European Radiology (2021) 31:9612–9619
https://doi.org/10.1007/s00330-021-08014-5

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Differential diagnosis of benign and malignant vertebral fracture on CT using deep learning

Yuan Li[1] · Yang Zhang[2] · Enlong Zhang[3] · Yongye Chen[1] · Qizheng Wang[1] · Ke Liu[1] · Hon J. Yu[2] · Huishu Yuan[1] · Ning Lang[1] · Min-Ying Su[2]

Received: 7 January 2021 / Revised: 21 April 2021 / Accepted: 26 April 2021 / Published online: 16 May 2021

Abstract

Objectives To evaluate the performance of deep learning using ResNet50 in differentiation of benign and malignant vertebral fracture on CT.

Methods A dataset of 433 patients confirmed with 296 malignant and 137 benign fractures was retrospectively selected from our spinal CT image database. A senior radiologist performed visual reading to evaluate six imaging features, and three junior radiologists gave diagnostic prediction. A ROI was placed on the most abnormal vertebrae, and the smallest square bounding box was generated. The input channel into ResNet50 network was 3, including the slice with its two neighboring slices. The diagnostic performance was evaluated using 10-fold cross-validation. After obtaining the malignancy probability from all slices in a patient, the highest probability was assigned to that patient to give the final diagnosis, using the threshold of 0.5.

Results Visual features such as soft tissue mass and bone destruction were highly suggestive of malignancy; the presence of a transverse fracture line was highly suggestive of a benign fracture. The reading by three radiologists with 5, 3, and 1 year of experience achieved an accuracy of 99%, 95.2%, and 92.8%, respectively. In ResNet50 analysis, the per-slice diagnostic sensitivity, specificity, and accuracy were 0.90, 0.79, and 85%. When the slices were combined to ve per-patient diagnosis, the sensitivity, specificity, and accuracy were 0.95, 0.80, and 88%.

Conclusion Deep learning has become an important tool for the detection of fractures on CT. In this study, ResNet50 achieved good accuracy, which can be further improved with more cases and optimized methods for future clinical implementation.

Key Points

- *Deep learning using ResNet50 can yield a high accuracy for differential diagnosis of benign and malignant vertebral fracture on CT.*
- *The per-slice diagnostic sensitivity, specificity, and accuracy were 0.90, 0.79, and 85% in deep learning using ResNet50 analysis.*
- *The slices combined with per-patient diagnostic sensitivity, specificity, and accuracy were 0.95, 0.80, and 88% in deep learning using ResNet50 analysis.*

Keywords Deep learning · Spinal fractures · Tomography, X-ray computed · Diagnosis, differential

Abbreviations

AI	Artificial Intelligence
CNN	Convolutional neural network
FDA	Food and Drug Administration
FN	False negative
FP	False positive
PSTM	Paraspinal soft tissue mass
ROI	Region of interest
TN	True negative
TP	True positive

✉ Ning Lang
langning800129@126.com

[1] Department of Radiology, Peking University Third Hospital, 49 North Garden Road, Haidian District, Beijing 100191, People's Republic of China

[2] Department of Radiological Sciences, University of California, Irvine, CA, USA

[3] Department of Radiology, Peking University International Hospital, Beijing, People's Republic of China

Introduction

Vertebral fractures are common, increase in prevalence with age, and can lead to acute and chronic pain [1]. The major etiologies are trauma and osteoporosis (benign) and neoplastic infiltration (malignant), which need to be confirmed for deciding the treatment strategy [2]. Patients with malignancy need to receive treatment soon to control the cancer spread, and also to prevent further bone damages for maintaining a good quality of life [3]. Benign fractures can be caused by trauma, osteoporosis, or in combination, and in general, patients can achieve a good outcome with proper treatments. In the elderly, osteoporotic vertebral compression fractures are common [4], and they may coexist with malignant fractures. This coexistence complicates the evaluation and the differential diagnosis of the fractures.

Differential diagnosis of benign and malignant vertebral fractures in the clinical setting is often carried out using advanced imaging modalities such as CT, MRI, PET, and SPECT [5]. Among these, CT can be easily performed, which is often used to confirm and further characterize fractures suspected on plain X-ray [6, 7]. With its high spatial resolution, CT can also be used to distinguish benign from malignant fractures based on morphological features such as osseous integrity and fracture margins/patterns [5]. Although some typical malignancy features can be demonstrated in CT [8, 9], they may not be well differentiated when mixed with osteoporosis or chronic damage in the growing elderly population [4, 5]. This can be particularly challenging for inexperienced radiologists.

Computer-aided analysis within the segmented vertebra has been applied to differentiate traumatic, osteoporotic, and malignant vertebral fractures, e.g., using histogram [10]. Recently, deep learning has become an active area of research demonstrating as a feasible technique for the management of fracture [11, 12]. The diagnostic application in plain radiography has been demonstrated in wrists, humerus, hip, femur, shoulders, hands, and feet [6, 7, 13–16]. The application in CT has also been reported, but limited in the detection of posttraumatic [17, 18] and osteoporotic fractures [19]. Besides, Deep learning has also been successfully applied to perform automatic vertebral segmentation in the spine [20–23] and opportunistic osteoporosis screening [24]. The aim of this study is to evaluate the feasibility of deep learning using ResNet50 algorithm in differential diagnosis of benign and malignant vertebral fracture on CT.

Materials and methods

Patient datasets

The patients were selected from the radiological reporting system of our hospital's spinal CT image database. A total of 433 patients were identified (mean age 59.4, age range of 14–93 years), 137 with benign (mean age 62.8, age range of 14–93 years) and 296 with malignant fracture (mean age 57.8, age range of 27–84 years). The malignant cases had either biopsy-proven cancer or known history of primary tumor with progressive disease. All benign cases had no known cancer history and have been followed up with stable disease. The CT images were acquired in axial orientation using a GE Discovery CT 750HD scanner with 120 kV, 137~543 mAs, and 3-mm thickness. The acquired images were reformatted to sagittal view (3-mm thickness) for further analysis.

Visual assessment and scoring

One senior radiologist with 20 years of experience carried out a visual review to score 6 features on the CT images, using a binary classification of 0 (absent) vs. 1 (present). These evaluated features were (1) vertebral compression (reduction in vertebral height more than 20%, compared with adjacent normal vertebrae); (2) bone cortex interruption; (3) paravertebral soft tissue mass; (4) transverse fracture line; (5) involving multiple segments; and (6) bone destruction. For testing the reading performance of junior radiologists with different levels of training, three radiologists with 5, 3, and 1 year of experience reviewed the datasets and gave benign or malignant diagnosis for each patient. The cases were shuffled in the reading session to avoid bias. All radiologists were blinded to the final diagnosis and the clinical history or outcome of the patients.

Deep learning architecture

In each patient, the vertebral segment that showed the worst abnormality was selected for deep learning analysis. A region of interest (ROI) was manually placed on a sagittal slice, and then the smallest square bounding box containing the entire abnormal area was generated by computer program and used as input. Deep learning was performed using the ResNet50 architecture, with the binary output of malignant and benign, as shown in Fig. 1. The input network included the slice along with its two adjacent neighboring slices. Therefore, the number of input channel was 3. The bounding box was re-sampled to a 64 × 64 matrix using linear interpolation, and then the pixel intensities were normalized to have mean of 0 and standard deviation of 1. In contrast to other convolutional neural networks (CNNs), such as VGG or AlexNet that learns features using large convolutional network architectures, the ResNet extracts residual features as subtraction of features learned from input of that layer using "skip connections." The ResNet50 architecture contained one 3 × 3 convolutional layer, one max-pooling layer, and 16 residual blocks. Each block contained one 1 × 1 convolutional layer, one 3 × 3 convolutional layer, and one 1 × 1 convolutional layer. The residual connection was from the beginning of the block to the end of the block. The output of the last block was connected to

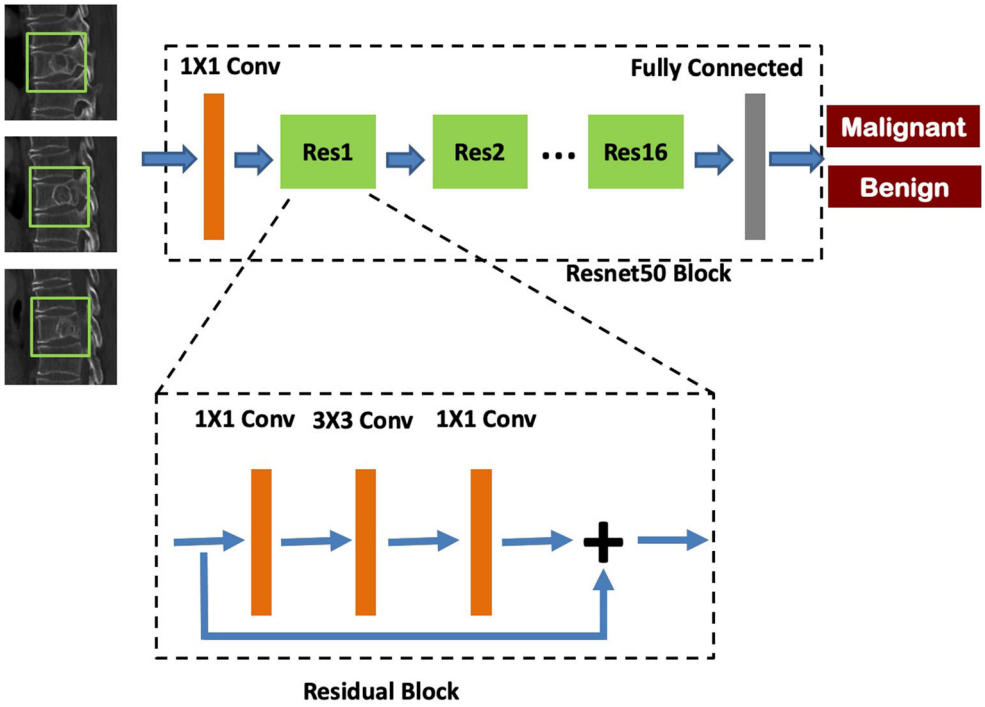

Fig. 1 Architecture of ResNet50, containing 16 residual blocks. Each residual block begins with one 1 × 1 convolutional layer, followed by one 3 × 3 convolutional layer, and ends with another 1 × 1 convolutional layer. The output is then added to the input via a residual connection. The total input channel is 3, including the slice along with its two adjacent neighboring slices. The output is binary, malignant, or benign

a fully connected layer with a sigmoid function to make the prediction, by giving a malignancy probability.

Training and evaluation methods

In each patient, 3–5 sagittal slices that clearly showed abnormality were analyzed. The dataset was further augmented 20 times by using random affine transformations, including translation, scaling, and rotation. To control for overfitting, L2 regularization term was added to the final loss function, and then, during the training process, the early stop was applied based on the lowest validation loss to obtain the optimized model. The loss function was cross-entropy. The training was implemented using the Adam (Adaptive Moment Estimation) optimizer [25]. The learning rate was set to 0.0001 with momentum term β to 0.5 to stabilize training. Parameters were initialized using ImageNet. The batch size was set to 32 and the number of epochs was set to 100.

The evaluation was performed using 10-fold cross-validation, 9 fold for training, and 1 fold set aside for testing. Each case had one chance to be included in the testing dataset. The output of deep learning was a malignancy probability for each input slice. Then, the results of all slices in a patient were combined to give a per-patient diagnosis, by assigning the highest probability among all slices to that patient.

Statistical analysis

All statistical analyses were performed using SPSS version 22.0. Data following normal distribution were shown as mean ± standard deviation. The presence of 6 evaluated imaging features was compared between malignant and benign groups using the χ^2 test. The diagnostic performance of ResNet50 was calculated based on per-slice and per-patient basis, according to the threshold of 0.5 (i.e., probability ≥ 0.5 as malignant). The sensitivity, specificity, and overall accuracy of three readers and the ResNet50 prediction were calculated.

Results

Radiologists' reading results

The presence and absence of 6 features determined by the senior radiologist are summarized in Table 1, and the final diagnosis of three junior readers is also included in Table 1. All six features show significant differences between benign and malignant groups with $p \leq 0.001$. The soft tissue mass and bone destruction were highly suggesting malignancy, and the presence of transverse fracture line was highly suggesting benign. The age of benign patients was significantly older than malignant patients (62.8 ± 16.1 vs. 57.8 ± 11.3, $p = 0.001$). There were more males in the malignant group, and more females in the benign groups, consistent with women's higher risk of osteoporosis when they aged. The reading performed by three radiologists with 5, 3, and 1 year of experience achieved an overall accuracy of 99%, 95.2%, and 92.8%, respectively.

Table 1 The reading of imaging features and predicted diagnosis made by three readers between malignant and benign groups

Parameters/imaging features	Malignant ($N = 296$)	Benign ($N = 137$)	p value
Age (years)	57.8 ± 11.3	62.8 ± 16.1	0.001
Gender (male/female)	170/126	53/84	< 0.001
Vertebral compression	250 (84.5%)	133 (97.1%)	< 0.001
Bone cortex interruption	290 (98.0%)	125 (91.2%)	0.001
Paravertebral soft tissue mass	276 (93.2%)	2 (1.5%)	< 0.001
Transverse fracture line	3 (1.0%)	109 (79.6%)	< 0.001
Multiple segments	161 (54.4%)	30 (21.9%)	< 0.001
Bone destruction	294 (99.3%)	1 (0.7%)	< 0.001
Reader-1 diagnostic prediction	294 (99.3%)	135 (98.5%)	
Reader-2 diagnostic prediction	281 (94.9%)	131 (95.6%)	
Reader-3 diagnostic prediction	275 (92.9%)	127 (92.7%)	

ResNet50 diagnostic results

Figure 2 shows two true-positive (TP) examples predicted by ResNet50. They illustrated the typical malignant features of paravertebral soft tissue mass and vertebral bone destruction. Figure 3 shows how the per-slice malignant probability is combined to give per-patient diagnosis. Although the probability of edge slice was much lower, even < 0.5 due to mixed normal findings, when all results were combined and the highest probability was assigned to the patient, the partial volume effect would not affect the final diagnosis. Figure 4 shows two true-negative (TN) cases, illustrating the typical benign features of vertebral compression and bone cortex interruption. Figure 5 shows two false-negative (FN) cases.

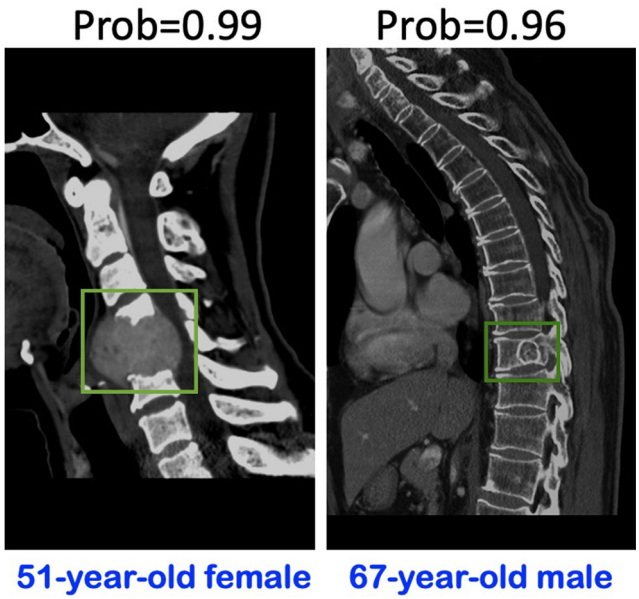

Fig. 2 Two malignant case examples, predicted as malignant by ResNet50 as true positive (TP) cases. The left case shows paravertebral soft tissue mass, and the right case demonstrates vertebral bone destruction, which are typical features of a malignant fracture

When the imaging features were subtle, deep learning was more likely to make wrong predictions, and in these cases, even experienced radiologists could misdiagnose them. Figure 6 shows two false-positive (FP) cases. They presented subtle vertebral compression and bone cortex interruption, and the atypical transverse fracture line might be mistaken as bone destruction, leading to wrong diagnosis. In the overall diagnostic performance made by ResNet50, the per-slice diagnosis showed a sensitivity of 0.90, specificity of 0.79, and accuracy of 85%. In per-patient diagnosis, sensitivity was improved to 0.95, specificity remained similar at 0.80, and the overall accuracy was 88%.

Discussion

Imaging plays a crucial role in the diagnosis of vertebral fracture. CT is an easily accessible and relatively inexpensive modality that can be used to characterize abnormality and further differentiate between benign and malignant fractures. When needed, more expensive secondary imaging such as MRI [26–28] and invasive procedures such as biopsies can be used to obtain a definitive diagnosis. As demonstrated in our reading results, prominent features such as transverse fracture line and the presence of soft tissue mass are very specific and can be used to diagnose benign and malignant fractures, respectively, with a high confidence. When these typical findings are not present, and further, when benign and malignant features are mixed, the diagnosis will be much more challenging. Age is a well-known risk factor for malignancy, and with the increasing prevalence of osteoporosis in the growing elderly population, vertebral fracture imaging presentations will be complicated, making evaluation and differential diagnosis more difficult. Artificial intelligence (AI) technology can provide a feasible diagnostic assistance tool. In this work, we demonstrated that deep learning can be applied to perform

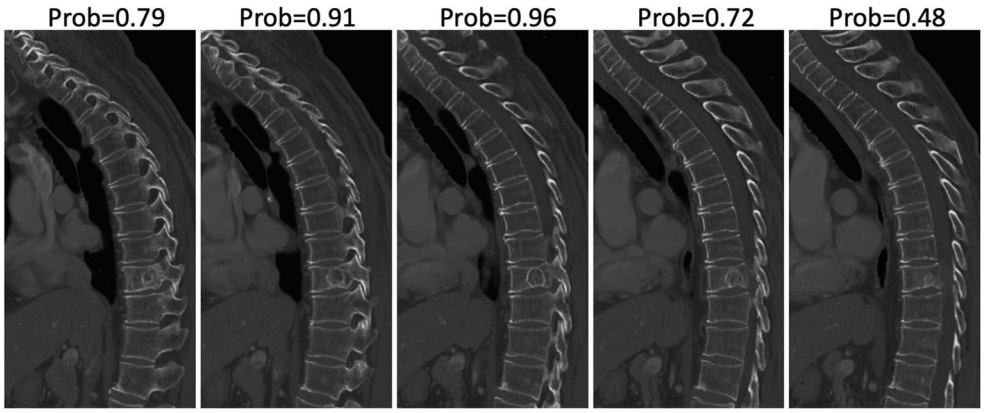

Fig. 3 The malignancy probability predicted by ResNet50 in 5 sagittal slices of the right case shown in Fig. 2. The probability of the edge slice is lower, which is due to the partial volume effect of mixed abnormal and normal findings. The highest probability of 0.96 among the 5 slices is assigned to this patient

differential diagnosis, and achieve a per-patient diagnostic sensitivity of 95%, specificity of 80%, and overall accuracy of 88%.

CT provides detailed morphological information about bone integrity and fracture margin, which plays a role in distinguishing benign and malignant spinal fractures. According to previous works [2, 8, 9], the frequently observed CT findings in benign fractures include cortical fractures of the vertebral body without cortical bone destruction, retropulsion of a bone fragment of the posterior cortex of the vertebral body into the spinal canal, fracture lines within the cancellous bone of the vertebral body, an intravertebral vacuum phenomenon, and a thin diffuse paraspinal soft tissue mass (PSTM). The significant malignant fracture findings include destruction of the anterolateral or posterior cortical bone of the vertebral body, destruction of the cancellous bone of the vertebral body, destruction of a vertebral pedicle, a focal PSTM, and an epidural mass.

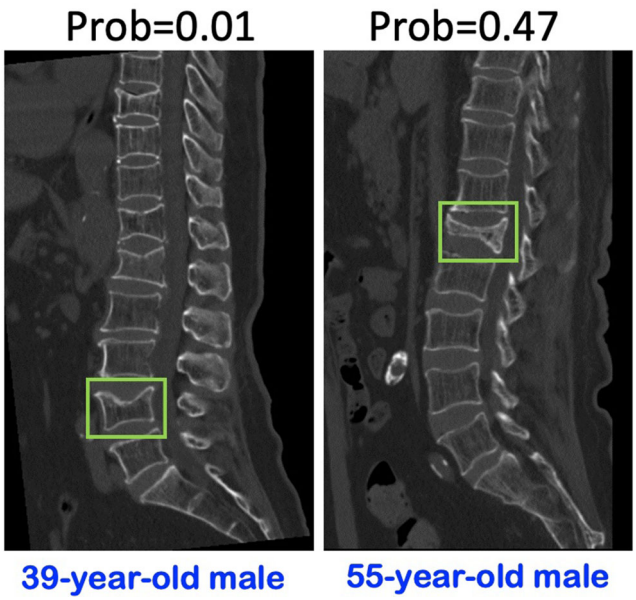

Fig. 4 Two benign case examples, predicted as benign by ResNet50 as true-negative (TN) cases. These two cases show vertebral compression and bone cortex interruption, which are typical features of benign fracture. The left case has obvious fractures in multiple segments

In our results shown in Table 1, soft tissue mass and bone destruction were highly suggestive of malignancy; and the presence of a transverse fracture line was highly suggestive of benign fracture, consistent with reports in the literature.

The radiologists participating in this study were trained in a tertiary hospital dedicated to bone diseases, and their reading accuracy was very high at 93–99%. For radiologists who did not receive specific MSK training or for general physicians, the reading accuracy was expected to be lower. Many companies are developing AI tools for clinical applications, and the performance may be inferior to that of radiologists; however, this does not preclude its clinical use. Instead of claiming the

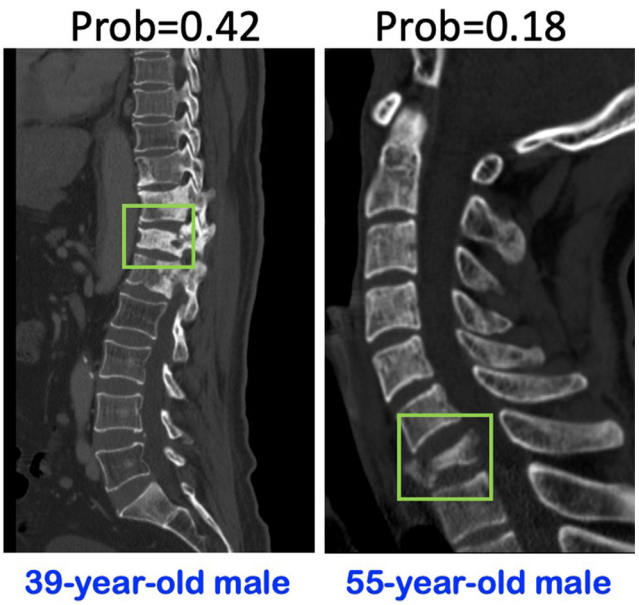

Fig. 5 Two malignant case examples, predicted as benign by ResNet50 as false-negative (FN) cases. The left case shows vertebral compression and vertebral body osteogenic bone destruction in multiple segments, which are features of malignant fracture. ResNet50 predicts a malignancy probability of 0.42 (< 0.5), probably because the involved multiple segments are not considered in the input bounding box. The right case shows vertebral compression and cortex interruption in only one segment, which can be easily misdiagnosed as benign even by experienced radiologists. The ResNet50 probability is also very low at 0.18

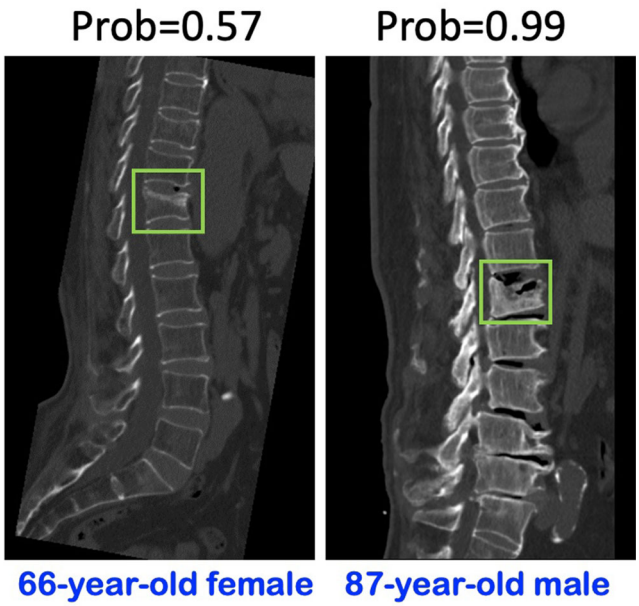

Fig. 6 Two benign case examples, predicted as malignant by ResNet50 as false-positive (FP) cases. These two cases show vertebral compression and bone cortex interruption which are features of benign fracture. However, they are not obvious, and the atypical transverse fracture line may be mistaken as bone destruction. These atypical features may lead to a false prediction by ResNet50

tools can make an accurate diagnosis and marketing them as fully automatic computer-aided diagnostic systems, most Food and Drug Administration (FDA)–cleared AI products were used to assess the probability of malignancy (or abnormality) and flag cases for priority reading. The provided information may also be used by radiologists or other physicians to improve their accuracy and workflow efficiency. The initial accuracy of 88% in our study is an encouraging starting point, and with more training cases and optimized methods, it can be further improved.

Deep learning using a convolutional neural network (CNN) has made an important impact on medical imaging analysis. The algorithms are designed to self-train to recognize important features using the training dataset, and then the features were given appropriate weights by adjusting their internal parameters to build models for testing using a new dataset [29]. Deep learning shows strong processing ability and information retention ability [30, 31], superior to machine learning or radiomics that rely on parameters extracted using pre-defined methods [32]. The cascading architecture of CNN can easily transform the extracted features to high-level features which are more adaptive to the classification applications. Thus, CNN can usually reach a higher accuracy than other machine learning algorithms. The ResNet50 network was used in the present study. Due to its architecture using residual connections, the convergence of the training process can be achieved efficiently compared to other CNN models.

In recent years, deep learning has also become popular in the detection of fractures and differential diagnosis, as summarized in two recent review papers [11, 12]. Kim et al [14] re-trained the top layer of the Inception v3 network using lateral wrist radiographs to produce a model for the classification of new studies as either "fracture" or "no fracture." Chung et al [6] used ResNet-152 for automated detection and classification of the different types of proximal humerus fracture on plain shoulder radiographs. A square ROI was needed as the input into CNN, which was drawn manually to include the humeral head and neck centered and constituted approximately 50% of the square image. Gan et al [7] applied a similar deep learning strategy for the detection of distal radius fractures. Instead of using manually cropped square ROI encasing the whole distal radius as input, a Faster R-CNN network was added as an auxiliary algorithm to train a model for automatic ROI selection, and these inputs were used in CNN with Inception-v4 algorithm for detection of fracture. Compared to the analysis on radiographs, studies on CT were limited. Tomita et al [19] applied ResNet34 to extract features for automatic detection of osteoporotic vertebral fractures on CT. Pranata et al [18] applied ResNet and VGG for automated classification and detection of calcaneus fractures on CT and concluded that ResNet achieved better performance for involving a deeper neural network architecture. Husseini et al [33] proposed a novel grading loss for learning representations that respect Genant's fracture grading scheme. The proposed loss function achieves a fracture detection F1 score of 81.5%, a 10% increase over a naive classification baseline on a publicly available spine dataset. In the present study, the input ROI was selected by a radiologist to focus on one most suspicious abnormal vertebra, which needed to be resolved for developing an automatic diagnostic system. Deep learning has also been successfully applied to perform automatic vertebral segmentation in the spine [20–23]. In the future, it may be possible to use the combination of three different networks to perform a streamlined diagnosis, first to segment all vertebra, second to select abnormal ROI, and last to give classification.

For differentiation of benign vs. malignant spinal fracture, MRI provides more information for the characterization of soft tissue [26, 28, 34]. Li et al [27] developed a scoring system based on the combination of CT and MRI findings to differentiate osteoporotic vs. malignant vertebral fracture and achieved a high accuracy of 98.3%. For MRI, T1- and T2-weighted images are typically utilized, and a similar deep learning diagnostic model may be built by including more images as inputs. For CT, there is only one set of images unlike MRI; therefore, in order to consider more information, we included the slice with its two neighboring slices as inputs. As illustrated in Fig. 3, on the edge slices when the abnormal region was small, the partial volume effect (i.e., mixed abnormal and normal tissues) would decrease the malignant probability and compromise the per-slice diagnostic accuracy. However, when all slices of a patient were considered, only

the highest malignant probability was used, and the edge slices would not affect the final diagnostic results.

This study had several limitations. First, the number of patients included in this study was small. Second, since this was a retrospective study, the selection bias could not be avoided. The malignant cases were more difficult to obtain, and we had to search for a much longer time period to collect them compared to benign cases. Third, to limit variations and potential confounding factors, only patients with metastatic cancer were selected in the malignant group, and the results might not be applicable to other primary bone cancers or lymphoproliferative diseases such as lymphoma and multiple myeloma. Fourth, the slice thickness of 3 mm may be a limitation to this study. Bauer et al [35] found that, while still sufficient for the detection of osteoporotic vertebral fractures (benign), the thinner axial slices could increase reliability. Lastly, as a proof-of-principle preliminary study, we did not include an independent testing dataset. When a new dataset is available, the developed model in this work can be readily applied to test its generalization ability.

In conclusion, this study investigated the application of deep learning for differential diagnosis of benign and malignant vertebral fracture on CT. Deep learning using ResNet50 can yield a high accuracy, but future studies with a larger number of training and testing cases should be performed to further improve the accuracy. In parallel, there are research efforts in performing automatic segmentation of vertebrae in the spine and identifying abnormal areas, which can be integrated with the proposed classification network to develop a streamlined, fully automatic, AI diagnostic tool. Not only deep learning can assist such tool radiologists and other physicians to make accurate diagnoses with higher confidence, but also it can be integrated with the clinical workflow and improve working efficiency.

Acknowledgements We thank Lixiang Gao and Xiaoxi Ji for helping us with visual assessment and scoring.

Funding This study has received funding from the National Natural Science Foundation of China (No. 81971578; 81901791; 81871326); Beijing Natural Science Foundation (No. Z190020); and Peking University International Hospital Research Grant (No. YN2019QN03).

Declarations

Guarantor The scientific guarantor of this publication is Prof. Ning Lang, MD.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- Retrospective
- Diagnostic or prognostic study
- Performed at one institution

References

1. Kendler DL, Bauer DC, Davison KS et al (2016) Vertebral fractures: clinical importance and management. Am J Med 129:221. https://doi.org/10.1016/j.amjmed.2015.09.020
2. Mauch JT, Carr CM, Cloft H, Diehn FE (2018) Review of the imaging features of benign osteoporotic and malignant vertebral compression fractures. AJNR Am J Neuroradiol 39:1584–1592. https://doi.org/10.3174/ajnr.A5528
3. Takigawa T, Tanaka M, Sugimoto Y, Tetsunaga T, Nishida K, Ozaki T (2017) Discrimination between malignant and benign vertebral fractures using magnetic resonance imaging. Asian Spine J 11:478–483. https://doi.org/10.4184/asj.2017.11.3.478
4. Wong CC, McGirt MJ (2013) Vertebral compression fractures: a review of current management and multimodal therapy. J Multidiscip Healthc 6:205–214. https://doi.org/10.2147/JMDH.S31659
5. Schwaiger BJ, Gersing AS, Baum T, Krestan CR, Kirschke JS (2016) Distinguishing benign and malignant vertebral fractures using CT and MRI. Semin Musculoskelet Radiol 20:345–352. https://doi.org/10.1055/s-0036-1592433
6. Chung SW, Han SS, Lee JW et al (2018) Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop 89:468–473. https://doi.org/10.1080/17453674.2018.1453714
7. Gan K, Xu D, Lin Y et al (2019) Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. Acta Orthop 90:394–400. https://doi.org/10.1080/17453674.2019.1600125
8. Laredo JD, Lakhdari K, Bellaiche L, B Hamze, Janklewicz P, Tubiana JM (1995) Acute vertebral collapse: CT findings in benign and malignant nontraumatic cases. Radiology 194:41–48. https://doi.org/10.1148/radiology.194.1.7997579
9. Kubota T, Yamada K, Ito H, Kizu O, Nishimura T (2005) High-resolution imaging of the spine using multidetector-row computed tomography: differentiation between benign and malignant vertebral compression fractures. J Comput Assist Tomogr 29:712–719. https://doi.org/10.1097/01.rct.0000175500.41836.24
10. Lv M, Zhou Z, Tang Q et al (2020) Differentiation of usual vertebral compression fractures using CT histogram analysis as quantitative biomarkers: a proof-of-principle study. Eur J Radiol 131:109264. https://doi.org/10.1016/j.ejrad.2020.109264
11. Yang S, Yin B, Cao W, Feng C, Fan G, He S (2020) Diagnostic accuracy of deep learning in orthopaedic fractures: a systematic review and meta-analysis. Clin Radiol 75:713–717. https://doi.org/10.1016/j.crad.2020.05.021
12. Kalmet P, Sanduleanu S, Primakov S et al (2020) Deep learning in fracture detection: a narrative review. Acta Orthop 91:215–220. https://doi.org/10.1080/17453674.2019.1711323
13. Olczak J, Fahlberg N, Maki A et al (2017) Artificial intelligence for analyzing orthopedic trauma radiographs. Acta Orthop 88:581–586. https://doi.org/10.1080/17453674.2017.1344459

14. Kim DH, MacKinnon T (2018) Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol 73:439–445. https://doi.org/10.1016/j.crad.2017.11.015
15. Adams M, Chen W, Holcdorf D, McCusker MW, Dl Howe P, Gaillard F (2019) Computer vs human: deep learning versus perceptual training for the detection of neck of femur fractures. J Med Imaging Radiat Oncol 63:27–32. https://doi.org/10.1111/1754-9485.12828
16. Badgeley MA, Zech JR, Oakden-Rayner L et al (2019) Deep learning predicts hip fracture using confounding patient and healthcare variables. NPJ Digit Med 2:31. https://doi.org/10.1038/s41746-019-0105-1
17. Raghavendra, S UBN, Gudigar A, Rajendra Acharya U (2018) Automated system for the detection of thoracolumbar fracture using a CNN architecture. Future Generation Computer Systems: S167739X-S17321544X
18. Pranata YD, Wang KC, Wang JC et al (2019) Deep learning and SURF for automated classification and detection of calcaneus fractures in CT images. Comput Methods Programs Biomed 171:27–37. https://doi.org/10.1016/j.cmpb.2019.02.006
19. Tomita N, Cheung YY, Hassanpour S (2018) Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. Comput Biol Med 98:8–15. https://doi.org/10.1016/j.compbiomed.2018.05.011
20. Chu C, Belavý DL, Armbrecht G, Bansmann M, Felsenberg D, Zheng G (2015) Fully automatic localization and segmentation of 3D vertebral bodies from CT/MR images via a learning-based method. PLoS One 10:e143327. https://doi.org/10.1371/journal.pone.0143327
21. Lessmann N, van Ginneken B, de Jong PA, Išgum I (2019) Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. Med Image Anal 53:142–155. https://doi.org/10.1016/j.media.2019.02.005
22. Lffler MT, Sekuboyina A, Jacob A et al (2020) A vertebral segmentation dataset with fracture grading. Radiology: Artificial Intelligence. https://doi.org/10.1148/ryai.2020190138
23. Sekuboyina A, Bayat A, Husseini M et al (2020) VerSe: A vertebrae labelling and segmentation benchmark for multi-detector CT images. https://arxiv.org/abs/2001.09193
24. Fang Y, Li W, Chen X et al (2021) Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks. Eur Radiol 31(4):1831–1842. https://doi.org/10.1007/s00330-020-07312-8
25. Kingma DP BJ (2015) Adam: a method for stochastic optimization. International Conference on Learning Representations: International Conference on Learning Representations, 1–13
26. Torres C, Hammond I (2016) Computed tomography and magnetic resonance imaging in the differentiation of osteoporotic fractures from neoplastic metastatic fractures. J Clin Densitom 19:63–69. https://doi.org/10.1016/j.jocd.2015.08.008
27. Li Z, Guan M, Sun D, Xu Y, Li F, Xiong W (2018) A novel MRI- and CT-based scoring system to differentiate malignant from osteoporotic vertebral fractures in Chinese patients. BMC Musculoskelet Disord 19:406. https://doi.org/10.1186/s12891-018-2331-0
28. Romeo V, Ugga L, Stanzione A, Cocozza S, Cuocolo R, Brunetti A (2019) Differential diagnosis of benign and malignant vertebral compression fractures using conventional and advanced MRI techniques. BJR Open 1:20180033. https://doi.org/10.1259/bjro.20180033
29. Korbar B, Olofson AM, Miraflor AP et al (2017) Deep learning for classification of colorectal polyps on whole-slide images. J Pathol Inform 8:30. https://doi.org/10.4103/jpi.jpi_34_17
30. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003
31. Lee JG, Jun S, Cho YW et al (2017) Deep learning in medical imaging: general overview. Korean J Radiol 18:570–584. https://doi.org/10.3348/kjr.2017.18.4.570
32. LeCun Y KKFC (2010) Convolutional networks and applications in vision. International Symposium on Circuits and Systems (ISCAS 2010): 253–256
33. Husseini M, Sekuboyina A, Loeffler M, Navarro F, Menze BH, Kirschke JS (2020) Grading loss: a fracture grade-based metric loss for vertebral fracture detection. Medical Image Computing and Computer Assisted Intervention (MICCAI 2020):733–742
34. Wang KC, Jeanmenne A, Weber GM, Thawait SK, Carrino JA (2011) An online evidence-based decision support system for distinguishing benign from malignant vertebral compression fractures by magnetic resonance imaging feature analysis. J Digit Imaging 24:507–515. https://doi.org/10.1007/s10278-010-9316-3
35. Bauer JS, Müller D, Ambeka A (2006) Detection of osteoporotic vertebral fractures using multidetector CT. Osteoporos Int 17(4):608–615. https://doi.org/10.1007/s00198-005-0023-8