

RESEARCH

Open Access



# Integrating manual annotation with deep transfer learning and radiomics for vertebral fracture analysis

Jing Wang<sup>1†</sup>, Zhirui Dong<sup>1†</sup>, Huanxin He<sup>1†</sup>, Zhiyang Gao<sup>1</sup>, Yukai Huang<sup>1</sup>, Guangcheng Yuan<sup>2</sup>, Libo Jiang<sup>2\*</sup> and Mingdong Zhao<sup>1\*</sup>

## Abstract

**Background** Vertebral compression fractures (VCFs) are prevalent in the elderly, often caused by osteoporosis or trauma. Differentiating acute from chronic VCFs is vital for treatment planning, but MRI, the gold standard, is inaccessible for some. However, CT, a more accessible alternative, lacks precision. This study aimed to enhance CT's diagnostic accuracy for VCFs using deep transfer learning (DTL) and radiomics.

**Methods** We retrospectively analyzed 218 VCF patients scanned with CT and MRI within 3 days from Oct 2022 to Feb 2024. MRI categorized VCFs. 3D regions of interest (ROIs) from CT scans underwent feature extraction and DTL modeling. Receiver operating characteristic (ROC) analysis evaluated models, with the best fused with radiomic features via LASSO. AUCs compared via Delong test, and clinical utility assessed by decision curve analysis (DCA).

**Results** Patients were split into training (175) and test (43) sets. Traditional radiomics with LR yielded AUCs of 0.973 (training) and 0.869 (test). Optimal DTL modeling improved to 0.992 (training) and 0.941 (test). Feature fusion further boosted AUCs to 1.000 (training) and 0.964 (test). DCA validated its clinical significance.

**Conclusion** The feature fusion model enhances the differential diagnosis of acute and chronic VCFs, outperforming single-model approaches and offering a valuable decision-support tool for patients unable to undergo spinal MRI.

**Keywords** Deep transfer learning, Vertebral compression, Automatic segmentation, Fractures, Radiomics

## Background

Vertebral compression fractures (VCFs) are common and highly disabling injuries [1–3], and their incidence is becoming a significant global public health concern that cannot be ignored. Both acute and chronic VCFs are crucial factors that must be considered when determining whether to opt for conservative or surgical treatment [4–6]. In clinical practice, magnetic resonance imaging (MRI) is crucial for diagnosing acute fractures, due to its superior ability to visualize bone marrow edema (BME) in VCFs [7]. However, its high cost, lengthy scanning time, and numerous contraindications can restrict its use in certain populations. Computed tomography

<sup>†</sup>Jing Wang, Zhirui Dong and Huanxin He contributed equally to this work.

\*Correspondence:

Libo Jiang

jianglibo@zs-hospital.sh.cn

Mingdong Zhao

zhao\_mingdong@fudan.edu.cn

<sup>1</sup>Department of Orthopaedic Surgery, Jinshan Hospital, Fudan University, Shanghai, China

<sup>2</sup>Department of Orthopaedic Surgery, Zhongshan Hospital, Fudan University, Shanghai, China



(CT), while more accessible and efficient, struggles with detecting small fractures or lesions in areas of degenerative bone sclerosis, leading to missed diagnoses or delays in treatment.

Radiomics, a discipline dedicated to extracting numerous quantitative features from medical images, enables further analysis of image details invisible to the naked eye through sophisticated algorithmic models [8]. This field has proven valuable in assessing the microstructural changes of trabecular bone [9]. Radiomics, which leverages CT imaging and machine learning, has demonstrated significant potential in effectively assessing both acute and chronic VCFs and in distinguishing their benign or malignant nature [10–12]. This method relies on predefined features rather than automatic learning from the data. Additionally, the subjectivity of medical professionals can influence feature extraction and selection. In contrast, deep transfer learning [13] (DTL) offers an innovative approach to automatically learning complex features from highly nonlinear and complex medical imaging data. This strategy enables the application of deep learning features to smaller datasets by fine-tuning pretrained deep learning networks to adapt to new tasks, becoming a research hotspot in recent years [14–16]. Despite promising advances in both fields, the integration of these techniques for diagnosing acute VCFs remains underexplored, particularly regarding the optimization of cropping strategies and the utilization of indirect diagnostic indicators. These indicators, such as soft tissue swelling and blurred fat planes, play a critical role in fracture diagnosis alongside directly visible fracture lines. This type of diagnosis can draw on research methods used to study tumor microenvironments [17–19].

For instance, the approach involves using various cropping techniques on 3D regions of interest (ROIs) and integrating DTL to extract and model features of these areas. By analyzing the effectiveness of these feature models, we can better understand the benefits of different cropping strategies, revealing potential methods to enhance diagnostic precision.

In this study, we aimed to process vertebral compression fracture (VCF) medical images with high precision and reliability. To achieve this, we adopted manual segmentation for feature extraction, which allowed for more accurate and consistent data preparation. This approach, combined with advanced deep transfer learning (DTL) and radiomics analyses, was key to achieving optimal experimental results. This study aims to:

1. Develop and evaluate a diagnostic model that combines radiomics and DTL to analyze CT images for acute and chronic VCF differentiation.
2. Assess the impact of various cropping strategies on feature extraction and model performance.

3. Validate the clinical utility of the proposed framework in enhancing diagnostic precision and facilitating treatment decisions.

## Methods

### Study population

After a review by the hospital's Institutional Review Board, the requirement for informed consent was waived. The primary cohort of this study comprised medical records retrieved from the institutional picture archiving and communication systems (PACS) database, spanning from October 2022 to February 2024.

The inclusion criteria for the present study were as follows: ① diagnosis of benign VCFs, including traumatic or osteoporotic fractures; and ② complete original imaging data, including CT and MRI scans of the vertebral bodies, with a time interval of no more than three days between the two examinations. The exclusion criteria were as follows: ① suspected infection, tumor-related pathological fractures, or history of spinal surgery, ② poor image quality or the presence of foreign bodies, and ③ patients with incomplete clinical records. The detailed screening process and workflow of the feature fusion model are illustrated in Figs. 1 and 2, respectively.

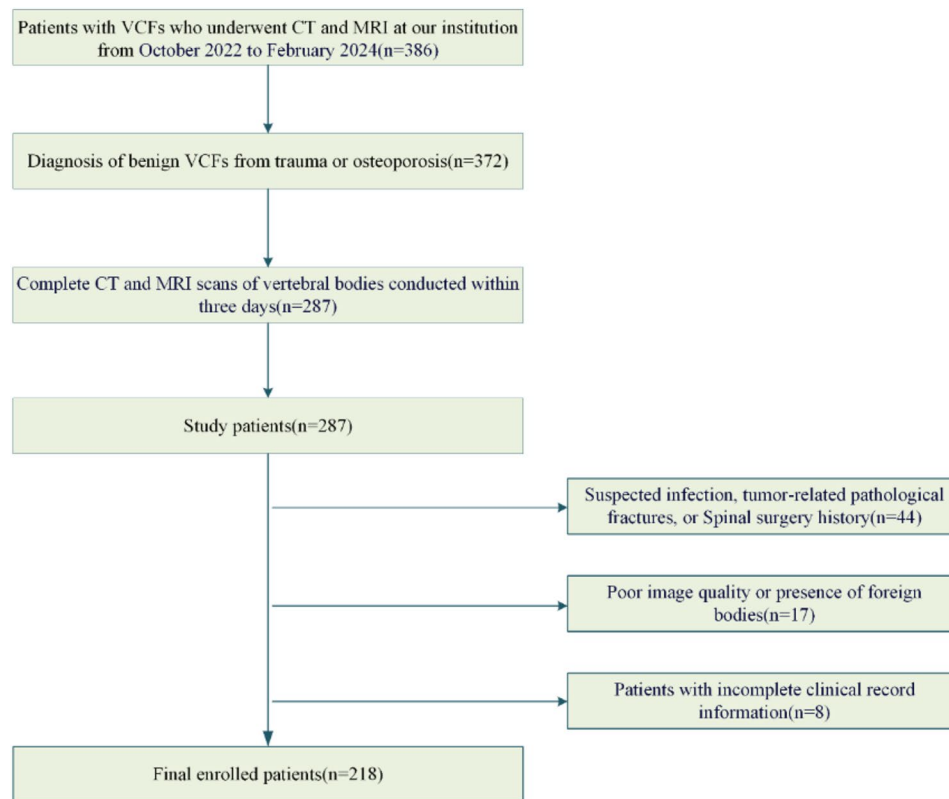
Finally, we retrospectively analyzed 218 patients admitted to our hospital with thoracic and lumbar compression fractures. Patients who fulfilled the inclusion criteria were randomly assigned to a training cohort ( $n = 175$ ) or a validation cohort ( $n = 43$ ).

### CT examination and clinical baseline characteristics

All patients' age, sex, and bone mineral density data were collected from the medical records. All CT and MR images were retrieved from PACS. These image files were saved in the Digital Imaging and Communications in Medicine (DICOM) format. All images obtained using Siemens CT scanners (Siemens Healthineers, Germany) were reconstructed using bone window settings and sagittal plane images with a slice thickness of 1 mm were produced for subsequent detailed analysis. Acute and chronic VCFs are differentiated by the presence of BMEs, assessed using fat-suppressed T2WI, STIR, or TIRM sequences, where the presence of high-signal linear or nodular regions in these sequences signifies the acute nature of the fracture [20, 21]. Figure 3A-B show MR images (T2-weighted) of chronic and acute fractures, respectively.

### Image segmentation and preprocessing

Accurate segmentation of fractured vertebrae is highly important for image analysis. First, interpolating the original images to a uniform voxel spacing of 1.0 mm × 1.0 mm × 1.0 mm is necessary. All images, with a slice thickness of 1.0 mm, were reconstructed utilizing a



**Fig. 1** Patient selection process for VCFs study, October 2022 to February 2024

bone window setting of 1200 for width and 500 for window level. Subsequently, these reconstructed images were processed and thoroughly analyzed based on the acquired bone window settings. In this study, radiologists with ten years of extensive skeletal and muscle imaging expertise meticulously performed manual segmentation.

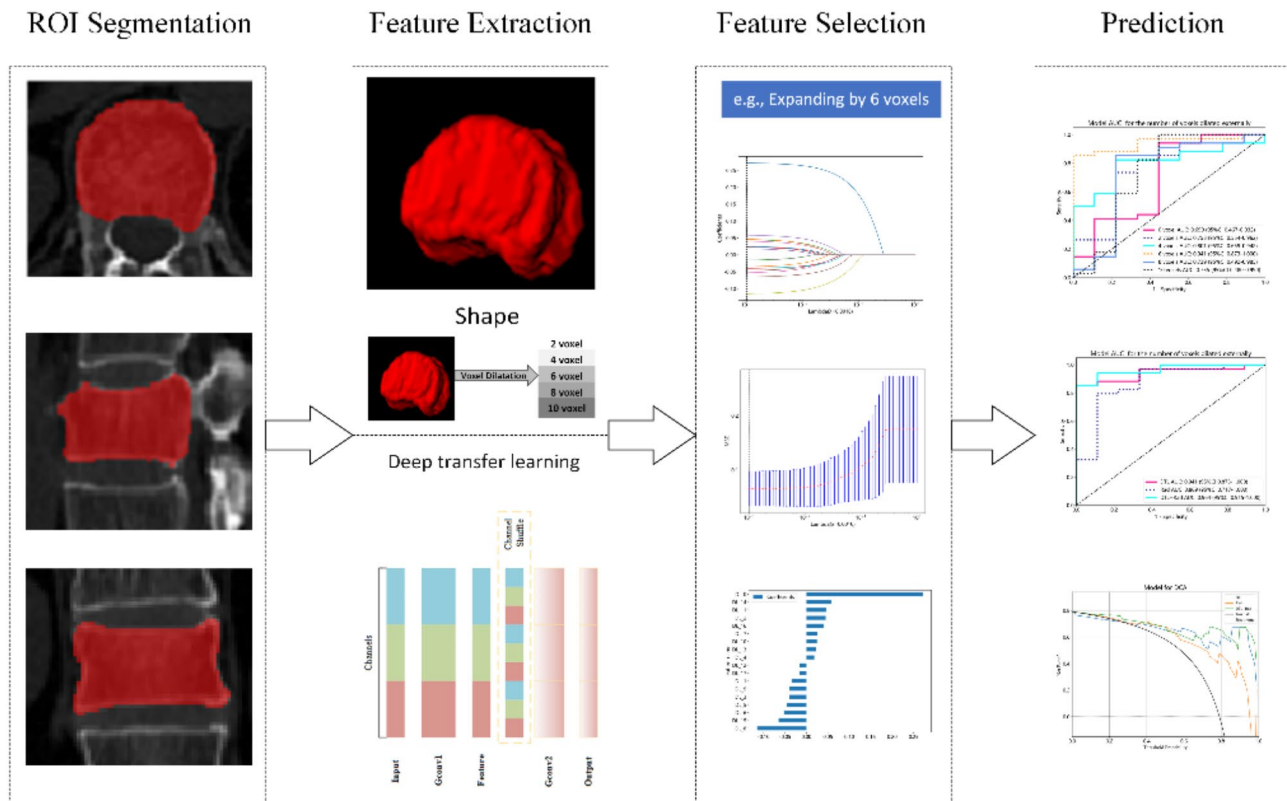
Initially, Radiologist A imported the CT images into the ITK-snap software (version 3.8.0) for three-dimensional visualization. With the utmost precision, the edges of the fractured vertebrae were carefully identified and manually delineated in the sagittal images. This process entailed meticulous excluding adjacent structures such as intervertebral discs, pedicles, and adipose tissues. This thorough delineation was repeated for each image layer, ensuring a comprehensive and accurate representation of the fractured vertebral region. One week later, a representative subset of 30 patients from the training sequence was randomly selected for reproducibility assessment. Radiologist A and B independently resketched the patients' vertebrae. Intraclass correlation coefficients (ICCs) were meticulously calculated to rigorously evaluate the consistency of vertebral delineation within and between observers.

### Feature extraction

PyRadiomics (<http://pypi.org/project/pyradiomics/>), based on the Python 3.6 platform, was utilized to conduct traditional radiomics feature extraction. This process involved extracting 1834 radiological features from CT images, encompassing first-order statistics, shape features, and texture features derived from various matrices, such as the Grey-Level Co-occurrence Matrix (GLCM), Grey-Level Size Zone Matrix (GLSZM), Grey-Level Run Length Matrix (GLRLM), Neighbouring Grey-Tone Difference Matrix (NGTDM), and Grey-Level Dependence Matrix (GLDM).

The extraction process incorporated the following parameters and filters:

- **Laplacian of Gaussian (LoG):** Applied with sigma values of 1.0, 2.0, and 3.0 to emphasize edges and fine structures.
- **Wavelet:** Decomposed using standard wavelet filters to capture multiscale texture information.
- **LBP3D (Local Binary Patterns in 3D):** Applied to capture texture information in three-dimensional space.
- **Transformations:** Exponential, Square, SquareRoot, Logarithm, and Gradient filters were employed to enhance image details and detect subtle patterns.



**Fig. 2** Workflow depicting the integration of deep transfer learning for feature extraction from ROI segmentation, feature selection, and predictive modeling, combined with traditional radiomics

- **Binwidth:** The binwidth for discretizing image intensity values was set to 5.

A detailed description of these radiomics features and their specific settings can be found in the PyRadiomics documentation (<http://pyradiomics.readthedocs.io>). In addition to traditional radiomics features, we also employed a DTL methodology for feature extraction in this study, incorporating the following pivotal steps:

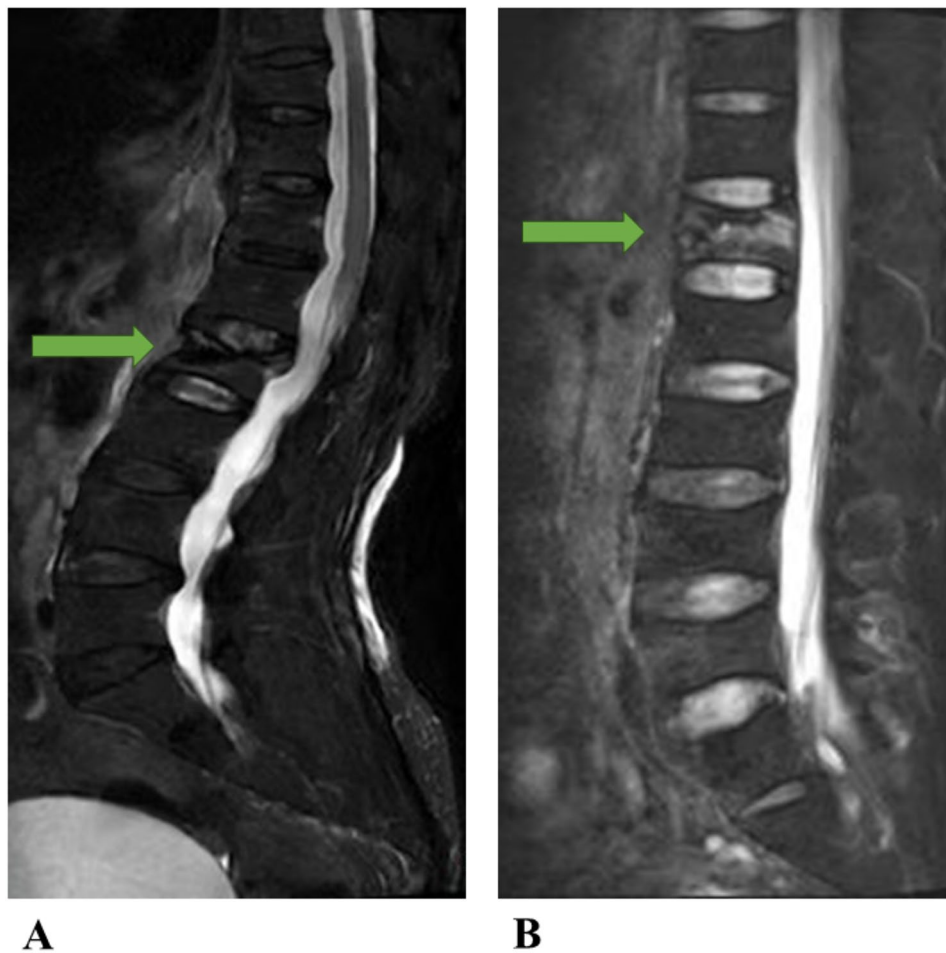
1. **Image Dataset Cropping:** We applied four specific cropping techniques to focus on the ROI: the smallest 3D ROI, and expansions of this region by 2,4,6,8 and 10 voxels. This precision in image preparation is foundational for targeted feature extraction.
2. **Model Selection and Optimization:** Regarding the training of the models, we trained six separate models, each corresponding to one of the cropped ROI cubes at different expansion levels. These models were pre-trained on ImageNet using ShuffleNet [22] (Fig. 4), a lightweight and efficient convolutional neural network (CNN) architecture, known for its excellent balance between performance and computational efficiency. This architecture was then fine-tuned on our dataset, serving as the

backbone for feature extraction. The models were initialized with a learning rate of 0.001, and ‘Adam’ optimization was used to optimize performance, striking a balance between accuracy and computational resource consumption. Additionally, since we used manually delineated masks to define the training regions, the models were trained exclusively within these labeled areas, eliminating the risk of learning irrelevant features from the surrounding non-target regions.

3. **Feature Transfer and Dimensionality Reduction:** A strategic selection of the model’s layer (layer3.3.conv3) for feature extraction is followed by reducing these features to an 18-dimensional space. This approach aims to enhance model generalization, minimizing overfitting risks, and maintaining a harmonious balance of extracted features.

**Feature selection and fusion**

The extracted features were first Z-score normalized, with the mean and standard deviation calculated using only the training set data to prevent data leakage. The transformation was then applied to the test dataset. To address feature redundancy, the Pearson correlation coefficient was computed, and any pair of features with



**Fig. 3** **A** This patient has experienced persistent lower back pain for a month, exacerbated recently by a lumbar sprain. MRI reveals extensive low signals in L2 vertebra, with medium to high signals at the fracture site, and high signals in subcutaneous tissues, diagnosed as chronic L2 fracture with subcutaneous contusion. **B** This patient was admitted for treatment 6 h after experiencing lower back pain due to a vehicular accident. MRI revealed speckled high signals in the L1 vertebra, leading to a diagnosis of an acute fracture in L1. (X-rays and CT scans indicated lumbarization of the S1 vertebra)

a correlation coefficient greater than 0.9 was reduced to a single feature. This step ensured that only features with good reproducibility and low redundancy were retained.

To assess the reproducibility of the extracted features, we calculated the intraclass correlation coefficient (ICC) for both interobserver and intraobserver agreement. Features with ICC values greater than 0.8 were considered stable and reliable for further analysis.

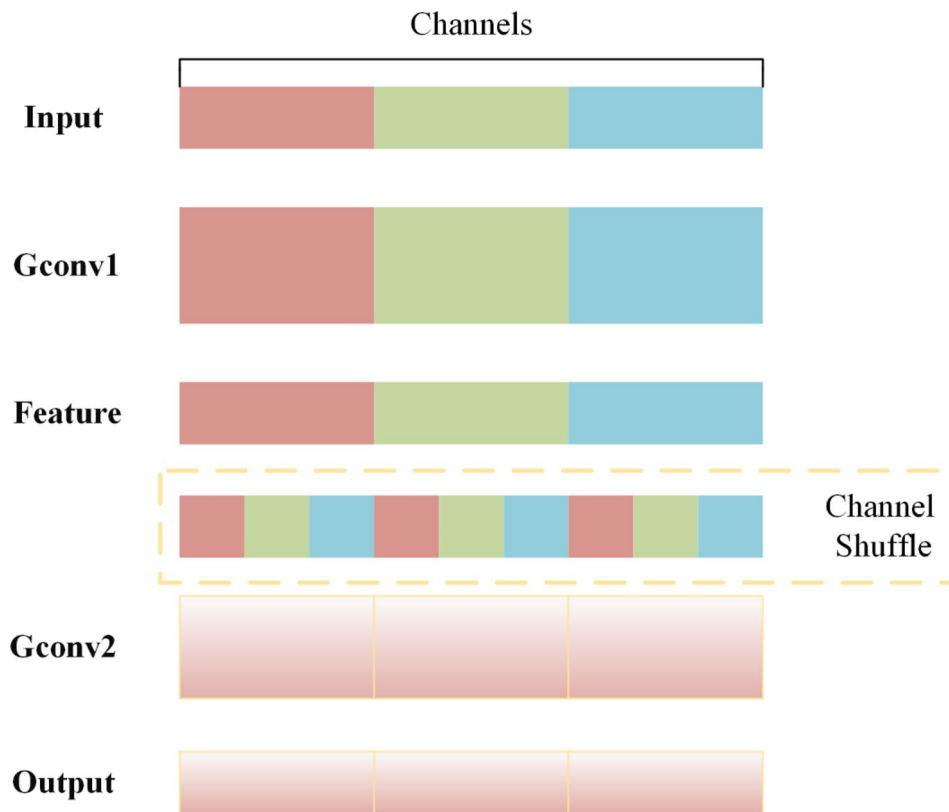
Feature selection was then performed using a multivariate logistic regression (LR) model in a wrapper-based approach, followed by further refinement using the least absolute shrinkage and selection operator (LASSO). LASSO was used to enhance the representativeness of the selected features by reducing overfitting and retaining only the most informative features. To determine the optimal value of the regularization parameter ( $\lambda$ ), we employed grid search with cross-validation. After LASSO, correlation analysis was conducted again to eliminate any remaining redundant features, removing

those that exhibited nonzero correlations with other features.

Finally, only the features with nonlinear relationships, as determined by the multivariate LR model, were retained and subjected to early fusion. The radiomic features and the features extracted via deep transfer learning (DTL) were then combined to create a comprehensive feature set for subsequent modeling.

#### Model construction and validation

After extracting high-quality features, logistic regression (LR) was used for modeling. We employ a grid search to meticulously explore the parameter space of the model to optimize model performance and determine the best combination of parameters. Finally, we leveraged the receiver operating characteristic (ROC) curve and area under the curve (AUC) value as evaluation metrics to assess the classification capabilities of our chosen model. The ROC curve provides a visual representation of the



**Fig. 4** This figure shows the ShuffleNet process from input through two grouped convolutions (Gconv1 and Gconv2) separated by a channel shuffle, highlighting efficient feature mixing and processing across channel groups

performance of models across varying classification thresholds. The AUC provides a quantitative measure of the model's overall performance. These two indicators gave us added confidence in the selected model's excellent classification abilities, ensuring its effectiveness in practical applications.

#### Statistical analysis

1. All the statistical tests were performed using R software version 4.2.2.
2. The clinical utility of the decision curve analysis (DCA) was assessed.
3. The Delong test was used to compare the AUCs of various prediction models.
4. First, the Shapiro–Wilk test was used to check for a normal distribution. If the data were normally distributed, a t test was performed for continuous variables; otherwise, the Mann–Whitney U test was used. For discrete variables, the chi-square test was used. Significance was set at  $P < 0.05$ .

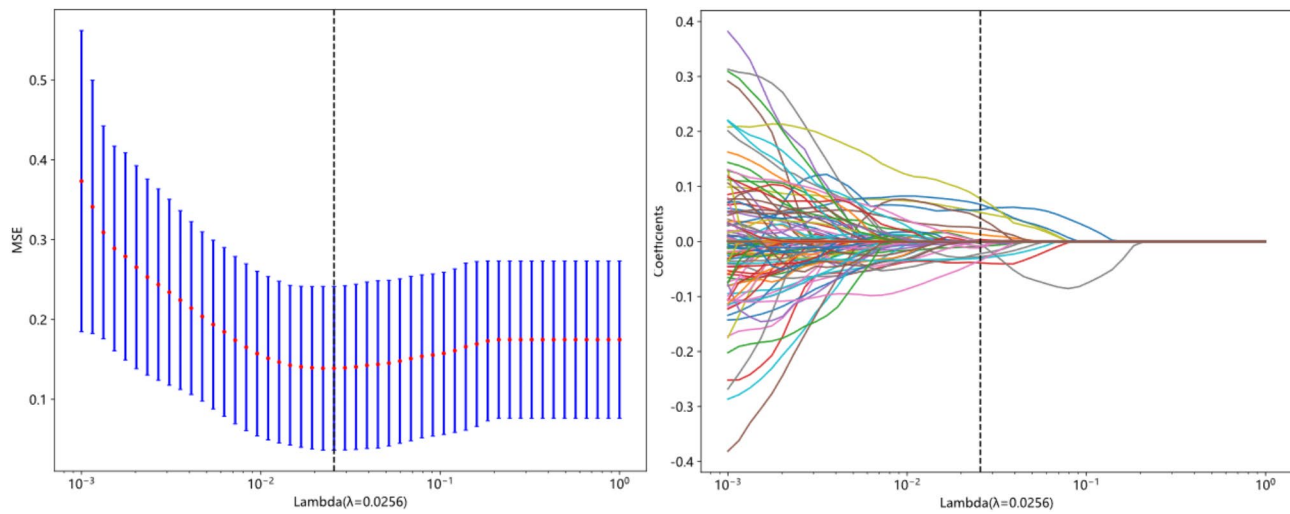
**Table 1** Baseline clinical profiles of patients in the training and test groups

|                | Train( $n = 175$ ) | Test( $n = 43$ ) | $p$ -value |
|----------------|--------------------|------------------|------------|
| Age(years)     | 63.86 ± 15.26      | 60.21 ± 17.37    | 0.33       |
| BMD            | -2.30 ± 1.27       | -2.20 ± 1.22     | 0.67       |
| Females/Males  | 95/80              | 24/19            | 0.99       |
| Acute/ Chronic | 137/38             | 34/9             | 0.87       |

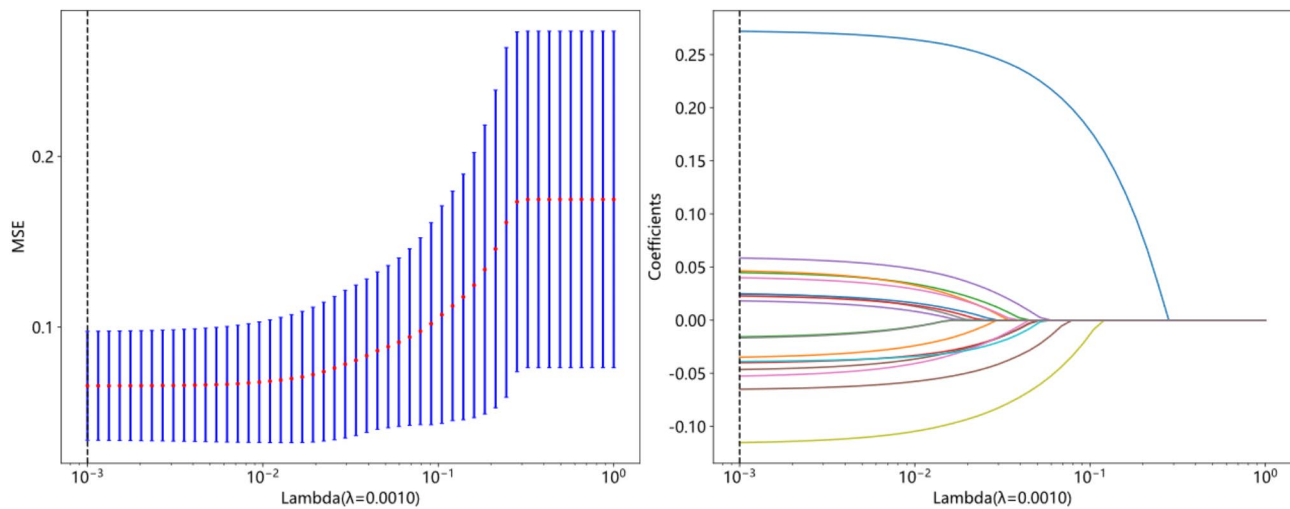
#### Results

##### Clinical baseline characteristics

In this study, we analyzed a cohort of 218 patients, comprising 109 males and 119 females, aged between 21 and 97 years, with a mean age of 63.14 years and a standard deviation of 15.72 years. The cohort included 171 patients with acute vertebral compression fractures (VCFs) and 47 patients with chronic VCFs. For the training set, there were 137 acute and 38 chronic VCF cases, while the test set consisted of 34 acute and 9 chronic VCF cases. The baseline characteristics are summarized in Table 1, which shows the profiles for both the training and test groups. There were no significant differences in age, bone mineral density (BMD), sex distribution, or the proportion of acute versus chronic fractures between the training and test groups, as confirmed by statistical tests ( $p > 0.05$ ).



**Fig. 5** The optimal  $\lambda$  value of 0.0256 enabled radiomic features using the least absolute shrinkage method, followed by generating a histogram of features importance scores based on the selected features



**Fig. 6** The optimal  $\lambda$  value of 0.0010 enabled DTL features (expansion of the 3D ROI by 6 voxels) using the least absolute shrinkage method, followed by generating a histogram of features importance scores based on the selected features

### Feature selection

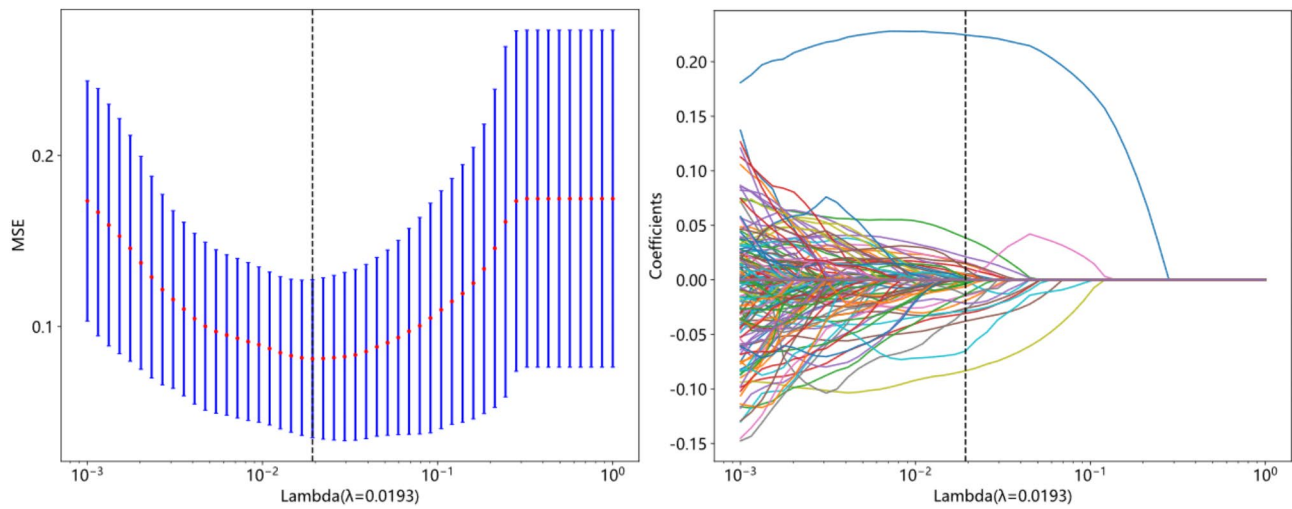
LASSO regression analysis was used to reduce the dimensionality of the radiomic features. Selecting the penalty coefficient ( $\lambda = 0.0256$ ), feature selection, and the curve of the variation in the feature coefficient with  $\lambda$  are shown in Fig. 5. A total of 28 features were selected after the final screening.

Dimensionality reduction in DTL features was performed using LASSO regression analysis, reducing the feature count to 18. The penalty coefficient ( $\lambda = 0.0010$ ), which demonstrated the best performance characterized by the minimal expansion of the 3D ROI by 6 voxels, was selected. The feature selection process and the curve illustrating the variation in the feature coefficients with respect to  $\lambda$  are shown in Fig. 6.

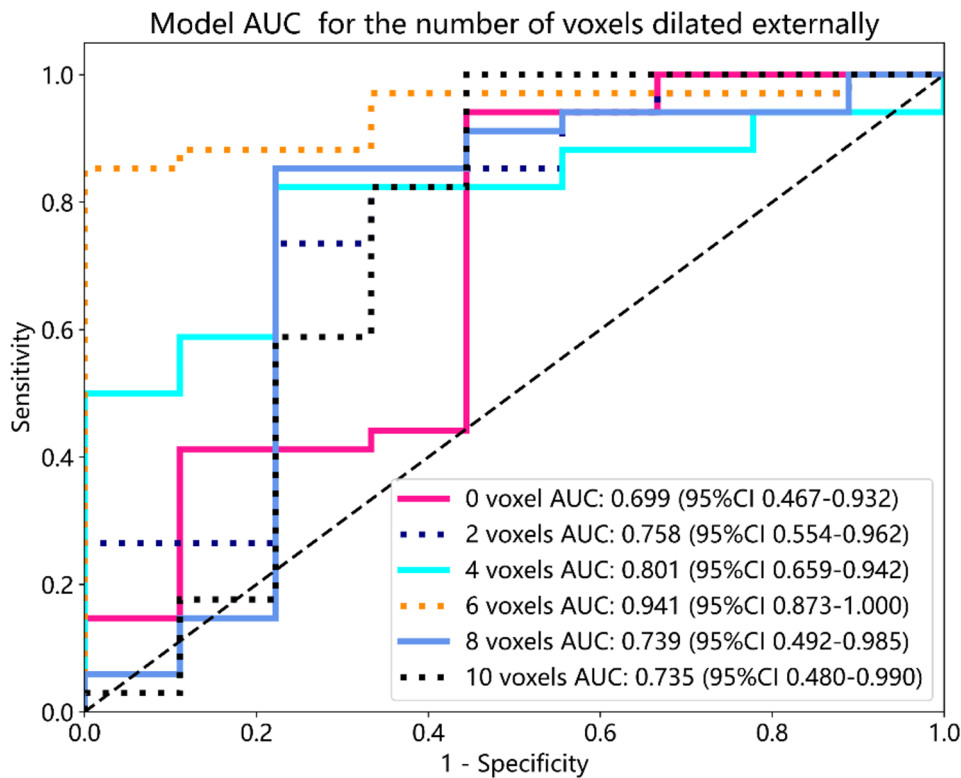
LASSO regression was used to perform dimensionality reduction in feature fusion. Selecting the penalty coefficient ( $\lambda = 0.0193$ ), the feature screening process, and the graph of the variation in the feature coefficient with  $\lambda$  are shown in Fig. 7. After the final screening of the feature fusion, 18 radiomic features and 15 DTL features were retained.

### The predictive performance of the model

In our research, we thoroughly explored various methodologies for processing ROIs and extracting sophisticated deep learning features through advanced DTL techniques. Our empirical results clearly demonstrate that the most effective approach entails meticulously cropping the ROI to its minimal bounding cube and enlarging it by 6 voxels, which optimally prepares it for feature



**Fig. 7** The optimal  $\lambda$  value of 0.0193 enabled features fusion using the least absolute shrinkage method, followed by generating a histogram of features importance scores based on the selected features



**Fig. 8** The AUCs of DTL prediction models in the test cohort

extraction. We successfully achieved a remarkable training set AUC of 0.992 (95% CI, 0.9828-1.0000) and a test set AUC of 0.941 (95% CI, 0.8727-1.0000) by employing LR as our modeling framework. The AUCs of all DTL prediction models in the test cohort are presented in Fig. 8. (The remaining details are in supplement.1)

Additionally, in traditional radiomics, we also employed the widely acknowledged LR model. This

model demonstrated a high level of predictive stability, with an AUC of 0.973 (95% CI, 0.9525–0.9933) in the training cohort. In the testing cohort, it achieved an AUC of 0.869 (95% CI, 0.7670-1.0000), validating its robust predictive ability.

Finally, we integrated traditional omics features with DTL features extracted after minimal ROI expansion of 6 voxels and modeled them using LR, which has emerged



as the optimal approach. The training set yielded an AUC of 1.000 (95% CI, 1.000–1.0000), while the testing set yielded an AUC of 0.964 (95% CI, 0.9149–1.0000) (Fig. 9). Although the Delong test did not reveal significant differences among the three models, DCA demonstrated that the feature fusion prediction model was optimal (Fig. 10).

## Discussion

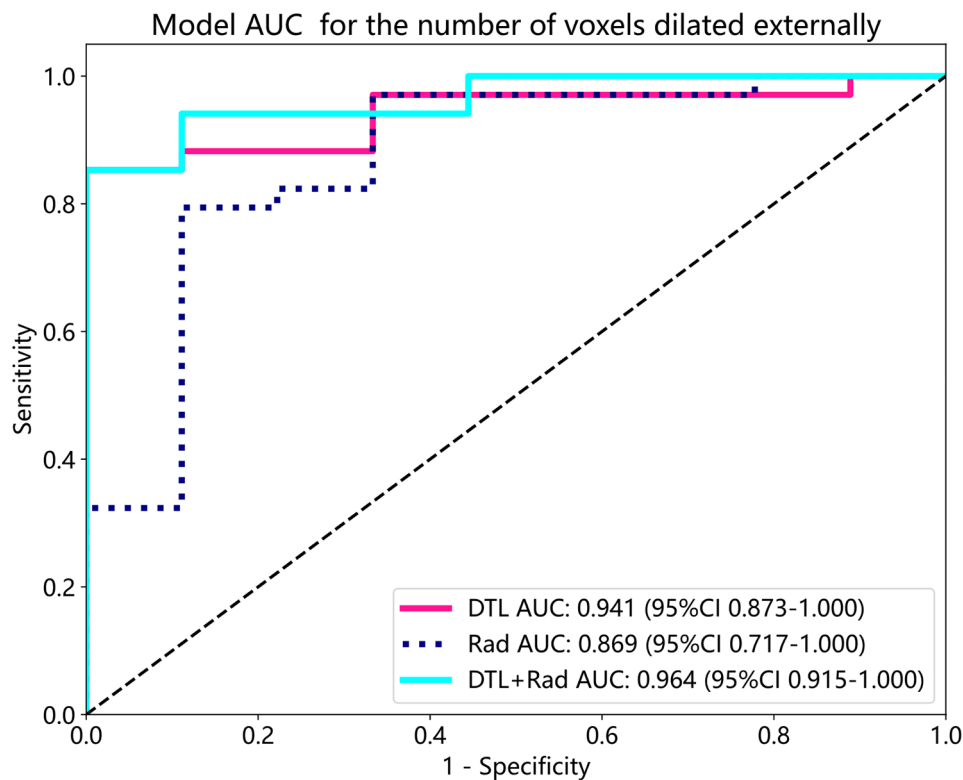
In a socioeconomically evolving society and an aging population, the increasing costs of treating VCFs present a formidable challenge [23]. This is particularly evident in minor trauma incidents, where middle-aged and older adults are frequently screened for suspected VCFs. Assessing the freshness of fractures is critical for determining liability in accidents. Reliance on costly MRI scans can delay diagnosis, treatment, and the assignment of responsibility. Physicians also lack confidence in diagnosing acute VCFs solely with CT imaging. Therefore, it is essential to identify a convenient and cost-effective method to differentiate between acute and chronic VCFs.

In previous studies, dual-energy CT was utilized to evaluate BMEs in VCFs [20, 21, 24] and achieved promising outcomes. However, the high cost and operational complexity of dual-energy CT technology have impeded its widespread adoption in primary healthcare facilities. Additionally, traditional radiomic features derived from CT images were used to predict acute and chronic VCFs [10, 12]. This method

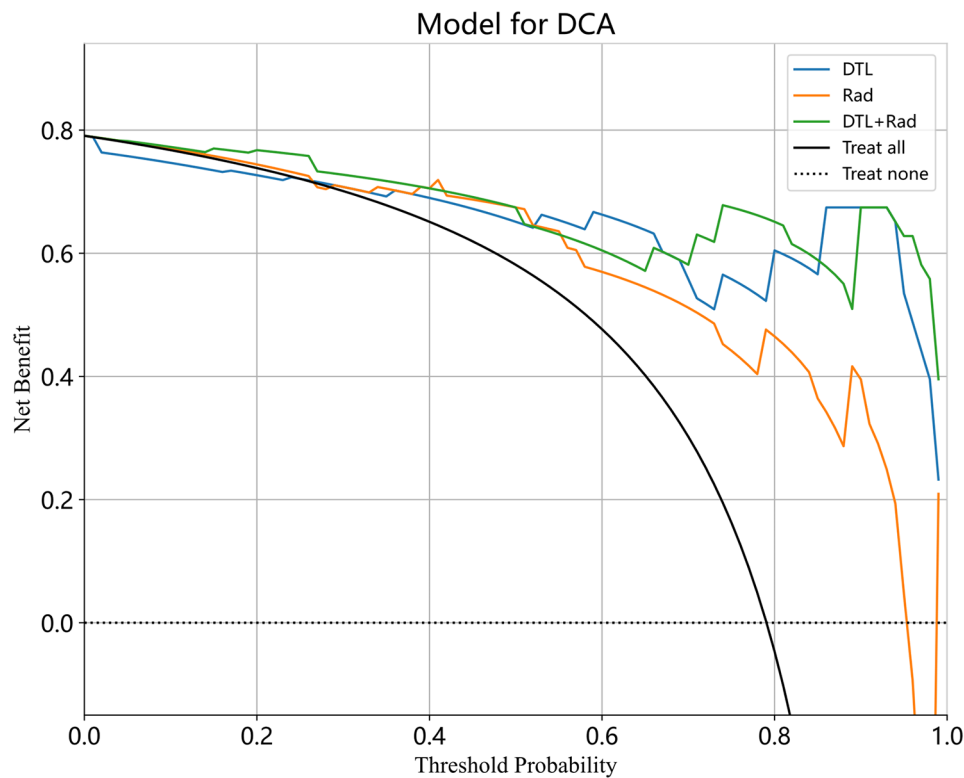
heavily relies on the operator's experience and judgment, employing predefined, straightforward features, which limits the ability of the model to handle complex or atypical imaging patterns and results in poor reproducibility. In addition, DTL, in conjunction with X-ray imaging was used for feature extraction and prediction [25, 26]. Despite its cost-effectiveness, X-ray imaging is inferior to CT in delivering detailed image information, making it unsuitable as the first-choice examination for suspected spinal fractures. Moreover, Zhang [27] applied DTL to models by extracting features from the largest sagittal area of interest (ROI) on CT images; however, this approach may overlook the conditions and spatial relationships of surrounding tissues.

Furthermore, while previous studies [10] have often used traditional radiomics combined with clinical data to construct nomograms, we chose not to use a nomogram in our study. Our primary objective was to leverage more advanced, data-driven approaches for predictive modeling. While nomograms are valuable for integrating clinical data, they rely on predefined features and assumptions, which may not fully capture the complexity of medical imaging.

In contrast, our approach combines traditional radiomics with deep transfer learning (DTL), allowing us to detect subtle patterns and complex features that are often missed by conventional methods, thereby enhancing the accuracy and adaptability of our model.



**Fig. 9** The AUCs of various models in the test cohort



**Fig. 10** Decision curve analysis was developed with different prediction models

In traditional radiomics, PyRadiomics is used to extract quantitative features from medical imaging data for analysis and interpretation. In contrast, DTL utilizes pretrained models to detect subtle patterns that may not be captured by human observation or traditional algorithms. This suggests the effectiveness of DTL in improving feature extraction from CT images, particularly in enhancing the visualization and interpretation of subtle indicators within complex anatomical areas such as the spine, thereby potentially increasing the diagnostic accuracy for conditions such as spinal fractures or other complex disorders.

Notably, in certain scenarios, such as fractures, changes may be induced in the surrounding soft tissues, with indirect signs including tissue swelling, hematoma formation, and alterations in the texture of surrounding muscle groups. For instance, the radiological sign of “fat stranding” (FS) in CT scans, characterized by soft tissue density streaks within fat areas, is typically associated with inflammation, infection, or trauma and is common in the abdomen, pelvis, retroperitoneum, chest, neck, and subcutaneous tissues [28–30]. This sign significantly aids physicians in accurately diagnosing acute and chronic diseases. Although FS provides critical diagnostic insights into various pathologies, its manifestations in VCF CT images may not be clear or specific. Inspired by research on the tumor microenvironment, we extended the ROI to the area surrounding

vertebral fractures in this study. Drawing on previous literature [31], we progressively expanded the 3D ROI by 2 voxels at a time, extracting features and constructing models based on these features to explore the effects of feature extraction and model performance at various expansion levels. In the research, modeling performance improvement was observed with the increase in the minimum vertical expansion of body voxels. Specifically, an expansion of 6 mm was optimal and is believed to enhance DTL efficacy in capturing essential features surrounding fractures. This optimal setting significantly refines the details and accuracy of fracture imaging, offering novel insights into the classification and prognosis of both acute and chronic VCFs. Traditionally, the focus of studies in these areas has predominantly been on tumors; however, the findings suggest that the scope of VCF studies can be broadened by DTL, providing a more nuanced understanding of these conditions. Moreover, expansions beyond this 6 mm threshold led to a decrease in the AUC, likely due to the introduction of excessive irrelevant or noisy information that obscured essential features. This observation highlights the critical balance required in configuring imaging parameters, emphasizing the need for precision to optimize diagnostic outcomes without diminishing data quality.

Then, we implemented a feature fusion model that combines traditional modeling approaches with DTL features, which includes a 6 mm expansion. This fusion

model demonstrated promising predictive ability for VCFs, achieving an AUC of 0.964 in the testing cohort. While the DeLong test did not confirm statistically significant differences between the fusion model and the individual methods, the decision curve analysis (DCA) suggested that the fusion model may provide greater clinical net benefit across a range of threshold probabilities. Although the difference in AUC between the single models and the feature fusion model was not statistically significant, DCA showed that the feature fusion model provided greater patient benefits. Meticulously integrating classical statistical methods with advanced DTL techniques has substantially improved the performance of our predictive models and catalyzed significant advancements and innovations in radiological analysis.

While the current study provides valuable insights into the application of deep transfer learning (DTL) and radiomics for diagnosing acute and chronic vertebral compression fractures (VCFs), several limitations warrant further attention.

Firstly, the relatively small sample size of VCF cases in this study limits the generalizability of our findings. Although the power analysis indicates that the current sample size is sufficient to detect significant effects, a larger cohort would help to confirm the robustness of the proposed diagnostic models. We suggest conducting multicenter studies with larger sample sizes to validate the predictive performance of the feature fusion model across diverse patient populations and clinical settings. Additionally, prospective data collection would be essential to overcome potential biases inherent in retrospective study designs, ensuring more reliable and clinically relevant results.

Moreover, our study focused primarily on CT imaging, which, while widely available, has limitations in detecting subtle soft tissue changes around fractures. Future research could explore the integration of MRI or hybrid imaging techniques with CT to provide a more comprehensive diagnostic approach. Furthermore, investigating the potential of other advanced imaging modalities, such as dual-energy CT, could offer new avenues for improving diagnostic accuracy and distinguishing between acute and chronic fractures.

These efforts would enhance the clinical applicability and robustness of diagnostic models, paving the way for more efficient and accurate identification of VCFs in diverse clinical settings.

Looking ahead, we envision several key directions for future research. First, studies could explore the refinement of cropping strategies, particularly in the context of DTL, to optimize feature extraction from surrounding tissues. This could further enhance the diagnostic ability to differentiate acute from chronic fractures based on soft tissue changes. Additionally, incorporating clinical data such as patient demographics, clinical history, and

laboratory results into predictive models could improve their performance, making them more applicable in real-world settings. As DTL and radiomics evolve, they hold significant potential for personalized medicine. Future studies could explore integrating these models into clinical decision-making to guide treatment and improve patient outcomes, bridging the gap between advanced imaging and routine clinical practice.

## Conclusion

This study successfully developed a feature fusion model that significantly improves the differentiation between acute and chronic vertebral compression fractures (VCFs), directly addressing the gaps identified in the introduction regarding the limitations of current diagnostic methods. By integrating traditional radiomics with deep transfer learning (DTL) and optimizing cropping strategies, our model enhances diagnostic accuracy and provides a reliable alternative for patients who are unable to undergo spinal MRI. The findings demonstrate the model's potential to aid in clinical decision-making, particularly in settings where MRI availability is limited. The superior performance of the fusion model, compared to single-model approaches, highlights its practical utility in routine clinical practice, where rapid and accurate diagnosis is crucial for timely treatment. Furthermore, this approach offers significant promise for future research, particularly in the development of more personalized diagnostic tools and treatment plans, as well as expanding its application to other complex musculoskeletal conditions.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12880-025-01573-9>.

Supplementary Material 1  
Supplementary Material 2  
Supplementary Material 3  
Supplementary Material 4  
Supplementary Material 5  
Supplementary Material 6  
Supplementary Material 7  
Supplementary Material 8  
Supplementary Material 9  
Supplementary Material 10  
Supplementary Material 11

## Acknowledgements

We sincerely thank all the patients who participated in this study and express our deep gratitude to the Onekey platform for its invaluable support.

## Author contributions

All authors contributed to the study conception and design. The first draft of the manuscript was written by Jing Wang. Wang and Dong contributed

equally to this article. All authors commented on previous versions of the manuscript, and all authors read and approved the final manuscript. The corresponding authors are Libo Jiang and Mingdong Zhao, with Mingdong Zhao being the final corresponding author.

#### Funding

This work was supported by grants from National Natural Science Foundation of China (grant nos. 81971315), Natural Science Foundation of Shanghai (grant no. 23ZR1409000).

#### Data availability

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

#### Declarations

##### Ethics approval and consent to participate

This study was approved by the Institutional Review Board of Jinshan Hospital, Fudan University Medical Ethics Committee (approval number JIEC 2024-S59). The requirement for informed consent was waived by the board because all patients signed a general informed consent upon admission, which covers the use of their data for research purposes. Additionally, the participants' identities were separated from the research data to ensure anonymity, and patients retained the right to withdraw from the study at any time. The study was conducted in accordance with the relevant guidelines and regulations.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 25 August 2024 / Accepted: 28 January 2025

Published online: 06 February 2025

#### References

1. Alsoof D, Anderson G, McDonald C L, et al. Diagnosis and management of Vertebral Compression fracture [J]. *Am J Med*. 2022;135(7):815–21.
2. Hoyt D, Urits I. Current concepts in the management of Vertebral Compression fractures [J]. *Curr Pain Headache Rep*. 2020;24(5):16.
3. Khan M A, Jennings J W Bakerjc, et al. ACR appropriateness Criteria® Management of Vertebral Compression fractures: 2022 update [J]. *J Am Coll Radiol*. 2023;20(5S):102–24.
4. Carli D, Venmans A, Lodder P, et al. Vertebroplasty versus active control intervention for Chronic Osteoporotic Vertebral Compression fractures: the VERTOS V Randomized Controlled trial [J]. *Radiology*. 2023;308(1):e222535.
5. Parreira P C S, Maher C G, Megale R Z, et al. An overview of clinical guidelines for the management of vertebral compression fracture: a systematic review [J]. *Spine J*. 2017;17(12):1932–8.
6. Beall D P, De Leacy R. A. Management of Chronic Vertebral Compression fractures with Vertebroplasty: focus on clinical symptoms [J]. *Radiology*. 2023;308(1):e231243.
7. Frellesen C, Azadegan M. Dual-energy computed tomography-based Display of Bone Marrow Edema in Incidental Vertebral Compression fractures: diagnostic accuracy and characterization in oncological patients undergoing routine staging computed tomography [J]. *Invest Radiol*. 2018;53(7):409–16.
8. Fritz B, Yi P H, Kijowski R et al. Radiomics and Deep Learning for Disease Detection in Musculoskeletal Radiology: an overview of Novel MRI- and CT-Based approaches [J]. *Invest Radiol*. 2023;58(1).
9. Muehlethaler U J, Mannil M, Becker A S, et al. Vertebral body insufficiency fractures: detection of vertebrae at risk on standard CT images using texture analysis and machine learning [J]. *Eur Radiol*. 2019;29(5):2207–17.
10. Yang H, Yan S, Li J, et al. Prediction of acute versus chronic osteoporotic vertebral fracture using radiomics-clinical model on CT [J]. *Eur J Radiol*. 2022;149:110197.
11. Duan S, Hua Y, Cao G, et al. Differential diagnosis of benign and malignant vertebral compression fractures: comparison and correlation of radiomics and deep learning frameworks based on spinal CT and clinical characteristics [J]. *Eur J Radiol*. 2023;165:110899.
12. Kim A Y, Yoon M A, Ham S J, et al. Prediction of the acuity of Vertebral Compression fractures on CT using Radiologic and Radiomic features [J]. *Acad Radiol*. 2022;29(10):1512–20.
13. Lu L, Wang X, Carneiro G, et al. Deep learning and convolutional neural networks for medical imaging and clinical informatics [M]. Springer. 2019.
14. Han Y, Wang Z, Chen A, A deep transfer learning-based protocol accelerates full quantum mechanics calculation of protein [J]. *Brief Bioinform*. 2023;24(1).
15. Al Zorgani M M, Ugaill H, Pors K, et al. Deep transfer learning-based Approach for glucose Transporter-1 (GLUT1) expression Assessment [J]. *J Digit Imaging*. 2023;36(6):2367–81.
16. Yaqoob S, Cafiso S, Morabito G, et al. Deep transfer learning-based anomaly detection for cycling safety [J]. *J Saf Res*. 2023;87:122–31.
17. Jiang Y, Zhou K, Sun Z, et al. Non-invasive tumor microenvironment evaluation and treatment response prediction in gastric cancer using deep learning radiomics [J]. *Cell Rep Med*. 2023;4(8):101146.
18. Yu Y, He Z. Magnetic resonance imaging radiomics predicts preoperative axillary lymph node metastasis to support surgical decisions and is associated with tumor microenvironment in invasive breast cancer: a machine learning, multicenter study [J]. *EBioMedicine*. 2021;69:103460.
19. Wang X, Xie T, Luo J, et al. Radiomics predicts the prognosis of patients with locally advanced breast cancer by reflecting the heterogeneity of tumor cells and the tumor microenvironment [J]. *Breast Cancer Res*. 2022;24(1):20.
20. Kaup M, Wichmann J L, Scholtz J-E, et al. Dual-energy CT-based Display of Bone Marrow Edema in Osteoporotic Vertebral Compression fractures: Impact on Diagnostic Accuracy of radiologists with varying levels of experience in correlation to MR Imaging [J]. *Radiology*. 2016;280(2):510–9.
21. Petritsch B, Kosmala A, Weng A M, et al. Vertebral Compression fractures: third-generation dual-energy CT for detection of bone marrow edema at visual and quantitative analyses [J]. *Radiology*. 2017;284(1):161–8.
22. Zhang X, Zhou X, Lin M et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2018 [C].
23. Court-Brown C M, McQueen MM. Global Forum: fractures in the Elderly [J]. *J Bone Joint Surg Am*. 2016;98(9):e36.
24. Wang C-K, Tsai J-M, Chuang M-T, et al. Bone marrow edema in vertebral compression fractures: detection with dual-energy CT [J]. *Radiology*. 2013;269(2):525–33.
25. Zhang J, Xia L, Tang J, et al. Constructing a deep learning Radiomics Model based on X-ray images and Clinical Data for Predicting and distinguishing Acute and chronic osteoporotic vertebral fractures. A Multicenter Study [J]. *Acad Radiol*. 2023.
26. Chen W, Liu X, Li K, et al. A deep-learning model for identifying fresh vertebral compression fractures on digital radiography [J]. *Eur Radiol*. 2022;32(3):1496–505.
27. Zhang J, Liu J, Liang Z, et al. Differentiation of acute and chronic vertebral compression fractures using conventional CT based on deep transfer learning features and hand-crafted radiomics features [J]. *BMC Musculoskelet Disord*. 2023;24(1):165.
28. Abramson Z, Sheyn A, Goode C, et al. Perimandibular Fat Stranding sign: a diagnostic aid for subtle mandibular fractures [J]. *AJR Am J Roentgenol*. 2022;218(5):917–8.
29. Lin H-A, Tsai H-W, Chao C-C, et al. Periappendiceal fat-stranding models for discriminating between complicated and uncomplicated acute appendicitis: a diagnostic and validation study [J]. *World J Emerg Surg*. 2021;16(1):52.
30. Hedgire S, Baliyan V, Zucker E J, et al. Perivascular Epicardial Fat stranding at coronary CT angiography: a marker of Acute Plaque rupture and spontaneous coronary artery dissection [J]. *Radiology*. 2018;287(3):808–15.
31. Ding J, Chen S, Serrano Sosa M, et al. Optimizing the Peritumoral Region Size in Radiomics Analysis for Sentinel Lymph Node Status prediction in breast Cancer [J]. *Acad Radiol*. 2022;29(Suppl 1 Suppl 1):223–8.

#### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.