking, and consulted co-annotation (*Figure S1*,*S2*). First, three spine surgeons implemented diagnoses and annotations for every CT slice independently. Second, we implemented consistency checking for all annotations. We compared the annotations on each CT slice from the three surgeons to find inconsistent annotations. Finally, for the inconsistent annotations, all three surgeons rechecked the patient's spinal CT images and corresponding MRI T1 and T2 images together to reach consensus.

Parameter settings for vertebra detection module

Vertebra detection module (VDModule) was developed to detect all vertebrae in teach CT slice. For the object detection task, the Bboxes are rectangular and are sensitive to random rotation operation. Rotation augmentation in an arbitrary degree would affect the precision of Bboxes and make the ground truth less reliable, thereby likely hurting model performance. We ensured the reliability of the Bboxes after augmentation by adopting several specific operations, including horizontal and vertical flipping, rotating 90 degrees, or a multiple of 90 degrees. We enlarged the image and Bbox datasets eight-fold using offline data augmentation (*Figure S3*).

Other parameter settings were set as follows: batch size 3; maximum epoch 5; learning rate 10^{-3}; and stochastic gradient descent with momentum (SGDM) optimizer. The DL architectures and experiments were implemented on a computer with MATLAB 2021a and configured with an Nvidia GeForce GTX 1080 Ti GPU with 11 GB of memory.

Calculation of Bboxes for normal category vertebrae

In the annotation process of the vertebra classification task, only the vertebrae with injury were annotated in each image. The remaining vertebrae were the "normal" category. We automatically calculated the Bboxes of the normal category vertebrae. First, the Bboxes of all vertebrae were determined using the developed VDModule for each CT slice. Second, we abandoned the detected Bboxes that largely overlapped with at least one annotated disease Bbox (IoM ≥0.5). The remaining annotated Bboxes that had little or no spatial overlap with the disease Bboxes were deemed normal category vertebrae.

Development of the vertebra extraction module

The vertebra extraction module (VEModule) was used to construct the vertebra image patch dataset for the vertebra classification task. Each Bbox of a CT slice underwent three steps in the image patch extraction process. First, the Bbox was

 https://dx.doi.org/10.21037/qims-23-685

augmented five-fold through translation and scaling. The horizontal and vertical translation range was [-10 to 10] pixels. The value of the scaling range was [0.9–1.1]. Second, the CT image was cropped using each Bbox and the image size was modified to 196×196 pixels using zero-padding and center-crop operations. Third, the image patch was rescaled to 224×224 pixels to fit the subsequent DL models. For the training dataset, all augmented image patches were used for training. For the testing dataset, the average diagnostic results of the augmented image patches were used as the final diagnostic results.

Parameter settings for vertebra classification module

We developed vertebra classification module (VCModule) to classify six vertebra categories. We used random over-sampling and under-sampling strategies to solve the class imbalance problem, and used data augmentation techniques to solve the overfitting problem.

In the training dataset, the samples for different vertebra diseases were highly imbalanced. We used random over-sampling and random under-sampling strategies to establish a more balanced training dataset for various vertebra categories. Over-sampling can lead to an overfitting problem, especially for classes with few original samples. Hence, we used data augmentation techniques to solve the overfitting problem. The data augmentation consisted of multiple image processing operations, including image rotation, image translation, noise addition, and brightness and contrast modification. Other parameter settings were as follows: maximum epoch, 15; batch size, 64; learning rate, 10-2; learning rate decays with a ratio of 0.2 every 5 epochs; and SGDM optimizer.

Table S1 Distribution of diseased thoracic and lumbar vertebrae

Vertebra location	Luhe Hospital cohort: patient count (percentage)	Xuanwu Hospital cohort: patient count (percentage)
T1	0 (0.0%)	0 (0.0%)
T2	0 (0.0%)	0 (0.0%)
T3	0 (0.0%)	0 (0.0%)
T4	1 (0.1%)	0 (0.0%)
T5	9 (0.9%)	0 (0.0%)
T6	34 (3.2%)	0 (0.0%)
T7	41 (3.9%)	1 (2.2%)
T8	49 (4.7%)	3 (6.5%)
T9	39 (3.7%)	2 (4.3%)
T10	39 (3.7%)	1 (2.2%)
T11	86 (8.2%)	3 (6.5%)
T12	282 (26.8%)	11 (23.9%)
L1	285 (27.1%)	14 (30.4%)
L2	148 (14.1%)	7 (15.2%)
L3	107 (10.2%)	11 (23.9%)
L4	84 (8.0%)	3 (6.5%)
L5	24 (2.3%)	1 (2.2%)

T, thoracic vertebra; L, lumbar vertebra.

 https://dx.doi.org/10.21037/qims-23-685

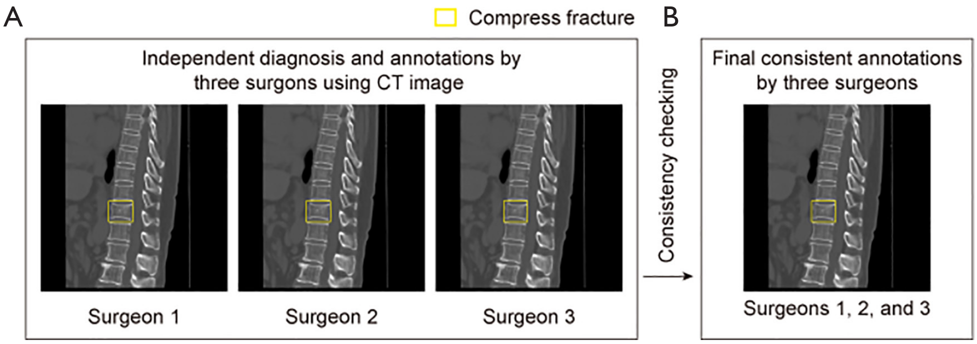

Figure S1 Annotation process for vertebrae with consistent initial diagnoses. (A) Consistent diagnosis and annotations on a computer tomography (CT) slice by three surgeons independently. (B) Final consistent annotations by the three surgeons.

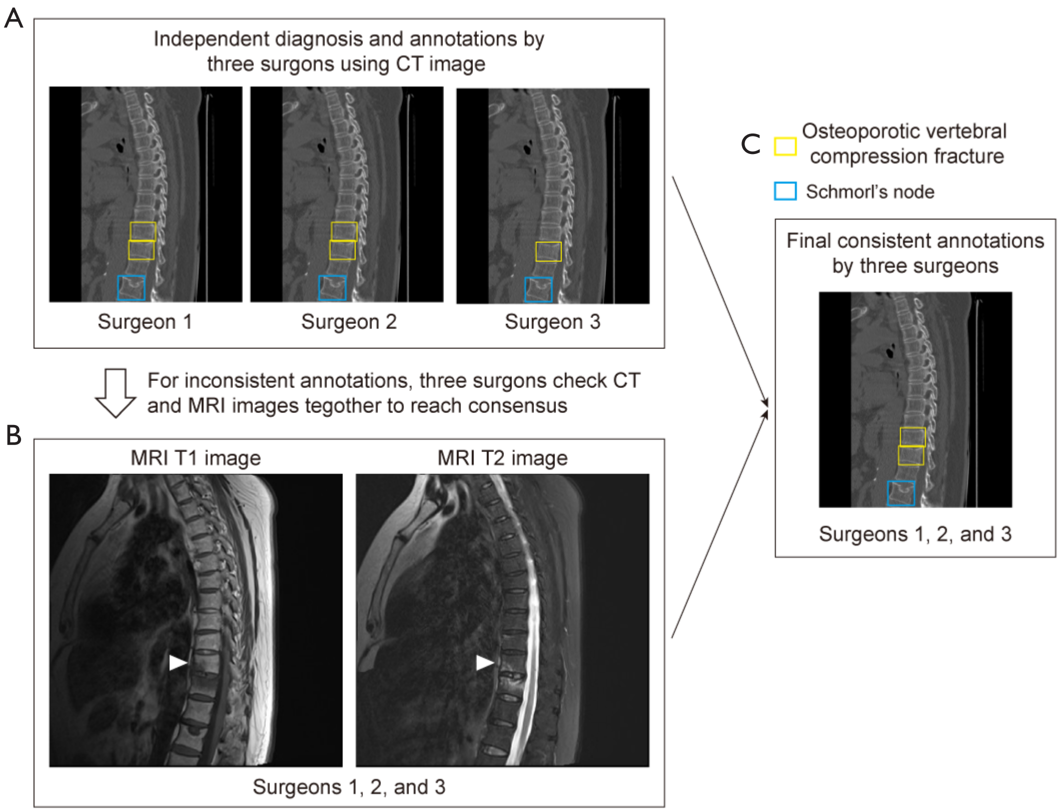

Figure S2 Annotation process for vertebrae with inconsistent initial diagnoses. (A) Inconsistent diagnosis and annotations on a computer tomography (CT) slice by three surgeons independently. All surgeons diagnosed thoracic vertebra T11 and lumbar vertebra L1 as osteoporotic vertebral compression fracture (OVCF) and Schmorl's node (SN), respectively. Surgeons 1 and 2 diagnosed thoracic vertebra T10 as OVCF, but surgeon 3 diagnosed T10 as Normal. (B) Consulted co-annotation by the three surgeons for inconsistent annotations based on CT images and the corresponding MRI T1 and T2 images. The arrows indicate that T10 has obvious OVCF characteristics on MRI T1 and T2 images. (C) Final consistent annotations by the three surgeons. T10, T11 and L1 were diagnosed as OVCF, OVCF, and SN, respectively.

 https://dx.doi.org/10.21037/qims-23-685

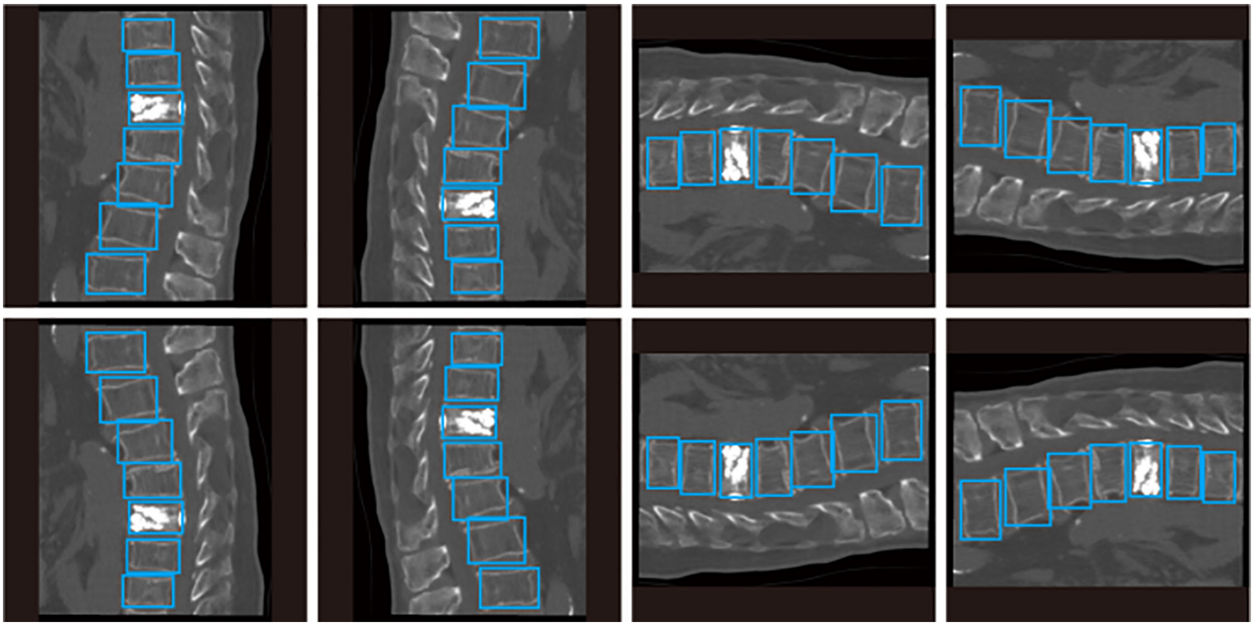

Figure S3 Eight types of augmentation operations used for both image and bounding boxes. Each blue rectangle represents one annotated or detected vertebra.

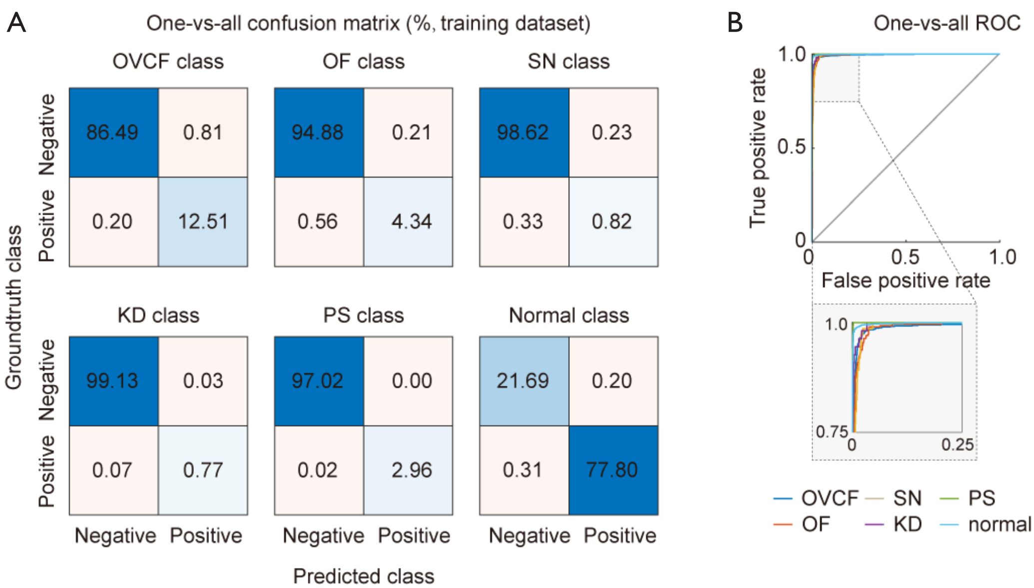

Figure S4 Performance of the DL-based vertebra diagnostic system in the training dataset. (A) One-*vs.*-all confusion matrix plot for ResNet50-based multi-output model. (B) One-*vs.*-all ROC curve. Insert plot, enlarged graph in the selected range.

https://dx.doi.org/10.21037/qims-23-685