Development and evaluation of a deep learning-based model for simultaneous detection and localization of rib and clavicle fractures in trauma patients' chest radiographs

Chi-Tung Cheng [1,2] Ling-Wei Kuo [1,2] Chun-Hsiang Ouyang,[1,2] Chi-Po Hsu,[1,2] Wei-Cheng Lin,[3] Chih-Yuan Fu,[1,2] Shih-Ching Kang,[1,2] Chien-Hung Liao [1,2]

[1]Department of Trauma and Emergency Surgery, Chang Gung Memorial Hospital Linkou, Taoyuan, Taiwan
[2]Department of medicine, Chang Gung university, Taoyuan, Taiwan
[3]Department of Electrical Engineering, Chang Gung University, Taoyuan, Taiwan

Correspondence to
Dr Chien-Hung Liao; Surgymet@gmail.com

ABSTRACT

Purpose To develop a rib and clavicle fracture detection model for chest radiographs in trauma patients using a deep learning (DL) algorithm.

Materials and methods We retrospectively collected 56 145 chest X-rays (CXRs) from trauma patients in a trauma center between August 2008 and December 2016. A rib/clavicle fracture detection DL algorithm was trained using this data set with 991 (1.8%) images labeled by experts with fracture site locations. The algorithm was tested on independently collected 300 CXRs in 2017. An external test set was also collected from hospitalized trauma patients in a regional hospital for evaluation. The receiver operating characteristic curve with area under the curve (AUC), accuracy, sensitivity, specificity, precision, and negative predictive value of the model on each test set was evaluated. The prediction probability on the images was visualized as heatmaps.

Results The trained DL model achieved an AUC of 0.912 (95% CI 87.8 to 94.7) on the independent test set. The accuracy, sensitivity, and specificity on the given cut-off value are 83.7, 86.8, and 80.4, respectively. On the external test set, the model had a sensitivity of 88.0 and an accuracy of 72.5. While the model exhibited a slight decrease in accuracy on the external test set, it maintained its sensitivity in detecting fractures.

Conclusion The algorithm detects rib and clavicle fractures concomitantly in the CXR of trauma patients with high accuracy in locating lesions through heatmap visualization.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Rib and clavicle fractures frequently occur as part of thoracic injuries, and their accurate diagnosis is often challenging when relying solely on chest plain film radiography.

⇒ Although deep learning (DL) has demonstrated significant breakthroughs in clinical assistance, its application within the trauma field remains relatively limited.

WHAT THIS STUDY ADDS

⇒ This study introduces chest X-ray-FrNET, a DL algorithm that demonstrates a high sensitivity and acceptable accuracy in the detection of rib and clavicle fractures.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The development of a versatile and multitask DL algorithm has the potential to greatly influence clinical practice by providing valuable support to physicians in the management of patients with rib and clavicle fractures.

⇒ This innovation may lead to improvements in patient care, diagnostic accuracy, and overall outcomes in the field of fracture assessment.

INTRODUCTION

Traumatic rib fractures are the most frequently observed injuries resulting from widespread thoracic trauma.[1] These fractures hold clinical significance as they are associated with considerable pulmonary morbidity, mortality, and may result in long-term disabilities.[2 3] Rapid assessment and management are crucial to patient outcomes.[4] Historically, therapeutic options for rib fractures have been limited to conservative management, including analgesia, pulmonary hygiene, oxygen delivery, and allowing time for wound healing.[5] Unlike rib fractures, clavicular fractures are also common insults after chest trauma. Misdiagnosis is not common, but when it occurs, it can lead to long-term limitations in range of motion and decreased quality of life.[6 7] Providing high-quality clinical care and effective trauma treatment relies not only on the expertise of physicians but also on the valuable insights obtained from various imaging modalities. Prompt and accurate diagnosis, coupled with appropriate management, is critical for the survival of trauma patients. Plain film chest X-ray (CXR) is a commonly employed imaging modality in evaluating patients with traumatic injuries initially. It holds diagnostic value for assessing orthopedic injuries in the chest region, including fractures on ribs, clavicle fractures, and other related injuries. However, a high rate of misdiagnosis of rib fractures has been reported, above 50%,[8 9] which can have consequences such as inadequate pain control and the development of respiratory complications, including post-traumatic pneumonia resulting from undetected rib fractures.[10–12] CT is considered the gold standard for rib fracture detection. However, these modalities may pose additional challenges, such as increased radiation exposure, higher medical costs, and are only

available in advanced medical institutions. Thus, improving the accuracy of CXR diagnosis is crucial to enhancing patient care quality and preventing unnecessary radiation exposure.

Deep learning (DL), a rapidly evolving subfield of machine learning, has gained significant attention in medical image analysis.[13] DL has demonstrated successful outcomes in various classification tasks, including classification of abnormalities in chest radiography,[14 15] and interpretation of neural images.[16 17] The application of high-performance DL in computer-aided diagnosis (CAD) holds the potential to streamline human labor, enhance diagnostic consistency and accuracy, personalize patient treatment, and improve patient-doctor relationships[18 19] However, a major challenge in developing DL algorithms for medical image analysis is acquiring large-scale annotations of medical images, which is often labor-intensive and requires specialized expertise.[20 21] In response, several studies have focused on training deep convolutional neural networks using weak labels,[22–24] which can be obtained automatically or semiautomatically from medical records at a low cost. In the context of fracture detection, some studies have used weakly supervised learning to identify specific categories of fractures in defined regions, demonstrating comparable accuracy to physicians.[25–28] Nevertheless, a comprehensive and valuable CAD should possess a universal capability to detect various abnormalities within a single image. Currently, only a limited number of algorithms have showcased the ability to simultaneously detect abnormalities spanning multiple categories in an image.

In this study, we introduced CXR-FxNet, a DL-based detection algorithm trained using a weak labeling data set and limited number of expert annotations. We apply the concept of knowledge distillation in algorithm design to empower the performance of fracture detection. CXR-FxNet exhibits the ability to detect multiple trauma-related radiographic findings on CXRs, including clavicle and rib fractures.

MATERIALS AND METHODS
Development data set and image level label acquisition
The development data set was established by retrospectively reviewing the data in the trauma registry of trauma center A (Chang Gung Memorial Hospital, Linkou) recorded from May 2008 to December 2016. The demographic and trauma-related data, including age, gender, date of injury, and final diagnosis, were recorded. The first anteroposterior CXR, which is essential for evaluating the image modality for trauma patients, taken after the patient's arrival, was acquired from the picture archiving and communication system (PACS) repository. To ensure the image quality, we established criteria to guarantee that all included CXRs encompassed essential landmarks, including the C-spines, bilateral shoulder joints, and both sides of the diaphragms. Images which do not fulfil these criteria will be excluded from the data set. The images were deidentified and converted to Portable Network Graphics format for further processing.

All the CXRs were paired with a patient record in the trauma registry. The weak image-level label was parsed from a simple test-matching Python script using the International Classification of Diseases, Ninth Revision, Clinical Modification diagnosis code and the text of the final diagnosis to identify the presence of rib or clavicle fracture in the registry. Part of those images from patients with rib or clavicle fractures were delivered to two trauma surgeons specializing in thoracic trauma including rib and clavicle fracture with 15 years and 12 years of experience for further fracture site annotation. The image set with precise fracture site labeling was named 'expert labeled set', and

the other images with the image-level label only were named 'weakly labeled set'. The images chosen for expert review were randomly selected. Each image was assigned a randomly generated number, and the annotator labeled them sequentially to minimize bias in the selection process.

Fracture site annotation
The assessment of the images was conducted in conjunction with clinical diagnoses, radiologist-generated reports, and findings obtained through advanced imaging modalities, including anteroposterior oblique projection and chest CT scans. The reviewers were tasked with delineating a bounding box encompassing each identified fracture site of the rib and clavicle, designating the said box as either a 'rib fracture' or 'clavicle fracture'. In instances where multiple fracture sites were found, multiple annotations were applied accordingly. In instances where multiple fracture sites were found, every fracture site will be labeled with a bounding box accordingly. If contemporary posterior and anterior fractures are presented, the fracture sites will be labeled by different bounding boxes.

Test data sets
To independently evaluate the algorithm performance, we collected CXR from the same trauma center in 2017 using a similar process as the development data set. Since the clavicle fracture is relatively easy to identify, we focused on evaluating model performance on rib fracture identification. By sample size calculation, with a target accuracy of our model at 75% on power 0.8 with a 0.05 significance level, we performed a random selection process to choose 150 patients with fracture as positive images, matched with 150 patients without fractures, and collected as an 'independent test set' from hospital A. We also acquired another CXR data set from hospitalized trauma patients in regional hospital B (Chang Gung Memorial Hospital, Taipei) from 2018 to 2022 as an 'external test set'. The test set was independently collected from Hospital A and Hospital B which consisted of all expert-labeled images. All the test set images were reviewed by the experts and annotated with bounding boxes to confirm the diagnosis and location. All the images from patients younger than 18 years old were excluded.

Algorithm design
The foundational architecture of CXR-FxNet is predicated on a 'knowledge distillation' DL paradigm.[29] The objective is to leverage the 'weakly labeled set' to its fullest potential, enhancing the model's performance derived from the comparatively limited 'expert labeled set'. The fundamental architecture of the DL network incorporates the Feature Pyramid Network with a DenseNet-121 backbone. We use two models simultaneously in the training process. First, the model was initially pretrained with the expert-labeled set. The performance was limited due to the relatively small number of images. Next, the 'weakly labeled set' and the 'expert labeled set' were both used in the semisupervised training. The teacher and student models were initialized with the pretrained weights. Then, weakly labeled images were applied to both models. The predicted locations generated by the teacher model were adjusted with a sharpening algorithm and compared with the student model. The student model has also trained with the expert-labeled images in the same step. All the above information was integrated to adjust the student model and teacher model. After repeated training, the model converged to the best performance as figure 1 presents.
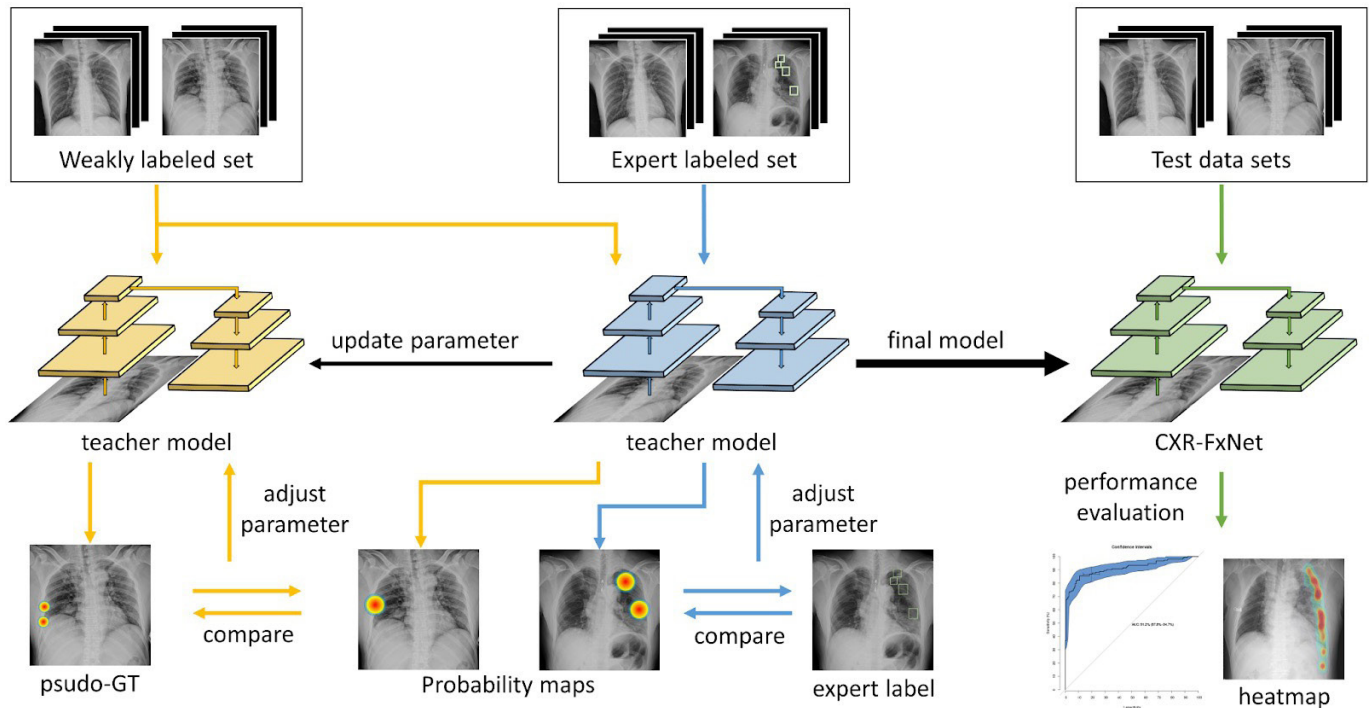
Figure 1 The foundational architecture of CXR-FxNet: A knowledge distillation deep learning paradigm to enhance a model's performance using both a limited 'expert labeled set' and a more extensive 'weakly labeled set'. The model, based on Feature Pyramid Network with a DenseNet-121 backbone, is pretrained with the expert-labeled set. Subsequent semisupervised training involves using both sets. The teacher model predicts locations on weakly labeled images, adjusting them with a sharpening algorithm. These predictions are then compared and integrated to fine-tune both the teacher and student models through repeated training until optimal performance is achieved. CXR, chest X-ray.

Statistical analysis and software

All the models were developed on a workstation with a single Intel Xeon E5-2650 v4 CPU (Central Processing Unit)@2.2 GHz, 128 GB RAM (Random Access Memory), and 4 NVIDIA Quadro RTX 8000 GPUs (Graph Processing Unit). Python V.3.6 and PyTorch V.1.6 under the operating system Ubuntu 18.04 LTS (Long Term Support). were used to design the algorithm. The statistical analysis was performed using R V.4.1.0 with the packages 'pROC' and 'table one'. The performance of the models was evaluated with the receiver operating characteristic (ROC) curve with area under the curve (AUC). The 95% CI of AUC was calculated using 2000 stratified bootstrap replicates. The accuracy, sensitivity, specificity, precision, and negative predictive value of the model on each test set were calculated with the two cut-off thresholds chosen on the independent test set with a high-sensitivity point and high-specificity point according to the clinical needs. The performance of the physicians was expressed with median with IQR and compared with CXR-FxNet with Mann–Whitney U test. The categorical parameters were compared with the χ^2 test. A vale of $p < 0.05$ indicated statistical significance.

RESULTS

We acquired a data set comprising 56 145 CXR images spanning from 2008 to 2016 at the trauma center A, as shown in figure 2. After the application of diagnostic codes and keyword matching, a total of 6886 images (15.2%) were identified as positive for rib/clavicle fractures, and the remaining 45 259 images (84.8%) were classified as negative. Among the positive cases, 991 CXRs were meticulously annotated by our domain experts, resulting in the delineation of 2740 bounding boxes corresponding to fracture sites, yielding an average of 2.8 boxes per image. There

are 146 patients who had both rib and clavicle fractures, 580 patients who had rib fractures only, and 185 patients who had clavicle fractures only. There are no patients who had bilateral clavicle fractures. Therefore, there are 331 clavicle bounding boxes and 2409 rib bounding boxes.

To construct our independent test set, we conducted a random selection process, drawing 300 CXRs from a patient pool of 6223 individuals treated at trauma center A in the year 2017. The selection was performed to maintain a balanced 1:1 ratio of CXRs with and without rib/clavicle fractures, as visually represented in figure 3. Detailed demographic attributes for each data set are concisely summarized in table 1. The CXR-FxNet showed an AUC of 91.2% (95% CI 87.7% to 94.7%) on the independent test set. The ROC curve is displayed in figure 4. The accuracy, sensitivity, specificity, precision, and negative predicted value on the high-sensitivity point are 83.3%, 87.5%, 79.1%, 81.1%, and 86.0%, respectively. The model achieved 97.3% specificity but a lower sensitivity of 71.1% on the high-specificity point. The details of the performance are displayed in table 2.

The external test set, obtained from hospital B, comprised 200 CXRs, evenly divided into 100 cases with rib/clavicle fractures and 100 without. Notably, the demographic characteristics, encompassing age and gender, exhibited significant disparities when compared with the data sets from hospital A. The patient cohort at hospital A skewed towards a younger age profile and was predominantly male. When evaluated on this external data set, CXR-FxNet demonstrated comparable sensitivity (88.9% vs 87.5%) but exhibited diminished accuracy compared with the independent test set (74.2% vs 83.3%) on the high-sensitivity cut-off. On the high-specificity cut-off, the model had comparable accuracy (80.8% vs 84.3%) but lower specificity (90.9% vs 97.3%) on the external data set. The CXR-FxNet detected
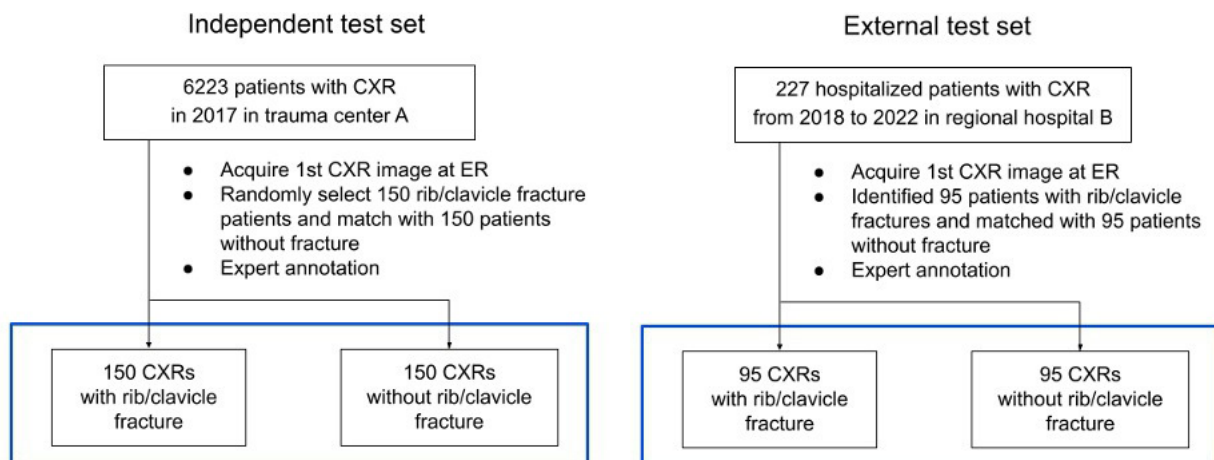
Development data set
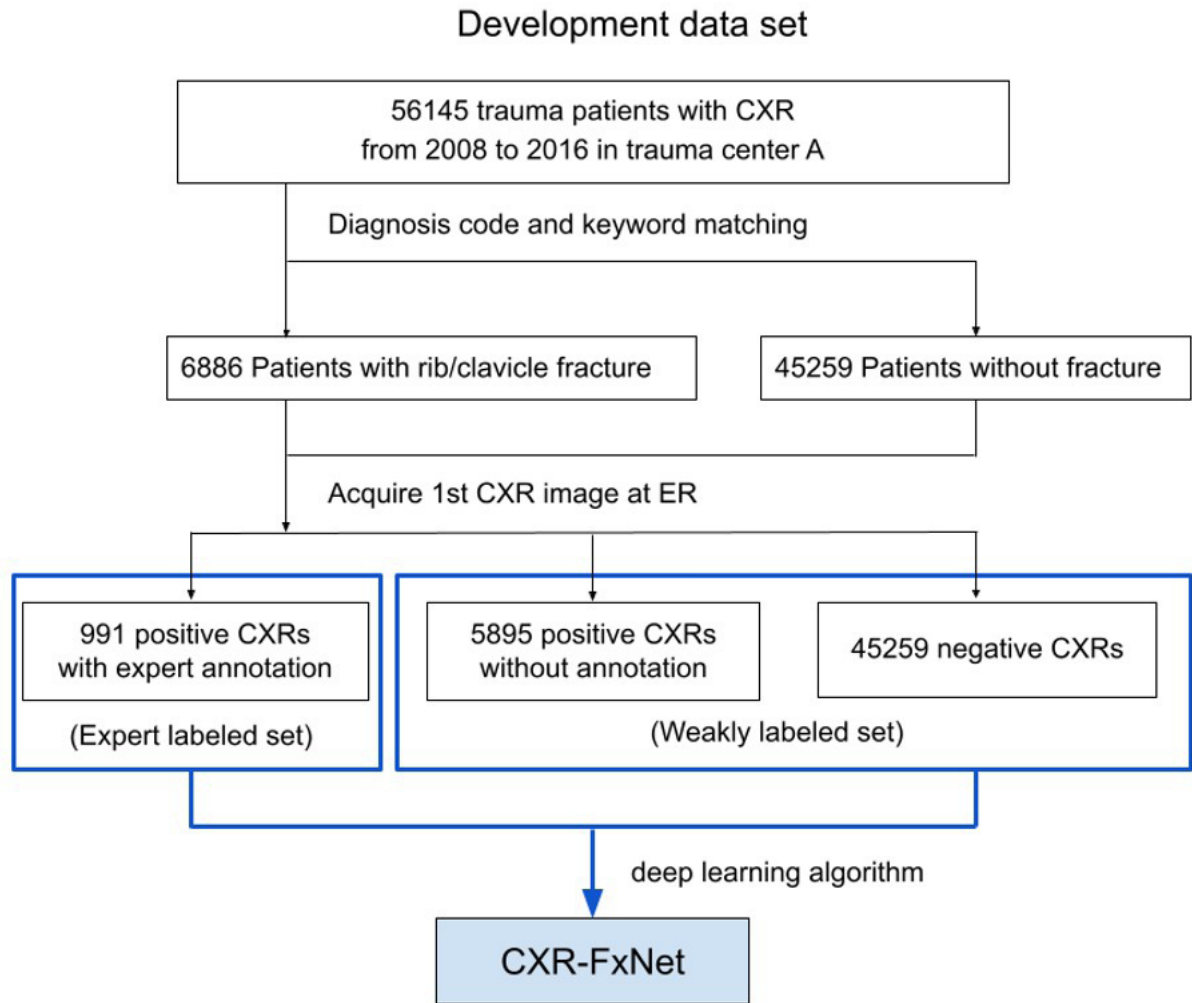
Figure 2 The selection of the development data set. CXR, chest X-ray.

all the clavicle fractures. Among the 32 clavicle fractures in the independent test set, at the high-sensitivity cut-off point, only 1 (3.1%) of the clavicle fractures was missed. At the high specificity point, only 3 (9.4%) of the clavicle fractures were missed. Figure 5 illustrates a heatmap generated by CXR-FxNet. Notably, this model has exhibited the capability to concurrently detect multiple fractures within a single CXR. Furthermore, the

model demonstrates the capacity to identify abnormalities even in cases where the fracture site exhibits no displacement.

DISCUSSION

A well-designed CAD algorithm can potentially reduce medical errors and facilitate accurate diagnoses.[25 30] However, there is

Figure 3 The test data set in this study for evaluation of the rib fracture detection performance. CXR, chest X-ray

Table 1 Characteristics of each data set

|	Development data sets		Test data sets	
	Weakly labeled set	Expert labeled set	Independent test set	External test set
Image numbers, n	55 154	991	300	198
Age, mean (SD)	46.63 (20.20)	51.36 (18.36)	52.45 (20.38)	60.40 (18.82)
Gender, n (%)				
Male	34 822 (63.1)	279 (28.2)	193 (64.3)	98 (49.5)
Female	20 332 (36.9)	712 (71.8)	107 (35.7)	100 (50.5)
Rib fracture, n (%)				
Yes	3161 (5.7)	726 (73.3)	150 (50.0)	89 (44.9)
No	51 834 (94.0)	265 (26.7)	150 (50.0)	109 (55.1)
Clavicle fracture, n (%)				
Yes	2674 (4.8)	331 (33.4)	32 (10.7)	19 (9.6)
No	52 328 (94.9)	660 (66.6)	268 (89.3)	181 (90.4)
Year of image acquisition	2008–2016	2008–2016	2017	2018–2022

SD, Standard deviation.

currently a lack of generalized and comprehensive algorithms for interpreting chest radiography in the trauma domain. Although DL algorithms have shown promise in detecting abnormalities in radiographs, there is still a gap between developing scientifically sound algorithms and their practical implementation in real-world settings.[31] In this study, we developed an algorithm based on a novel weak-supervised DL method that achieved high performance in identifying multitasks of trauma-related skeletal radiographic findings on CXRs to fit the clinical requirement. CXR-FxNet achieved an AUC of 91.2% in an independent data set and showed the ability to localize rib and clavicle fractures in CXR.

Accurate diagnosis is essential, as failure to do so could result in a bleak prognosis. The utilization of this algorithm presents an opportunity to make timely improvements in clinical performance and safety. DL has gained substantial traction in the

medical field, however, the application of DL in trauma assessment is still somewhat limited in real-world clinical scenarios.[32–34] DL algorithms in the medical field must exhibit performance comparable to that of physicians to generate meaningful clinical benefits.[35] Current available applications were still focused on detecting skeletal fractures of the pelvis and extremities.[28 36 37] Previous studies have demonstrated that algorithms can achieve similar performance to physicians in detecting various fractures on radiographs, including proximal humerus fractures,[38] wrist fractures,[25] and hip fractures.[39] This highlights the potential of CXR-FxNet to assist in the identification of these fractures. Indeed, the use of the CXR-FxNet algorithm can provide real-time recommendations to front-line physicians as they manage multiple trauma patients in a chaotic emergency environment, where misdiagnoses can occur.[40] Specifically, in the case of rib fractures, our algorithm has the capability to detect multiple rib and clavicle fractures as in figure 5. This feature proves particularly valuable in healthcare institutions that may lack access to consulting specialists or experienced medical staff.[41] By providing timely and accurate insights, the algorithm can enhance the diagnostic capabilities of front-line physicians and contribute to improved patient care in such settings.

In contrast to extremity radiographs, CXR shows complex anatomy, with frequent multiple injury sites and pathologies. The soft tissue components such as mediastinum and foreign catheters ex chest tube might induce misdiagnosis. In the contemporary medical environment,[28] developing separate algorithms for each type of anomaly present in a single image is not feasible. Consequently, there is a pressing need for universal solutions tailored to specific clinical scenarios in emergency CXR. Due to the complex anatomy, the development of DL is very rare in thoracic trauma. Most applications focus on chest CT algorithms for diagnosing rib fractures.[12 19 42–48] Although the models based on chest CT exhibited commendable performance, there were still certain limitations. First, medical costs, availability, and radiation exposure considerations limit the widespread use of CT in trauma evaluations, as it is not typically employed as the primary survey tool in most parts of the world. Second, the considerable volume of images and data associated with CT poses challenges. When we are training the DL algorithm using CT images, the data amount can be tens to hundreds of times larger compared with CXR. Consequently, the complexity of the calculations, the high computational power requirements, and the difficulty of integrating into the medical examination
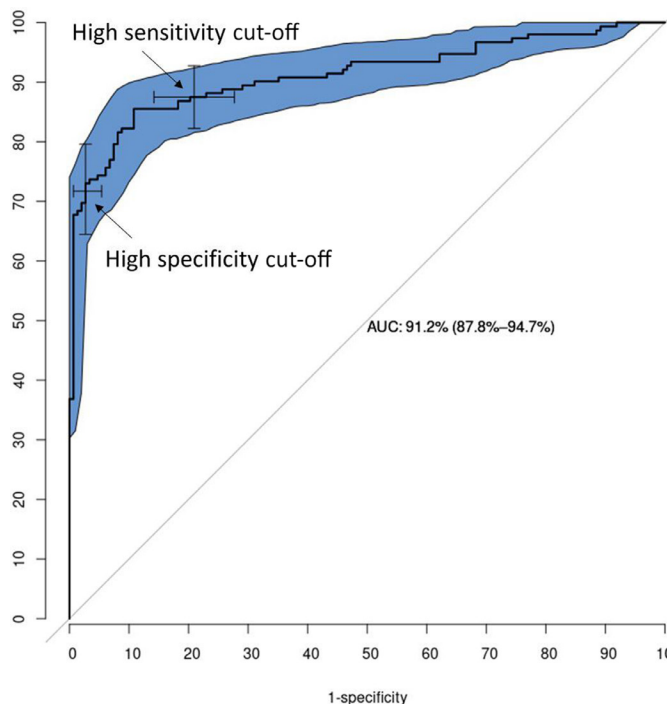
Figure 4 The ROC curve of CXR-FxNet. AUC, the area under curve; CXR, chest X-ray; ROC, receiver operating characteristic.

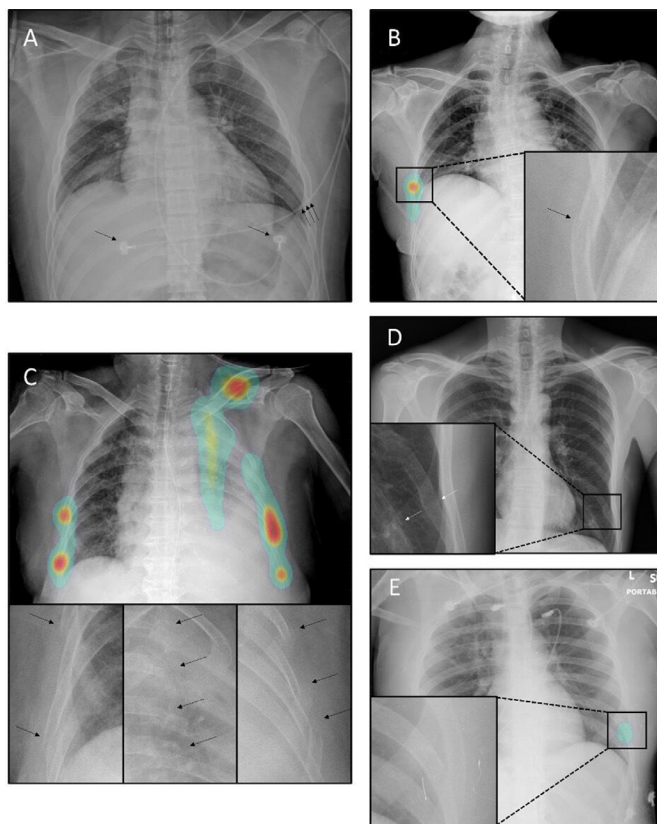**Table 2** The performance of CXR-FxNet on each test set

	High sensitivity cut-off		High specificity cut-off	
	Independent test set	External test set	Independent test set	External test set
Accuracy %, (95% CI)	0.833 (0.786 to 0.874)	0.742 (0.676 to 0.802)	0.843 (0.797 to 0.883)	0.808 (0.746 to 0.860)
Sensitivity %, (95% CI)	0.875 (0.812 to 0.923)	0.889 (0.810 to 0.943)	0.717 (0.638 to 0.787)	0.707 (0.607 to 0.794)
Specificity %, (95% CI)	0.791 (0.716 to 0.853)	0.596 (0.493 to 0.693)	0.973 (0.932 to 0.993)	0.909 (0.834 to 0.958)
Precision %, (95% CI)	0.811 (0.743 to 0.868)	0.688 (0.600 to 0.766)	0.965 (0.912 to 0.990)	0.886 (0.795 to 0.947)
NPV %, (95% CI)	0.860 (0.790 to 0.914)	0.843 (0.736 to 0.919)	0.770 (0.703 to 0.828)	0.756 (0.669 to 0.830)

CI, Confidence interval; CXR, chest X-ray; NPV, negative predictive value.

process are the limitations that these algorithms cannot be used on a global scale. Unlike CT, CXR is much more readily available in any hospital and it was looked at as the primary modality for evaluating trauma patients. Here we have introduced CXR-FxNet, which can offer some advantages. First, the CXR-FxNet algorithm demonstrates the capability to accurately identify and localize various trauma-related abnormalities. Its ability to detect multiple categories of abnormalities simultaneously, across multiple locations within an image, enhances physicians' confidence in the algorithm and facilitates its widespread adoption in clinical practice. Second, CXR-FxNet used CXR instead of

Figure 5 Visualization examples of CXR-FxNet. (A) True negative prediction. The arrow shows the model was not misleading by monitor leads and wires. (B) True positive prediction. The model detected one minimally displaced single fracture site. (C) True positive prediction of multiple fracture sites. The model simultaneously detected left clavicle fracture, left posterior third to sixth rib fractures, left lateral third to seventh rib fractures, and right lateral fifth and seventh rib fractures. (D) False negative prediction. The model missed left minimally displaced ninth and tenth rib fractures. (E) False positive prediction. The model mistakenly identified an artifact as a fracture site.

CT which helps reduce computational demands and standardize image quality. This approach allows for consistent diagnostic capabilities even in hospitals with limited medical information resources. In contrast to models relying on CT images, our DL model is more lightweight, accessible, and user-friendly, enabling a broader range of people to use it conveniently. For the institutes that can afford DL calculation server and PACS systems, the requirements and costs of information systems can decrease compared with high computation-requiring systems. For those unable to afford this additional equipment, the model can be set up on the cloud. We've also designed a website (website link: http://140.129.68.84:8081/) for easy online setup for public use and validation. The health providers can upload the CXR images taken with their cameras or mobile phones to the web and receive the DL model-assisted feedback within seconds. In this study, we also found an interesting result as previous research suggests that DL algorithms may be beneficial for younger and less experienced physicians. With the help of the DL algorithm, junior staff are able to locate fracture sites with performance comparable to that of experienced physicians.

The development of DL models in the medical field is often hindered by limited data size and the lack of clear labeling. The image-level label is relatively easy to acquire through medical records, but the detailed expert label on the image is excessively expensive. Weakly supervised methods have emerged as a potential solution, offering the ability to achieve a reasonably high baseline performance even with large but somewhat noisy data sets. In this study, we not only explored the use of weakly supervised methods relying solely on image-level information but also assessed the impact of incorporating bounding box annotated images on model performance. We tried the teacher-student knowledge distillation method in the current study to improve the model performance with few expert annotations. This evaluation aimed to analyze whether adding high-quality, detailed annotations could further enhance the model's accuracy compared with relying solely on weakly supervised methods. As a result, we found that adding more detailed information to the model reduced the need for training images and yielded better results.

Limitations

In addition to achieving excellent performance in the detection of rib and clavicle fractures, our algorithm represents the first study to successfully develop an algorithm capable of detecting such fractures from CXR, to the best of our knowledge. However, it is important to acknowledge the limitations of this algorithm. The primary limitation stems from the scarcity of training data available. DL algorithms are data-driven and rely on large data sets to effectively address problems. Despite implementing a weak labeling algorithm, this limitation could not be entirely overcome. Due to time and cost constraints, radiologists were

not used for image review and labeling. Two experienced trauma surgeons specializing in rib management undertook this task, with potential limitations in achieving standard labeling levels. No inter-rater reliability assessment is another limitation for data labeling for this study. Another limitation is the retrospective nature of this single-institute image review study. The population and image collection process were confined to a specific setting, potentially introducing biases that limit the direct applicability of our findings to other institutes with different population distributions. Moreover, the images were randomly selected based on the clinical diagnosis from the registry, so that the presence of selective bias cannot be completely excluded.

DL algorithms are often referred to as 'black boxes' because their primary function is to establish relationships between given data and outcomes. To address this issue, recent research has focused on interpretable DL techniques. In our study, we incorporated a visual heatmap highlighting areas of possible abnormality to aid doctors in understanding the algorithm's decision-making process. However, it is important to note that in real-world scenarios, physicians make diagnoses by radiographic findings and by clinical information such as patient histories and physical examinations. The true benefit of this algorithm should be evaluated in a prospective randomized clinical trial, considering the comprehensive clinical environment.

CONCLUSION

This study demonstrates that a universal trauma-related detection algorithm for CXR can be trained and scalable with limited weakly supervised annotations and performs well on both clinical scenario distribution data sets and balanced data sets. This is the first algorithm to detect rib and clavicle fractures simultaneously and can prevent misdiagnosis of these injuries in practical applications. Future prospective studies are needed to validate whether the application of this CXR-FxNet as a computer-aided diagnostic system in clinical scenarios leads to more accurate diagnosis and facilitates the management of trauma patients.

ORCID iDs
Chi-Tung Cheng http://orcid.org/0000-0002-2697-4642
Ling-Wei Kuo http://orcid.org/0000-0001-6489-3723
Chien-Hung Liao http://orcid.org/0000-0003-2812-9773

REFERENCES

1 Peek J, Ochen Y, Saillant N, Groenwold RHH, Leenen LPH, Uribe-Leitz T, Houwert RM, Heng M. Traumatic rib fractures: a marker of severe injury. A nationwide study using the National trauma data bank. *Trauma Surg Acute Care Open* 2020;5:e000441.

2 Bulger EM, Arneson MA, Mock CN, Jurkovich GJ. Rib fractures in the elderly. *J Trauma* 2000;48:1040–6;

3 Ziegler DW, Agarwal NN. The morbidity and mortality of rib fractures. *J Trauma* 1994;37:975–9.

4 Blum A, Gillet R, Urbaneja A, Gondim Teixeira P. Automatic detection of rib fractures: are we there yet *EBioMedicine* 2021;63:103158.

5 Fitzgerald MT, Ashley DW, Abukhdeir H, Christie DB 3rd. Rib fracture fixation in the 65 years and older population: a paradigm shift in management strategy at a level I trauma center. *J Trauma Acute Care Surg* 2017;82:524–7.

6 van Laarhoven JJEM, Hietbrink F, Ferree S, Gunning AC, Houwert RM, Verleisdonk EMM, Leenen LPH. Associated Thoracic injury in patients with a Clavicle fracture: a retrospective analysis of 1461 Polytrauma patients. *Eur J Trauma Emerg Surg* 2019;45:59–63.

7 Mouzopoulos G, Morakis E, Stamatakos M, Tzurbakis M. Complications associated with clavicular fracture. *Orthop Nurs* 2009;28:217–24;

8 Talbot BS, Gange CP, Chaturvedi A, Klionsky N, Hobbs SK, Chaturvedi A. Traumatic rib injury: patterns, imaging pitfalls complications, and treatment. *Radiographics* 2017;37:628–51.

9 Traub M, Stevenson M, McEvoy S, Briggs G, Lo SK, Leibman S, Joseph T. The use of chest computed tomography versus chest X-ray in patients with major blunt trauma. *Injury* 2007;38:43–7.

10 Ho S-W, Teng Y-H, Yang S-F, Yeh H-W, Wang Y-H, Chou M-C, Yeh C-B. Risk of pneumonia in patients with isolated minor rib fractures: a nationwide cohort study. *BMJ Open* 2017;7:e013029.

11 Tanaka H, Yukioka T, Yamaguti Y, Shimizu S, Goto H, Matsuda H, Shimazaki S. Surgical stabilization of internal pneumatic stabilization? A prospective randomized study of management of severe flail chest patients. *J Trauma* 2002;52:727–32;

12 Weikert T, Noordtzij LA, Bremerich J, Stieltjes B, Parmar V, Cyriac J, Sommer G, Sauter AW. Assessment of a deep learning algorithm for the detection of rib fractures on whole-body trauma computed tomography. *Korean J Radiol* 2020;21:891–9.

13 Langerhuizen DWG, Bulstra AEJ, Janssen SJ, Ring D, Kerkhoffs GMMJ, Jaarsma RL, Doornberg JN. Is deep learning on par with human observers for detection of Radiographically visible and occult fractures of the Scaphoid *Clin Orthop Relat Res* 2020;478:2653–9.

14 Meedeniya D, Kumarasinghe H, Kolonne S, Fernando C, Díez ID la T, Marques G. Chest X-ray analysis empowered with deep learning: a systematic review. *Appl Soft Comput* 2022;126:109319.

15 Santomartino SM, Hafezi-Nejad N, Parekh VS, Yi PH. Performance and usability of code-free deep learning for chest radiograph classification, object detection, and segmentation. *Radiol Artif Intell* 2023;5:e220062.

16 Daugaard Jørgensen M, Antulov R, Hess S, Lysdahlgaard S. Convolutional neural network performance compared to Radiologists in detecting intracranial hemorrhage from brain computed tomography: a systematic review and meta-analysis. *Eur J Radiol* 2022;146:110073.

17 Jyothi P, Singh AR. Deep learning models and traditional automated techniques for brain tumor Segmentation in MRI: a review. *Artif Intell Rev* 2023;56:2923–69.

18 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.

19 Jin L, Yang J, Kuang K, Ni B, Gao Y, Sun Y, Gao P, Ma W, Tan M, Kang H, *et al*. Deep-learning-assisted detection and segmentation of rib fractures from CT scans: development and validation of Fracnet. *EBioMedicine* 2020;62:103106.

20 Chea P, Mandell JC. Current applications and future directions of deep learning in musculoskeletal radiology. *Skeletal Radiol* 2020;49:183–97.

21 Pinto A, Berritto D, Russo A, Riccitiello F, Caruso M, Belfiore MP, Papapietro VR, Carotti M, Pinto F, Giovagnoni A, *et al*. Traumatic fractures in adults: missed diagnosis on plain radiographs in the emergency department. *Acta Biomed* 2018;89:111–23.

22 Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chestx-Ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common Thorax diseases. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017,

23 Wang H, Jia H, Lu L, Xia Y. Thorax-Net: an attention Regularized deep neural network for classification of Thoracic diseases on chest radiography. *IEEE J Biomed Health Inform* 2020;24:475–85.

24 Üreten K, Arslan T, Gültekin KE, Demir AND, Özer HF, Bilgili Y. Detection of hip osteoarthritis by using plain pelvic Radiographs with deep learning methods. *Skeletal Radiol* 2020;49:1369–74.

25 Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, Hanel D, Gardner M, Gupta A, Hotchkiss R, *et al*. Deep neural network improves fracture detection by Clinicians. *Proc Natl Acad Sci U S A* 2018;115:11591–6.

26 Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol* 2018;73:439–45.

27 Olczak J, Emilson F, Razavian A, Antonsson T, Stark A, Gordon M. Ankle fracture classification using deep learning: automating detailed AO foundation/orthopedic trauma Association (AO/OTA) 2018 Malleolar fracture identification reaches a high degree of correct classification. *Acta Orthop* 2021;92:102–8.

28 Cheng C-T, Wang Y, Chen H-W, Hsiao P-M, Yeh C-N, Hsieh C-H, Miao S, Xiao J, Liao C-H, Lu L. A Scalable physician-level deep learning algorithm detects universal trauma on pelvic radiographs. *Nat Commun* 2021;12:1066.

29 Wang Y, Zheng K, Cheng C-T, *et al*. Knowledge distillation with adaptive asymmetric label sharpening for semi-supervised fracture detection in chest X-rays. In: *Information Processing in Medical Imaging*. Springer International Publishing, 2021: 599–610.

30 Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer. *Am J Med* 2018;131:129–33.

31 Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. *NPJ Digit Med* 2018;1:40.

32 Laur O, Wang B. Musculoskeletal trauma and artificial intelligence: Current trends and projections. *Skeletal Radiol* 2022;51:257–69.

33 Dreizin D, Goldmann F, LeBedis C, Boscak A, Dattwyler M, Bodanapally U, Li G, Anderson S, Maier A, Unberath M. An automated deep learning method for tile AO/OTA pelvic fracture severity grading from trauma whole-body CT. *J Digit Imaging* 2021;34:53–65.

34 Farzaneh N, Stein EB, Soroushmehr R, Gryak J, Najarian K. A deep learning framework for automated detection and quantitative assessment of liver trauma. *BMC Med Imaging* 2022;22:39.

35 Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, Topol EJ, Ioannidis JPA, Collins GS, Maruthappu M. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 2020;368:m689.

36 Tanzi L, Vezzetti E, Moreno R, Moos S. X-ray bone fracture classification using deep learning: A baseline for designing a reliable approach. *Applied Sciences* 2020;10:1507.

37 Jones RM, Sharma A, Hotchkiss R, Sperling JW, Hamburger J, Ledig C, O'Toole R, Gardner M, Venkatesh S, Roberts MM, *et al*. Assessment of a deep-learning system for fracture detection in musculoskeletal radiographs. *NPJ Digit Med* 2020;3:144.

38 Chung SW, Han SS, Lee JW, Oh K-S, Kim NR, Yoon JP, Kim JY, Moon SH, Kwon J, Lee H-J, *et al*. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop* 2018;89:468–73.

39 Cheng C-T, Ho T-Y, Lee T-Y, Chang C-C, Chou C-C, Chen C-C, Chung I-F, Liao C-H. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. *Eur Radiol* 2019;29:5469–77.

40 Tammelin E, Handolin L, Söderlund T. Missed injuries in Polytrauma patients after trauma tertiary survey in trauma intensive care unit. *Scand J Surg* 2016;105:241–7.

41 Mattijssen-Horstink L, Langeraar JJ, Mauritz GJ, van der Stappen W, Baggelaar M, Tan ECTH. Radiologic discrepancies in diagnosis of fractures in a Dutch teaching emergency Department: a retrospective analysis. *Scand J Trauma Resusc Emerg Med* 2020;28:38.

42 Castro-Zunti R, Chae KJ, Choi Y, Jin GY, Ko S-B. Assessing the speed-accuracy trade-offs of popular convolutional neural networks for single-crop rib fracture classification. *Comput Med Imaging Graph* 2021;91:101937.

43 Liu X, Wu D, Xie H, Xu Y, Liu L, Tao X, Wang X. Clinical evaluation of AI software for rib fracture detection and its impact on Junior Radiologist performance. *Acta Radiol* 2022;63:1535–45.

44 Meng XH, Wu DJ, Wang Z, Ma XL, Dong XM, Liu AE, Chen L. A fully automated rib fracture detection system on chest CT images and its impact on Radiologist performance. *Skeletal Radiol* 2021;50:1821–8.

45 Zhang B, Jia C, Wu R, Lv B, Li B, Li F, Du G, Sun Z, Li X. Improving rib fracture detection accuracy and reading efficiency with deep learning-based detection software: a clinical evaluation. *Br J Radiol* 2021;94:20200870.

46 Kaiume M, Suzuki S, Yasaka K, Sugawara H, Shen Y, Katada Y, Ishikawa T, Fukui R, Abe O. Rib fracture detection in computed tomography images using deep Convolutional neural networks. *Medicine (Baltimore)* 2021;100:e26024.

47 Yao L, Guan X, Song X, Tan Y, Wang C, Jin C, Chen M, Wang H, Zhang M. Rib fracture detection system based on deep learning. *Sci Rep* 2021;11:23513.

48 Zhou QQ, Wang J, Tang W, Hu ZC, Xia ZY, Li XS, Zhang R, Yin X, Zhang B, Zhang H. Automatic detection and classification of rib fractures on Thoracic CT using Convolutional neural network: accuracy and feasibility. *Korean J Radiol* 2020;21:869–79.