



OPEN

## Rib fracture detection system based on deep learning

Liding Yao<sup>1,3</sup>, Xiaojun Guan<sup>1,3</sup>, Xiaowei Song<sup>1</sup>, Yanbin Tan<sup>1</sup>, Chun Wang<sup>2</sup>, Chaohui Jin<sup>2</sup>, Ming Chen<sup>2</sup>, Huogen Wang<sup>2</sup>✉ & Minming Zhang<sup>1</sup>✉

Rib fracture detection is time-consuming and demanding work for radiologists. This study aimed to introduce a novel rib fracture detection system based on deep learning which can help radiologists to diagnose rib fractures in chest computer tomography (CT) images conveniently and accurately. A total of 1707 patients were included in this study from a single center. We developed a novel rib fracture detection system on chest CT using a three-step algorithm. According to the examination time, 1507, 100 and 100 patients were allocated to the training set, the validation set and the testing set, respectively. Free Response ROC analysis was performed to evaluate the sensitivity and false positivity of the deep learning algorithm. Precision, recall, F1-score, negative predictive value (NPV) and detection and diagnosis were selected as evaluation metrics to compare the diagnostic efficiency of this system with radiologists. The radiologist-only study was used as a benchmark and the radiologist-model collaboration study was evaluated to assess the model's clinical applicability. A total of 50,170,399 blocks (fracture blocks, 91,574; normal blocks, 50,078,825) were labelled for training. The F1-score of the Rib Fracture Detection System was 0.890 and the precision, recall and NPV values were 0.869, 0.913 and 0.969, respectively. By interacting with this detection system, the F1-score of the junior and the experienced radiologists had improved from 0.796 to 0.925 and 0.889 to 0.970, respectively; the recall scores had increased from 0.693 to 0.920 and 0.853 to 0.972, respectively. On average, the diagnosis time of radiologist assisted with this detection system was reduced by 65.3 s. The constructed Rib Fracture Detection System has a comparable performance with the experienced radiologist and is readily available to automatically detect rib fracture in the clinical setting with high efficacy, which could reduce diagnosis time and radiologists' workload in the clinical practice.

Rib fractures are the most common consequences of traumatic chest injury<sup>1</sup>, most commonly caused by motor vehicle accidents, sport, falls etc. A recent study that included 4,168 patients with thoracic trauma in East China showed that 66.8% of the patients sustained rib fractures<sup>2</sup>. Rib fracture is also an important indicator of trauma severity, and patients with rib fractures have a higher admission rate and mortality than those without<sup>3</sup>. Many studies have highlighted that there is a high morbidity and mortality rate with even a single rib fracture, and as the number of rib fractures increases, patient's morbidity and mortality rates increase<sup>1,4</sup>. In addition, accurate detection of rib fracture can have medical-legal relevance<sup>5</sup>. Therefore, it is essential to accurately diagnose the location and the number of rib fractures in the clinical setting.

Plain X-ray and computed tomography (CT) are the most common imaging modalities used for rib fracture detection. In comparison to CT, the plain X-ray is convenient and fast, but the detection rate is relatively poor with more than 50% of rib fractures missed<sup>1,6</sup>. CT is the main imaging modality used to evaluate thoracic trauma for rib fractures and associated complications<sup>1</sup>. With chest CT, a large number of imaging sections and series are generated which consist of 12 pairs of ribs with heterogeneous shapes. All images must be evaluated sequentially, rib-by-rib and side-by-side, which is time-consuming and demanding<sup>5</sup>. Despite the best human effort, a misdiagnosis rate between 19.2 to 26.8% was reported with chest CT for rib fractures<sup>7,8</sup>, some of which may potentially lead to serious consequences<sup>9</sup>. Therefore, it is essential to develop an assistant machine learning detection system for rib fractures to minimize misdiagnosis.

Deep learning allows raw data to be fed into computer models and automatically processed with multiple pattern extraction and weighting levels<sup>10</sup>. Deep learning has been widely acknowledged for its great potential in complex pattern recognition and learning images with spatial hierarchy in multiple medical fields such as dermatology, radiology, ophthalmology, and pathology<sup>11</sup>, as well as in various fracture detections such as radial

<sup>1</sup>Department of Radiology, The Second Affiliated Hospital, Zhejiang University School of Medicine, No.88 Jiefang Road, Shangcheng District, Hangzhou 310009, Zhejiang, China. <sup>2</sup>Hithink RoyalFlush Information Network Co., Ltd, No. 18 Tongshun Street, Yuhang District, Hangzhou 310012, Zhejiang, China. <sup>3</sup>These authors contributed equally: Liding Yao and Xiaojun Guan. ✉email: wanghuogen@myhexin.com; zhangminming@zju.edu.cn

and ulnar fracture<sup>12</sup>, wrist fracture<sup>13</sup>, and thoracolumbar fracture etc.<sup>14</sup>. Therefore, it would also be feasible to construct a rib fracture detection system based on a deep learning model. Recently, Zhou et al.<sup>15</sup>, Weikert et al.<sup>16</sup> and Jin et al.<sup>17</sup> constructed several deep learning models for rib fracture detection with a high diagnostic sensitivity and specificity, in particular, such models have the advantage of dramatically reducing the diagnosis time. Since the performance of a deep learning model is highly dependent on the characteristics of the trained data, such as imaging resolution and fracture extent, the model's generalizability remains a constant issue. In general, existing deep learning models often require localized data input and refinement for validation before clinical translation.

From the deep learning perspective, rib fracture diagnosis is defined as an object detection problem, and a number of methods, for example, Faster-RCNN<sup>18</sup>, SSD<sup>19</sup>, YOLO-v3<sup>20</sup>, have been previously proposed. Because the rib fracture region is relatively small and imperceptible within many chest CT images, such detection system is more difficult compared to other fracture detection systems. Therefore, in order to construct a deep-learning model specific for rib fracture, this study had developed a three-step method to detect rib fracture in a patient population with mild to severe rib fractures: (1) a semantic segmentation model was trained and used to extract all bony features from chest CT; (2) a rib location model was trained to extract ribs and remove vertebrae, scapulae and sternum; (3) a classification model was trained to identify the fracture in the extracted ribs, and the performance was further tested using an independent dataset. Finally, the performance of radiologist-model collaboration was evaluated after the completion of deep-learning model.

## Materials and methods

**Data collecting.** This study was approved by the Ethics Committee of the Second Affiliated Hospital of Zhejiang University with the informed consent waived. All the methods were carried out in accordance with relevant guidelines and regulations. A total of 1707 patients from the Second Affiliated Hospital of Zhejiang University were enrolled in this study. All chest CT scans were collected using keyword searches in Picture Archiving and Communication Systems (PACS) between January 1, 2016 and March 31, 2019. The inclusion criteria were: (a) slice thickness after image reconstruction less than 2 mm, (b) at least 1 rib fracture present, (c) no significant artifact in the images. Finally, a total of 1707 patients with chest CT images were included in this study. To ensure that all data was analyzed at the same windowing level, the window center and window width parameters of all CT scans were set to 400 and 1600, respectively.

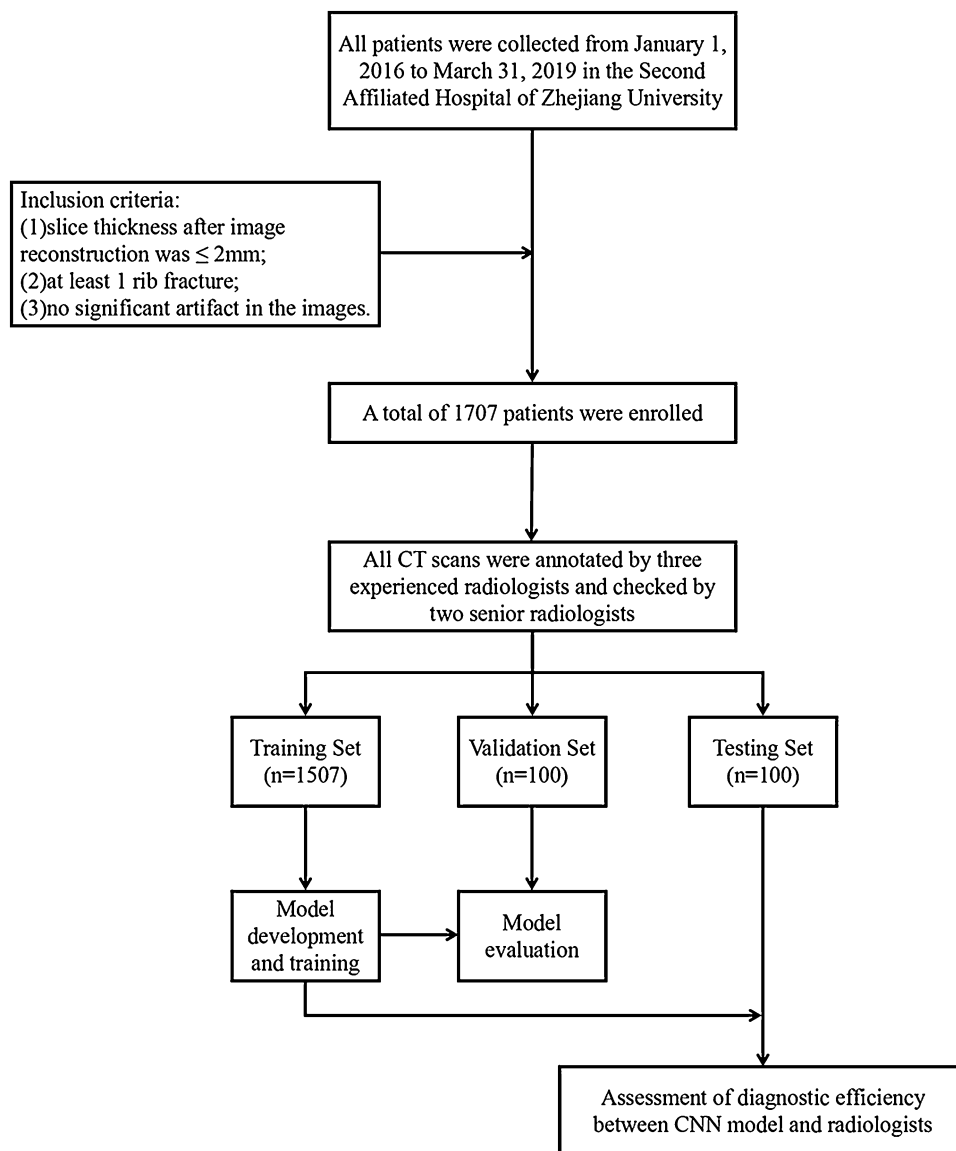
**Data annotation and pretreatment.** All CT images were annotated by three experienced radiologists (all with over 10 years of experience in CT diagnosis) and checked by two senior radiologists (both with more than 15 years of experience in CT diagnosis) as the ground truth for both the training, validation and testing sets of rib fracture detection model. In the situation where the annotation result of data was inconsistent, all five radiologists were invited to participate in a discussion, and a consensus decision was made. All CT images were annotated on a medical image processing and navigation software, 3D Slicer<sup>21</sup>, by drawing a mask around the bone region and a rectangular bounding box surrounding the rib fracture.

As shown in Fig. 1, patients were allocated into the training set (1507 cases, 581,701 slices, and 7362 fractures), the validation set (100 cases, 36,697 slices, and 473 fractures) and the testing set (100 cases, 37,183 slices, and 436 fractures) according to the examination time. The deep learning model for rib fracture detection was developed and trained on the training set. The hyper-parameters of the deep learning system were finetuned using the validation set. The diagnostic efficiency of the deep learning model was assessed on the testing dataset, together with the evaluation of the efficacy of the deep-learning model alone, radiologist alone and radiologist-model collaboration. These 100 cases for testing were composed of 62 male and 38 female subjects with a mean age of  $56.7 \pm 12.8$  years (range from 23–88 years). The overview of dataset was described in Table 1.

**Model construction.** The extents of rib fractures were highly varied and sometimes only a few abnormal pixels can be observed in very small fractures, which make them more difficult to detect. As illustrated in Fig. 2, we divided the fracture detection model into three-step algorithms, including bone segmentation, rib location, and rib fracture classification.

- (a) Bone segmentation: U-Net<sup>22</sup> was implemented and trained for bone segmentation. U-Net is an encoder-decoder architecture. The encoder was a typical Convolutional Neural Network (CNN) applied to extract semantic features from basic patterns to complex spatial patterns. In the decoder, a feature map resolution was gradually restored through a series of transposed convolution operations. Features from the encoder were reprocessed through feature cascade from levels of the encoder and decoder.
- (b) Rib location: Because ribs were difficult to distinguish from scapulae, sternum, and vertebrae when performing bone segmentation, we removed vertebrae and scapulae based on their characteristic shapes and locations as a critical preprocessing step.
- (c) Rib fracture classification: 3D DenseNet was employed for rib fracture classification, which was an extension of DenseNet<sup>23</sup> by extending 2D Convolution and 2D pooling into 3D. It connected each layer to every other layer in a feed-forward fashion, which alleviated the problem of gradient dissipation, facilitated the propagation and the reuse of features. An inception structure<sup>24</sup> was introduced to make the network learn features in different scales of receptive fields. Focal loss function<sup>25</sup> was calculated as the loss function during training, which enhanced the generalization ability of the model by minimizing the difference between the ground truth and training outcome, especially on hard cases.

The details of the procedures mentioned above were described in the Supplementary Materials.



**Figure 1.** Flow chart showing overall study process.

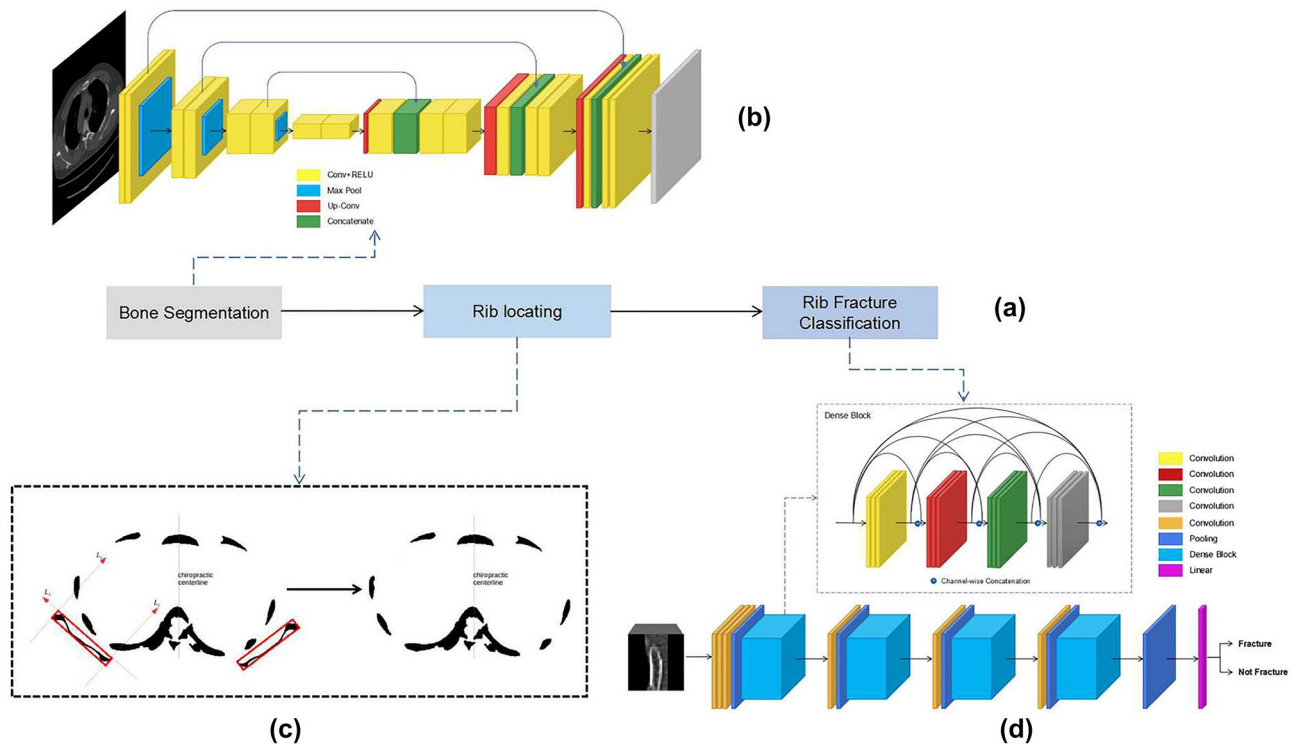
Cohorts	No. Patients	No. CT slices	No. Fractures
Training	1507	581,701	7362
Validation	100	36,697	473
Testing	100	37,183	436

**Table 1.** The overview of dataset.

### Model training.

- (a) The training of U-Net: U-Net was trained on a training set consisting of 4496 rib CT images. An input image of size  $512 \times 512$  was randomly cropped from a CT image. During training, each batch contained 8 samples. Data augmentation, such as flipping, contrast adjustment and random noise, was manipulated to avoid overfitting<sup>26</sup>. The optimization objective function was computed by the dice-loss<sup>27</sup> function, and the calculation formula of the dice loss function was:

$$L_{\text{dice}} = 1 - \frac{2 \sum_{i=1}^N P_i * g_i + \epsilon}{\sum_{i=1}^N P_i + \sum_{i=1}^N g_i + \epsilon}$$



**Figure 2.** The pipeline for detecting rib fractures from CT scans. The fracture detection task was divided into three stages, including bone segmentation, rib location, and rib fracture classification (a). (b) bone segmentation; (c) rib location; (d) rib fracture classification.

where  $N$  was the number of pixels in an input image,  $p_i$  was the predicted value, and  $g_i$  was the ground truth. A small constant  $\epsilon$  was also inserted to avoid the situation where the denominator was zero. The U-Net was trained using Adam optimizer with a warm-up strategy. The learning rate was set to 0.01 and then linearly decreased to 0.001 in 100 epochs.

- (b) The training of 3D DenseNet: A total of 91,574 fracture blocks and 50,078,825 normal blocks were extracted from the training set. One case contained a series of images which could be regarded as  $H \times W \times D$  volume data. Let  $h$  and  $w$  be the height and width of minimum circumscribed rectangles of all ribs on each image in the series, and a  $(h + 6) \times (w + 6) \times 5$  3D data block was sampled, and then the block was normalized into  $48 \times 48 \times 5$  in non-deformation way. To normalize the 3D data block in non-deformation way, a minimum circumscribed square of each 3D data block was firstly drawn. Then, the margin in the square was padded with pure black pixels to generate a new 3D data block. Finally, the new 3D data block was resized into  $48 \times 48 \times 5$ . During training, we adopted a sampling strategy to alleviate the imbalance between positive and negative samples by allowing each batch to contain 4 positive samples and 4 negative samples. Data augmentation, such as flipping, contrast adjustment and random noise, was applied to avoid overfitting. The optimization objective function was computed by focal loss function. The 3D DenseNet was further trained again using Adam optimizer with a warm-up strategy and the learning rate was initially set to 0.01 then decreased to 0.001 in 100 epochs.

### Model evaluation and statistical analysis.

- (a) Bone Segmentation: for bone segmentation model, the average IOU (intersection over union) and the Dice coefficient<sup>28</sup> were used for the evaluation of U-Net. IOU is a good metric for measuring overlap between two masks, and it can be defined as follows:

$$\text{IOU} = \frac{|A \cap B|}{|A \cup B|}$$

The Dice Similarity Coefficient (DSC) is a statistic used to evaluate the similarity of two masks. The DSC is defined as

$$\text{DSC} = \frac{2|A \cap B|}{|A| + |B|}$$

Cohorts	U-Net	3D DenseNet	
	No. CT images	No. Fracture blocks	No. Normal blocks
Training	4496	91,574	50,078,825
Validation	3145	5981	3,323,151
Testing	3568	5992	3,452,162

**Table 2.** The overview of dataset for the training of U-Net and 3D DenseNet.

where A is a set of rib regions annotated by the radiologists, and B is a set of segmented rib regions by the algorithm.

- (b) Rib fracture classification: to compare the efficiency of the trained deep learning model with that of radiologists in diagnosing rib fracture, we proposed three testing groups in our study, including the deep learning model alone, radiologist alone, and radiologist-model collaboration. Two radiologists (radiologist A with 3 years of experience in CT diagnosis; radiologist B with 8 years of experience in CT diagnosis) who were blind to patient information and imaging annotation were invited to participate in the study. Both radiologists were informed of the gold standard criteria for rib fracture classification before the test, and their detected fracture location and diagnosis time were recorded. After three months, both radiologists were invited to diagnose the same set of chest CT images with the assistance of the constructed detection model and same parameters were taken. Two sample t-test was calculated to compare the detection or diagnosis time required for deep learning model alone, radiologist alone, and radiologist-model collaboration.

Precision (positive predictive value, PPV)<sup>29</sup>, recall<sup>29</sup>, F1-score<sup>30</sup> and negative predictive value (NPV)<sup>29</sup> were selected as evaluation metrics for three testing groups. Precision was defined as the ratio of correctly predicted positive observations to the total predicted positive observations. Recall was the ratio of correctly predicted positive observations to the all observations in actual class. F1-score was the weighted average of Precision and Recall. NPV was the ratio of correctly predicted negative observations to the total predicted negative observations. These metrics can be formulated as follows:

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$\text{F1-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{NPV} = \frac{TN}{TN + FN}$$

where “TP” were these ribs classified by the algorithm as rib fractures and also by the radiologists, “FP” were these ribs classified by the algorithm as rib fractures but not by the radiologists, “TN” were corresponding to these ribs classified as not belonging to rib fractures both by the algorithm and the radiologists, and “FN” were these ribs classified as not belonging to rib fractures but they were corresponding to rib fractures bones according to the annotation of the radiologists.

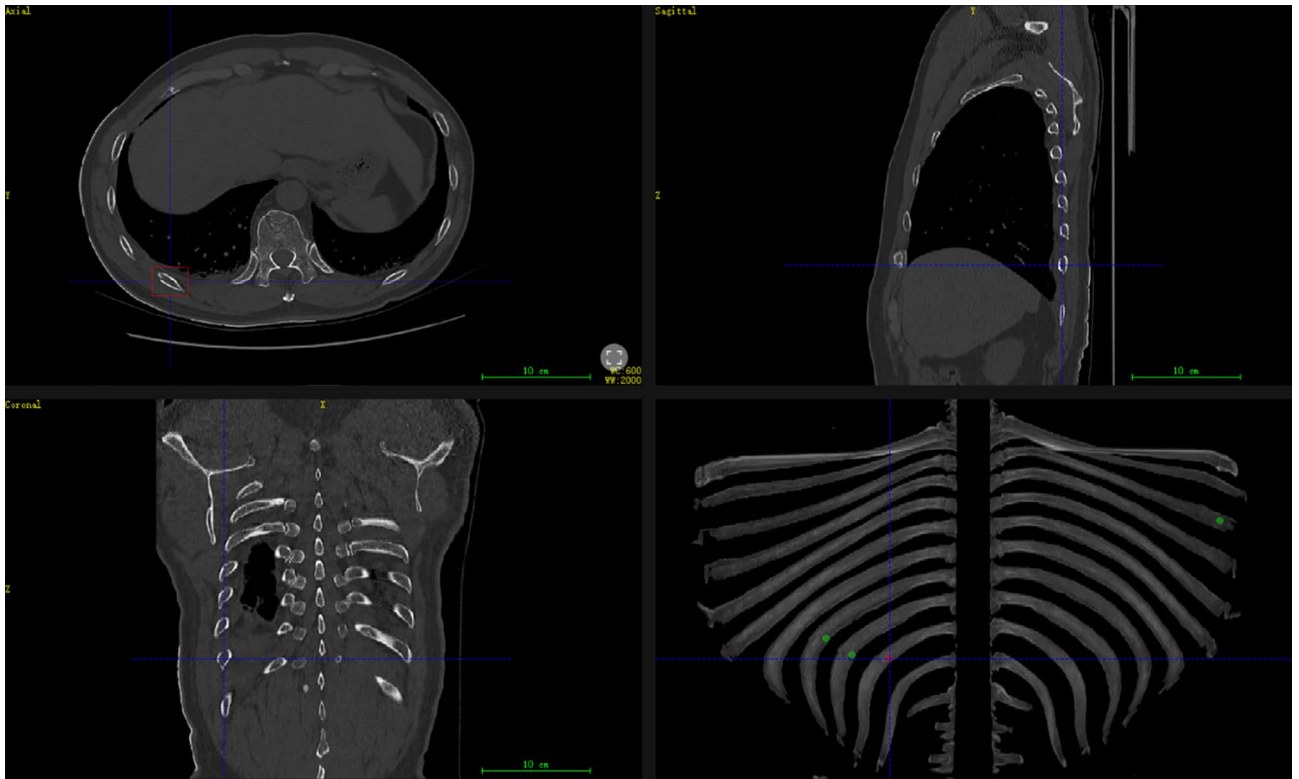
Finally, the sensitivity and average number of the false positives (FPs) per patient were analyzed using a free-response ROC (FROC) curve<sup>31</sup>, because FROC was a tool for characterizing the performance at all thresholds simultaneously. The diagnosis and detection time were recorded.

## Results

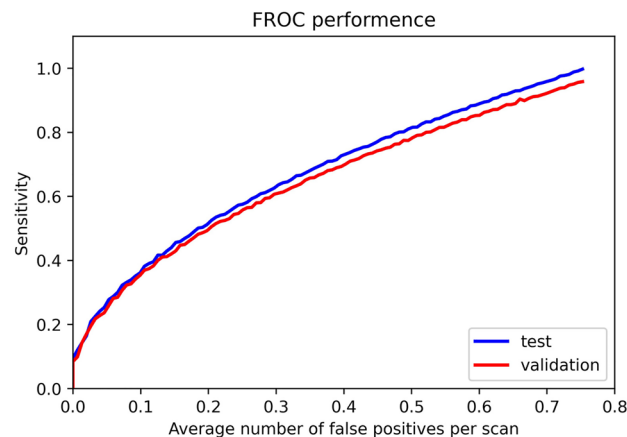
Due to the heavy workload of bone annotation, only 4496, 3145 and 3568 CT images were annotated for the training, validation and testing of U-Net in bone segmentation, respectively. The average IOU in the test dataset was 0.8462, and the DSC was 0.9167. It was worth mentioning that the edge of a bone was ambiguous and difficult to label accurately, so it was very hard to objectively evaluate the performance of bone segmentation. And the bone segmentation with elongated shape tended to be associated with low segmentation metrics. Despite the low IOU, U-Net could effectively remove the factors that interfere with the classification of rib fractures.

For the training of rib fracture classification, we recreated a new dataset. The overview of this dataset was shown in Table 2. For the testing of the Rib Fracture Detection System, a total of 100 cases were examined, including 436 rib fractures, with an average of 4.36 rib fractures in one case (Fractured ribs varied from 1 to 15). The result of rib fracture detection from Rib Fracture Detection System was represented in 3D view as shown in Fig. 3.

To evaluate the performance of the Rib Fracture Detection System, the results of the validation set and the testing set were both reported. In the validation set, the F1-score of the Rib Fracture Detection System was 0.888 and the precision and recall were 0.864 and 0.914 respectively in the 100 cases tested with an FROC cutoff value of 0.89 (FROC curve was shown in Fig. 4). On the testing set, the F1-score of the Rib Fracture Detection System was 0.890 and the precision, recall and NPV were 0.869, 0.913 and 0.969 respectively in the 100 cases tested with the same cutoff value of the FROC threshold of 0.89.



**Figure 3.** Rib 3D view. The red rectangular box was the selected fracture lesion, and the green parts were the other suspected fracture lesions detected by our Rib Fracture Detection System.



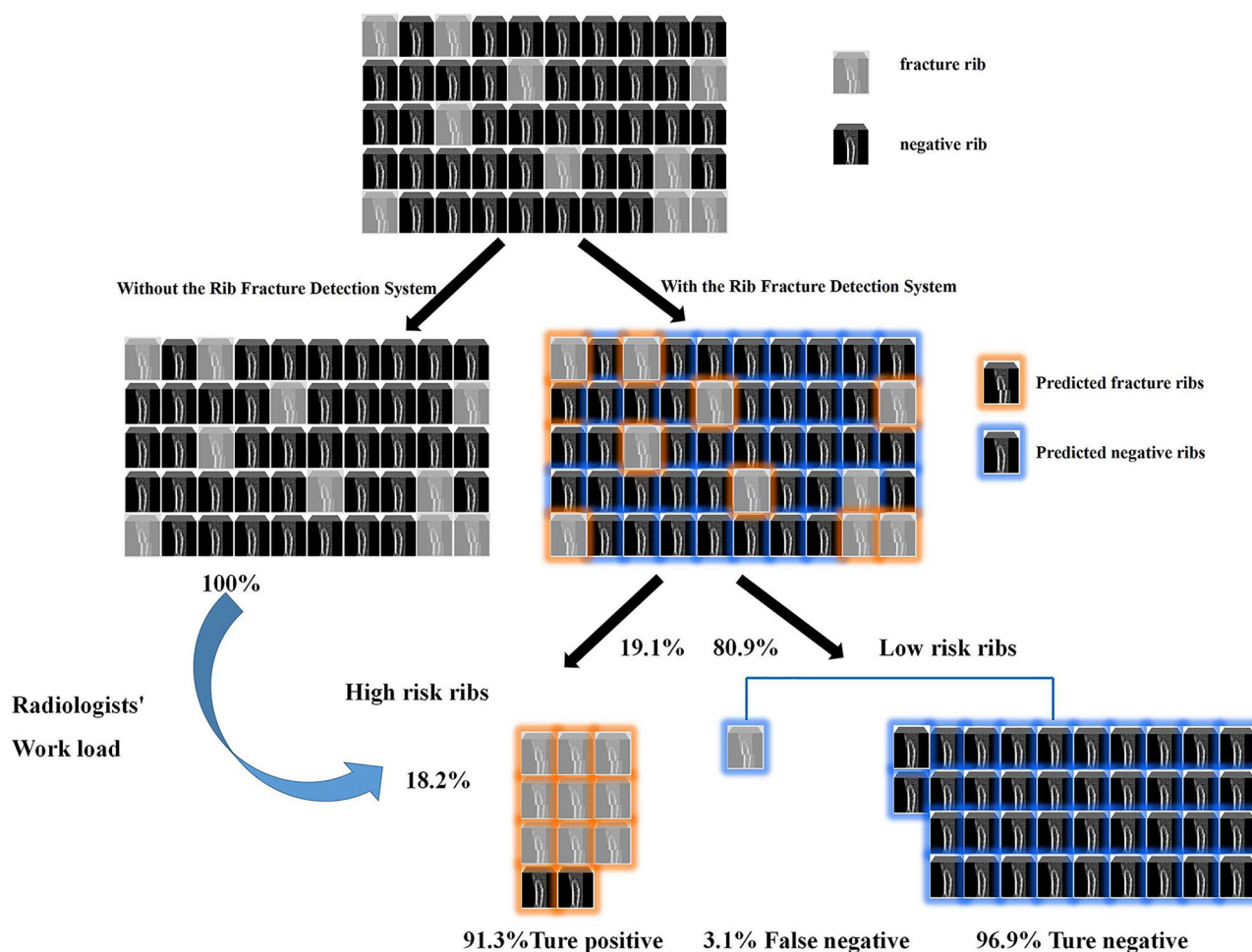
**Figure 4.** Free-response ROC (FROC) curve for our model.

The F1-score of radiologist A was 0.796 and the precision, recall and NPV were 0.935, 0.693 and 0.989, respectively. The F1-score of radiologist B was 0.889 and the precision, recall and NPV were 0.928, 0.853 and 0.985, respectively. Therefore, it was concluded that the Rib Fracture Detection System had a comparable performance to radiologists, but with higher recall. In addition, we tested the radiologist-model collaboration performance, and observed that, for both radiologists, their diagnostic performance were improved and their workload (diagnosis time) was significantly reduced (Table 3).

This Rib Fracture Detection System detected 458 suspected fractures and 398 true fractured ribs from these cases. An average of 4.58 suspected fractures and 3.98 ground truth annotated were detected in an average of 372 images in one case with our model, which meant if applying to the clinical setting, a radiologist only needs to review a small number of images containing suspected fractures using this model. The results demonstrated that this model had a NPV of 0.969. This emphasized the possibility of model triage, as a total of 80.9% of images were predicted as negative with 96.9% true-negatives, this left the other 19.1% images as high-risk images with a potential rib fracture. Therefore, radiologists could focus on these rib images to improve accuracy and workflow

Group	Model	Radiologist A	Radiologist B	Radiologist A-model collaboration	Radiologist B-model collaboration
F1-score	0.890	0.796	0.889	0.925	0.970
Recall	0.913	0.693	0.853	0.920	0.972
Precision	0.869	0.935	0.928	0.930	0.968
NPV	0.969	0.989	0.985	0.985	0.993
Time (seconds)	20 ± 5.8	242.6 ± 83.0*	153.6 ± 34.2*	207.0 ± 47.9 <sup>a</sup>	58.6 ± 31.4 <sup>a</sup>

**Table 3.** The comparison of the performance between the model, radiologist and radiologist-model collaboration. *NPV* negative predictive value. \*Indicated the p value of the comparison between model and radiologists was < 0.001. <sup>a</sup>Indicated the p value of the comparison before and after using the model was < 0.001.



**Figure 5.** Impact of the Rib Fracture Detection System in clinical practice for patients with suspicion of rib fracture in the Department of Radiology. In a cohort of patients with suspected rib fractures who underwent chest CT investigation, radiologists should pay close attention to all of the ribs without the help of our model in order to look for 18.2% of the fractured ribs. Since 80.9% of the ribs were diagnosed as non-fracture ribs by this model with a 96.9% true-negative rate, it demonstrated high accuracy in identifying true non-fractured ribs by the constructed model. As a result, with the assistance of this deep learning model, radiologists only had to pay more attention to 19.1% of the ribs that were categorized as high-risk for fracture, which significantly reduced their workload in detecting rib fracture.

and finally reduce workload. A schematic diagram was shown in Fig. 5 to illustrate how the Rib Fracture Detection System worked and reduced the workload.

Group	Model	Fast RCNN	Faster RCNN	YOLOv3
F1-score	0.890	0.863	0.870	0.877
Recall	0.913	0.874	0.889	0.894
Precision	0.869	0.853	0.852	0.861
NPV	0.969	0.925	0.932	0.942

**Table 4.** The comparison of the performance between our Rib Fracture Detection System with Fast RCNN, Faster RCNN, YOLOv3.

## Discussion

This study constructed a deep-learning Rib Fracture Detection System and its performance was assessed in comparison to radiologists. It was concluded that this deep-learning model had the ability to detect rib fractures with high precision and recall (precision: 0.869, recall: 0.913), which had a comparable diagnostic precision with both the junior and the experienced radiologists and was superior in term of recall ability. This indicated that this detection model would not only reduce the workload but also minimize misdiagnosis. Moreover, for both radiologists in the study, with the assistance of this detection model, their diagnostic effectiveness was significantly improved with decreased diagnosis time and increased recall.

Rib fracture detection is crucial for thoracic trauma evaluation in identifying associated injury, preventing complication, obviating potential medico legal issues, and helping patient pain management<sup>32</sup>. Thin-slice CT is the main imaging modality used to evaluate thoracic trauma, generating hundreds of images waiting to be analyzed. Consequently, radiologists have a high workload, and missed diagnosis becomes an inevitable issue, which is more apparent for junior radiologists<sup>13</sup>. To address this clinical challenge, this study suggested a Rib Fracture Detection System based on the deep learning model, which allowed to have superior recall and similar diagnostic precision to the radiologists. By using this detection system, diagnostic performance was significantly improved for both the experienced and the junior radiologists, and the workload was dramatically reduced, suggesting that such a detection system had a high potential for clinical translation.

Although several studies had recently reported the usefulness of deep-learning models in detecting rib fractures, each deep learning model required clinically suitable larger local data input with a balanced sample of mild to severe rib fractures. In some difficult mild fractures, only a few abnormal pixels could be observed, which could be easily missed. Therefore, in the present study, we performed a three-step model construction, including bone segmentation, rib location, and rib fracture classification. For rib fracture classification, 3D DenseNet was utilized as it was sensitive to even small fracture lesions with only a few abnormal pixels<sup>23</sup>. For rib location, the spine and scapulae removal could help eliminate the interference of the spine and scapulae. Based on our evaluation, 87 blocks in spine and scapulae were recognized as rib fractures in the rib fracture classification without the spine and scapulae removal. The sampling strategy and focal loss function were further applied to alleviate the imbalance between positive and negative samples. Focal loss is an improved version of Cross-Entropy Loss that tries to handle the imbalance between positive and negative samples by assigning more weights to hard examples. In addition, we adapted an annotation-checking-discussion workflow to ensure each sample was read by five radiologists and high accuracy of the ground truth. With these characteristics, this Rib Fracture Detection System achieved high precision and recall in rib fracture detection and improved radiologist work efficiency.

To further demonstrate the effectiveness of our proposed algorithm, we compared our algorithm with several state-of-the-art algorithms. We evaluated the performance of FracNet<sup>17</sup>, Fast RCNN<sup>16</sup>, Faster RCNN<sup>15</sup>, YOLOv3<sup>15</sup> on the testing set of our collected dataset. The comparison between our Rib Fracture Detection System with these approaches was listed in Table 4. These results demonstrated that the performance of our Rib Fracture Detection System outperformed FracNet, Fast RCNN, Faster RCNN, and YOLOv3.

Although the system performed well in most cases, this study still had several limitations. First, the deep learning model was developed and trained on CT data from a single large academic institution and the test set is relatively small, lacking multi-center or external data validation. Future research is required to determine if the same model trained can achieve high performance on larger or multi-institutional datasets. Second, this was a retrospective study and we would hope to include prospective data in the future with more radiologists as well as more chest CT cases. Third, without accurate masks for rib segmentation in the dataset, the quantitative evaluation of the precision of the spine and scapulae removal was not performed. Fourth, false positive rate is still high compared with the performance of radiologists and we hope we can reduce the false positive rate by collecting more dataset and introducing novel deep learning methods. And finally, this model could only detect rib fracture but cannot identify fracture degree and classify acute and healed rib fractures, which may be useful information to clinicians. Therefore, future studies were warranted to complete these tasks with this Rib Fracture Detection System.

## Conclusions

This study developed a Rib Fracture Detection System that achieved high performance in rib fracture detection on chest CT images based on the deep learning algorithm. By radiologist-model collaboration, radiologists can significantly reduce their workload, and minimize misdiagnosis. This Rib Fracture Detection System is readily available in the clinical setting.



## Data availability

Anonymized data can be made available upon reasonable request to the corresponding author.

Received: 14 May 2021; Accepted: 25 November 2021

Published online: 06 December 2021

## References

1. Talbot, B. S. *et al.* Traumatic rib injury: Patterns, imaging pitfalls, complications, and treatment. *Radiographics* **37**(2), 628–651 (2017).
2. Zhang, S. *et al.* Thoracic trauma: A descriptive review of 4168 consecutive cases in East China. *Medicine (Baltimore)* **98**(14), e14993 (2019).
3. Murphy, C. E. *et al.* Rib fracture diagnosis in the panscan era. *Ann. Emerg. Med.* **70**(6), 904–909 (2017).
4. Haines, K. L. *et al.* Rib fracture location should be evaluated when predicting morbidity and mortality in trauma patients. *Am. Surg.* **84**(9), 1462–1465 (2018).
5. Ringl, H. *et al.* The ribs unfolded: A CT visualization algorithm for fast detection of rib fractures: effect on sensitivity and specificity in trauma patients. *Eur. Radiol.* **25**(7), 1865–1874 (2015).
6. Traub, M. *et al.* The use of chest computed tomography versus chest X-ray in patients with major blunt trauma. *Injury Int. J. Care Injured.* **38**(1), 43–47 (2007).
7. Cho, S. H., Sung, Y. M. & Kim, M. S. Missed rib fractures on evaluation of initial chest CT for trauma patients: Pattern analysis and diagnostic value of coronal multiplanar reconstruction images with multidetector row CT. *Br. J. Radiol.* **85**(1018), e845–e850 (2012).
8. Banaste, N. *et al.* Whole-body CT in patients with multiple traumas: Factors leading to missed injury. *Radiology* **289**(2), 374–383 (2018).
9. Pinto, A. *et al.* Errors in imaging patients in the emergency setting. *Br. J. Radiol.* **89**(1061), 20150914 (2016).
10. Lecun, Y. *et al.* Deep learning. *Nature* **521**(7553), 436–444 (2015).
11. Esteva, A. *et al.* A guide to deep learning in healthcare. *Nat. Med.* **25**(1), 24–29 (2019).
12. Thian, Y. L. *et al.* Convolutional neural networks for automated fracture detection and localization on wrist radiographs. *Radiol. Artif. Intell.* **1**(1), e180001 (2019).
13. Lindsey, R. V. *et al.* Deep neural network improves fracture detection by clinicians. *Proc. Natl. Acad. Sci. U.S.A.* **115**(45), 11591–11596 (2018).
14. Raghavendra, U. *et al.* Automated system for the detection of thoracolumbar fractures using a CNN architecture. *Futur. Gener. Comput. Syst.* **85**, 184–189 (2018).
15. Zhou, Q. Q. *et al.* Automatic detection and classification of rib fractures on thoracic CT using convolutional neural network: Accuracy and feasibility. *Korean J. Radiol.* **21**(7), 869–879 (2020).
16. Weikert, T. *et al.* Assessment of a deep learning algorithm for the detection of rib fractures on whole-body trauma computed tomography. *Korean J. Radiol.* **21**(7), 891–899 (2020).
17. Jin, L. *et al.* Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet. *EBioMedicine* **62**, 103106 (2020).
18. Ren, S. *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(6), 1137–1149 (2017).
19. Wei, L., Dragomir, A., Dumitru, E. *et al.* SSD: Single shot MultiBox detector. In *European Conference on Computer Vision*. 21–37 (2016).
20. Redmon, J., Farhadi, A. YOLOv3: An incremental improvement. arXiv: Computer Vision and Pattern Recognition (2018).
21. Bruns, N. 3D slicer: Universelle 3D-visualisierung-software. *Unfallchirurg* **122**(8), 662–663 (2019).
22. Ronneberger, O., Fischer, P., Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 234–241 (2015).
23. Huang, G., Liu, Z., Van Der Maaten, L. *et al.* Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4700–4708 (2017).
24. Szegedy, C., Ioffe, S., Vanhoucke, V. *et al.* Inception-v4, inception-ResNet and the impact of residual connections on learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. 4278–4284 (2016).
25. Lin, T. Y., Goyal, P., Girshick, R. *et al.* Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*. 2980–2988 (2017).
26. Shorten, C. & Khoshgofaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**(1), 1–48 (2019).
27. Sudre, C. H., Li, W., Vercauteren, T. *et al.* Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. 240–248 (2017).
28. Milletari, F., Navab, N., Ahmadi, S. A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*. 565–571 (2016).
29. Goutte, C., Gaussier, E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European Conference on Information Retrieval*. 345–359 (2005).
30. Chicco, D. & Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **21**(1), 1–3 (2020).
31. Bandos, A. I., Rockette, H. E., Song, T. & Gur, D. Area under the free-response ROC curve (FROC) and a related summary index. *Biometrics* **65**(1), 247–256 (2009).
32. Kim, J. *et al.* Quantitative measurement method for possible rib fractures in chest radiographs. *Healthc Inform Res.* **19**(3), 196–204 (2013).

## Acknowledgements

The authors would like to thank Dr. Runlin Yang, from University of Melbourne at Melbourne, for her intellectual contributions of editing the grammar issues. We also would like to appreciate the Department of Intelligent Medicine at Hithink RoyalFlush Information Network Co. Ltd for technical support and statistical analysis. This work was supported by the Medical Health Science, the Technology Project of Zhejiang Provincial Health Commission (Grant Nos. 2017198897 and 2017205359).

## Author contributions

Y.L.D. designed the study, collected clinical data, performed statistical analyses, and wrote this manuscript; G.X.J. performed statistical analyses and revised the manuscript; T.Y.B. and S.X.W. collected clinical data and the chest CT data; W.C., J.C.H. and C.M. designed the study and performed statistical analyses. Z.M.M. and W.H.G.

designed the study, guided the implementation of the study and revised the manuscript. All authors reviewed the manuscript and approved its content.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-03002-7>.

**Correspondence** and requests for materials should be addressed to H.W. or M.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021