Research Article

A Classification Method for Thoracolumbar Vertebral Fractures due to Basketball Sports Injury Based on Deep Learning

XiaoGan Chen[1] **and Yu Liu** [2]

[1]*Wuxi Taihu University, Wuxi Jiangsu 214000, China*
[2]*Wuxi 9th People's Hospital Affiliated to Soochow University, Wuxi Jiangsu 214000, China*

Correspondence should be addressed to Yu Liu; liuyuwuxiwuxi@163.com

Objective. There are more and more basketball competitions, to propose a classification method of thoracolumbar fractures to assist in the diagnosis of basketball injuries, to analyze the feasibility of its clinical application, and to improve the recovery rate. *Methods*. From February 2015 to May 2022, 1130 CT images of thoracolumbar fractures admitted to our hospital and affiliated hospital units due to basketball injuries were collected, and the image labeling system uniformly labeled them. All CT images were classified according to the AO spine classification of thoracolumbar injuries. In the ABC-type classification, 935 CT images were used for training and validation to optimize the deep learning system, including 815 training sets and 120 validation sets; the remaining 198 CT images were used as test sets for comparing the deep learning system and clinician's diagnosis. In the classification of subtype A, a total of 523 CT scans can be performed for training and validation to optimize the deep learning system, including 500 training sets and 23 validation sets; the remaining 94 CT images are used as test sets for comparing depth learning systems and clinicians' diagnostic results. *Results*. The deep learning system had a correct rate of ABC classification of fractures in 86.4%, with a kappa coefficient of 0.850 ($P < 0.001$); the correct rate of subtype A was 85.3%, with a kappa coefficient of 0.815 ($P < 0.001$). *Conclusion*. The classification accuracy of thoracolumbar fractures based on deep learning is high. The method can assist in diagnosing CT images of thoracolumbar fractures and improve the current manual and complex diagnosis process.

1. Introduction

From the analysis of the status quo of clinical treatment of basketball injuries, basketball is a ball sport with strong collective antagonism. Offense and defense alternate instantaneously, sudden start and stop, jumping and squatting, body position changes, etc.; players constantly change. Action is a characteristic of basketball, so basketball is very prone to various sports injuries, and the incidence of fractures continues to rise. The incidence of medical disputes in fracture diagnosis remains high. Improving the timeliness, accuracy, and comprehensiveness of fracture diagnosis is necessary. Conventional radiographic diagnosis is difficult for occult or minor fractures in fracture diagnosis. Once missed or misdiagnosed, it may lead to complications such as compression plate fracture, deep infection, delayed fracture union, or nonunion. Improve fracture diagnosis techniques.

The current fracture diagnostic typing is based on simple typing of fracture morphology or neurological function score, which can only distinguish the general location, morphology, and degree of fracture. The diagnostic typing of the same patient varies greatly among clinicians. The current fracture diagnostic typing system is less intelligent and less consistent and lacks guidance on personalized treatment modalities, which is not conducive to forming and applying standardized treatment plans for spinal fractures. Therefore, a classification method is urgently needed to unify the typing and diagnose fracture types.

Deep learning technology is applied in the clinical medical field mainly for reading and diagnosing medical images. The deep learning technology can quickly mark abnormal structures or areas in the medical images of patients, providing important references for doctors' judgment and diagnosis. Current research on medical image processing based on

deep learning technology is mainly focused on large [1–3], liver [4], pancreas [5], prostate [6], multiorgan [7], and fracture analysis [8–10]. Artificial intelligence diagnosis technology based on medical imaging is still in the preliminary exploration stage in the field of orthopedics [11–14]. The study of orthopedic diagnosis problems from the perspective of artificial intelligence, especially for the identification and autonomous localization of thoracolumbar vertebrae [15–19], is still in urgent need of more in-depth research. This study proposes a novel classification method for fractures based on fast-region convolutional neural network (Faster RCNN) deep learning to improve the complex manual diagnosis process.

2. Materials and Methods

2.1. Study Subjects. The CT images of all cases this time include 1130 fracture CT images collected in our hospital and affiliated hospitals from February 2015 to May 2022, which were injured in basketball and collected in our hospital and affiliated hospitals. Images of fractures were observed. As sample data, exclude unique samples with special bone structures such as severe shadow shading, hidden fractures, and osteoporosis.

2.2. Faster RCNN-Based Fracture Classification. This study establishes a fracture classification method based on the Faster RCNN deep learning network. For the problem of various types of fractures, the Faster RCNN is applied to iteratively learn a large number of prelabeled CT sample data of fracture types, extract the different features of various types of fracture sample data as the input of the classifier, and achieve the classification of the corresponding fracture test sample data.

2.2.1. Data Preprocessing. To fit the input to the Faster RCNN, the CT images are first preprocessed. To standardize the images, including graying out the images, uniform size, and resolution, ① CT images are mainly in black and white tones and need to be grayed out. After the grayscale operation, the number of image channels is reduced, which can effectively improve the efficiency of image processing. This study uses the PIL library method that comes with Python to implement image grayscale. ② When the scale difference between images is significant, the image features will also vary greatly, which will significantly impact the training model's speed and results. This study also uses the self-contained PIL library in Python to process the size and resolution of the images so that they match the input size of the iterative learning network.

2.2.2. Dataset Construction. Based on data preprocessing, doctors who have been practicing clinical work for more than ten years at our hospital applied the Imaging Labeling system to classify and confirm CT images and label them to build the training and validation sets for deep learning networks.

According to the AO spine classification 2015, vertebral fractures are classified as type A (vertebral compression), type B (anterior-posterior distraction injury), type C (anterior-posterior rotational injury), A2 (fracture line involving both endplates but not the posterior wall of the vertebral body), A3 (vertebral fracture involving only a single endplate and involving both the posterior wall and the spinal canal), and A4 (vertebral fracture involving both upper and lower endplates and the posterior wall).

For the selected sample, the classification method proposed in this study can also deal with the presence of multiple fractures in the thoracolumbar spine, considering the practical situation where multiple different types of fractures may exist simultaneously on a single CT image.

Of the three basic types of ABC, a total of 935 thoracolumbar fracture images were selected for training and validation in this study, including 815 in the training set and 120 in the validation set, including 160 for type C, 319 for lumbar A, 48 for lumbar B, 350 for thoracic A, and 58 for thoracic B; 711 for males and 224for females; and 15 to 35 years of age, with an average of 20 years. A total of 198 thoracolumbar fracture images were selected as the test sample, of which 32 were C type, 70 were lumbar A type, 12 were lumbar B type, 80 were thoracic A type, and 15 were thoracic B type; 148 were male, and 47 were female; and the ages ranged from 17 to 36 years, with a mean of 25 years.

In the subtype classification study of type A fractures, a total of 523 images of thoracolumbar fractures were selected for training and validation, including 5489 images in the training set and 34 images in the validation set, including 46 images of type A1, 28 images of type A2, and 240 images of type A2. A3 type and A4 type were 175; 371 males and 151 females, aged 15 to 35 years old, with an average of 20 years old. 113 sheets were selected as test samples, including 15 sheets of type A1, 17 sheets of type A2, 34 sheets of type A3, and 47 sheets of type A4; 74 sheets for men and 20 sheets for women; aged 17 to 36 years old, with an average of 25 years old.

2.2.3. Applying Convolutional Neural Networks to Extract Features. Faster RCNN consists of 2 modules: region proposal network (RPN) candidate frame extraction module + Fast RCNN detection module. In Faster RCNN, the training image is input to the VGG16 convolutional neural network to extract the features of the image; the RPN layer generates the candidate frames, and then, each "candidate frame" is mapped to the feature map to obtain the feature map of the region of interest; the pooling layer of the region of interest divides each candidate region into $M \times N$ blocks (whereby M denotes pool height, and N denotes pool width). The pooling layer divides each candidate region into $M \times N$ blocks (M denotes the pool height and N denotes the pool width). The maximum pooling operation is performed for each block so that the candidate regions of different sizes are transformed into feature vectors of uniform size during the feature mapping process and then fed to the next layer. The pooling layer makes the output feature map consistent with the dimensionality of the fully connected layer after it. Finally, Softmax Loss (detecting classification probability) and Smooth L1 Loss (detecting edge regression) are used for classification and regression, as shown in Figure 1.

The convolution layer is also known as the feature extraction layer. The convolutional layer contains 3 layers of
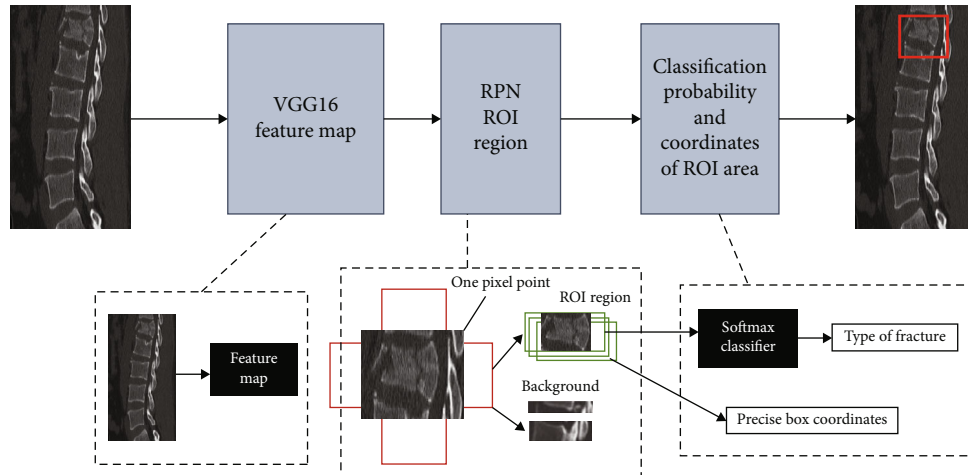
FIGURE 1: Faster RCNN learning classification process diagram. ROI: region of interest; Feature map: feature map.

convolution, pooling, and activation. To accommodate fracture image classification, this study's Faster RCNN learning model has 13 convolutional and activation layers and 4 pooling layers, as shown in Figure 2 for the specific structure.

The RPN operates as follows: a sliding scan is performed on the final convolutional feature map using a small network, and sliding window processing ensures that the regression and classification layers are associated with the entire feature space of the convolutional layer. This is then mapped to a low-dimensional vector, and finally, this low-dimensional vector is sent to the regression and classification layers to output classification diagnostics, as shown in Figure 3.

2.3. Evaluation Indicators. In this study, according to the classification results of the test set, the overall correct rate and kappa coefficient are used as the evaluation indexes of the overall classification effect, where the correct overall rate is the proportion of all correctly classified samples to the total samples, which is an assessment of the overall accuracy; the kappa coefficient reflects the consistency between the classification results of the classifier and the actual results. The Faster RCNN is a multiclassification network, for each class. For each category, we split it into multiple binary categories for calculation, such as for ABC-type fracture typing; when calculating the evaluation index, the problem of five categories is split into five binary categories, and the single category correct rate, sensitivity, specificity, positive predictive value, negative predictive value, and Jorden index are used as the evaluation index for the classification effect of a category, where the single category correct rate is the proportion of samples to the total samples for whether the classification is correct for a specific type. The sensitivity is the proportion of all positive cases that are correctly classified, which measures the capability of the classifier to identify positive cases; the specificity is the probability of correctly predicting the "true negative" case, which measures the ability of the model to identify negative cases; the positive predictive value is the proportion of the samples classified as positive cases that are actually positive cases. The loss value is a measure of the difference between the output

of the trained model and the actual result and is related to the choice of the loss function and the number of iterations.

2.4. Statistical Methods. SPSS 24.0 software was used to analyze the data statistically. Correctness, sensitivity, specificity, positive predictive value, negative predictive value, Youden index, and kappa coefficient were selected for statistical analysis according to data characteristics. The test level was $\alpha = 0.05$.

3. Results

3.1. ABC-Type Classification Test Results. The results of the classification of the ABC type are shown in Table 1. Of the 30 C-type fractures identified, 1 was not identified, 2 were incorrectly identified as thoracic A type, and 27 were correctly identified. Among the 71 lumbar type A fractures, 4 were incorrectly identified as thoracic type A, and 67 were correctly identified. Of the 12 lumbar type B fractures identified, 2 were incorrectly identified as lumbar type A and 10 were correctly identified. Of the 72 thoracic A fractures identified, 4 were not identified, 2 were incorrectly identified as lumbar A, and 66 were correctly identified. Of the 13 thoracic spine type B fractures identified, 2 were incorrectly identified as thoracic type A and 11 were correctly identified. The correct overall rate for this classification was 89.4%, with a kappa coefficient of 0.849 (P 0.001). Table 2 shows the single class correct rate, sensitivity, specificity, positive predictive value, negative predictive value, and Jorden index for each category calculated from this result. Examples of successful identifications are shown in Figures 4(a) and 4(b), and examples of incorrect identifications are shown in Figures 4(c) and 4(d). Figure 5 shows the relationship between the loss values and the number of iterations during ABC-type classification training.

3.2. Type A Subtype Classification Test Results. The results of the classification of type A subtypes are shown in Table 3. Of the 12 type A1 fractures, 1 was not identified, 1 was incorrectly identified as type A4, and 10 were correctly identified. Among 13 A2-type fractures, 4 were incorrectly identified as A4 type,
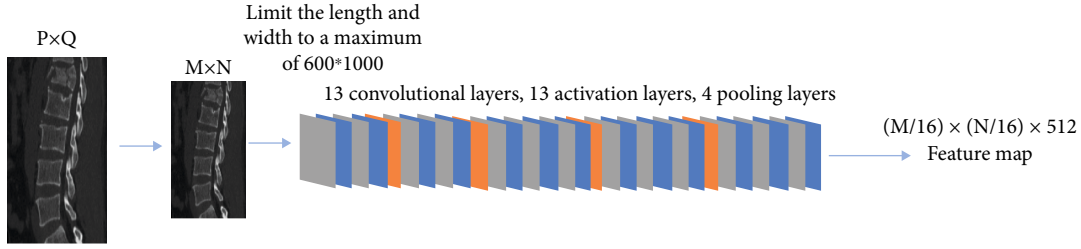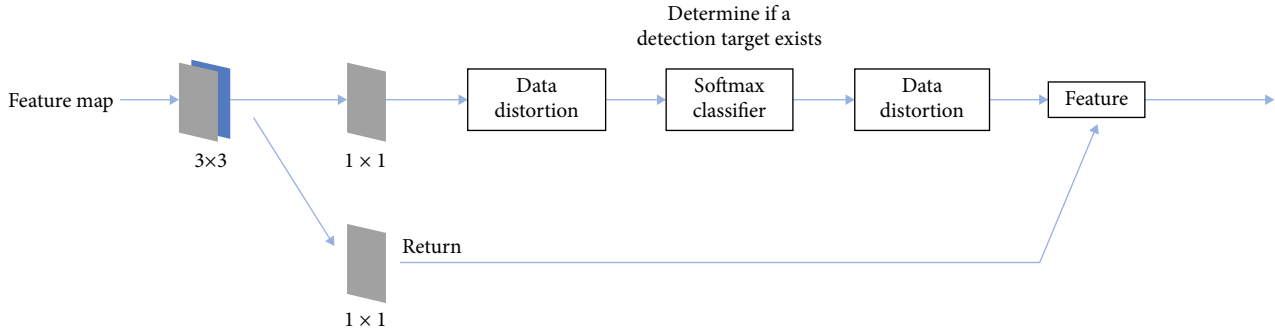
FIGURE 2: Convolutional layer specific structure.

FIGURE 3: RPN layer specific structure.

TABLE 1: ABC-type classification test results (sheets).

| Intelligent classification | Type C | Manual marking | | | | Total |
		Lumbar spine type A	Lumbar spine type B	Thoracic spine type A	Thoracic spine type B	
Type C	27	0	0	0	0	27
Lumbar spine type A	0	67	2	2	0	71
Lumbar spine type B	0	0	10	0	0	10
Thoracic spine type A	2	4	0	66	2	72
Thoracic spine type B	0	0	0	0	11	11
No identification	1	0	0	4	0	5
Total	30	71	12	72	12	197

TABLE 2: ABC typing test evaluation.

Type of fracture	Single category correct rate	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Youden index
Type C	0.979	0.883	1.000	1.000	0.979	0.875
Lumbar spine type A	0.952	0.967	0.957	0.932	0.964	0.893
Lumbar spine type B	0.988	0.777	1.000	1.000	0.986	0.770
Thoracic spine type A	0.909	0.902	0.920	0.867	0.951	0.826
Thoracic spine type B	0.981	0.786	0.998	1.000	0.979	0.788

and 9 were correctly identified. Of the 38 A3 fractures, 1 was not identified, 2 were incorrectly identified as A4, and 35 were correctly identified. Of the 41 A4 fractures, 3 were not identified, 1 was incorrectly identified as A3, and 37 were correctly identified. The correct overall rate for this classification was 87.5%, with a kappa coefficient of 0.817 ($P < 0.001$). Table 4 shows the single class correctness, sensitivity, specificity, positive predictive value, negative predictive value, and Jorden index calculated from this result. Figure 6 shows examples of successful and incorrect identifications. Figure 7 shows the relationship between the loss value and the number of iterations during subtype A classification training.
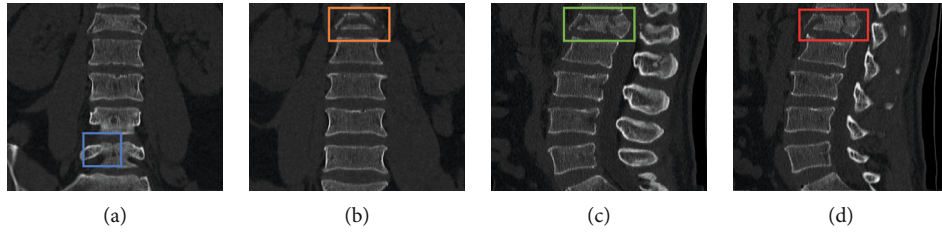
FIGURE 4: Example of ABC-type classification test results. (a) Correct identification of C fracture (sagittal). (b) Correct identification of A fracture (coronal). (c) Misidentification of lumbar A as thoracic A (sagittal). (d) Misidentification of thoracic B as lumbar B (sagittal) as lumbar type B (sagittal).
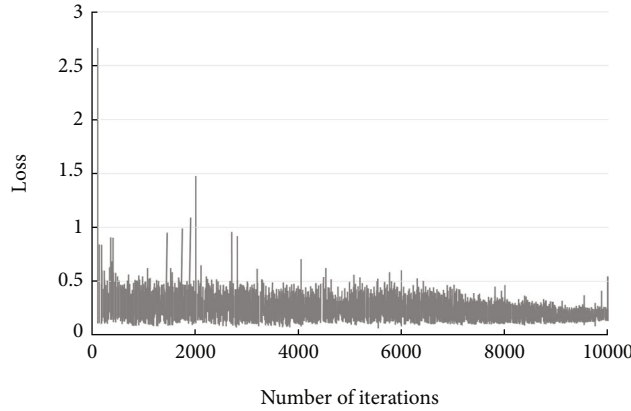

FIGURE 5: ABC-type classification loss.

TABLE 3: Class A subtype classification test results (sheets).

Type of fracture	Single category correct rate	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Youden index
Type A1	0.985	0.875	1.000	0.999	0.979	0.877
Type A2	0.947	0.940	0.958	0.919	0.964	0.896
Type A3	0.983	0.773	0.988	1.000	0.986	0.772
Type A4	0.916	0.902	0.922	0.8655	0.940	0.825

TABLE 4: Type A subtyping test evaluation.

Intelligent classification	Manual marking				
	Type A1	Type A2	Type A3	Type A4	Total
Type A1	10	0	0	0	10
Type A2	0	12	0	0	12
Type A3	0	0	35	3	38
Type A4	6	7	2	46	61
No	1	0	1	3	5
Total	17	19	38	52	126

4. Discussion

In this study, deep learning technology was applied to the classification of human thoracolumbar fractures, and a large number of CT images were screened, preprocessed, renamed, and sample annotated; after which Faster RCNN was used to learn and train the sample data of thoracolumbar fractures, establish the classification model and method of thoracolumbar fractures, and apply the CT images of human thoracolumbar fractures to test experiments.

For the classification of type ABC fractures, the overall correct classification rate of the classification model was 86.4%, with a kappa coefficient of 0.850. The sensitivity and specificity of the classification of type A fractures were higher, above 90%, for both thoracic and lumbar spine. Also, there were more unmarked samples for type A compared with type B and C. This may be caused by the bias of irregular shading in the type A dataset. Moreover, the unmarked cases are rarely found in the coronal plane and are mainly concentrated in the sagittal plane. This is because types A3 and A4 are both burst fractures with a richer fracture pattern, leading to increased difficulty in classification.

Type C ranked 2nd in the Jorden index. Of the 904 training and 35 validation images, 160 were type C, a smaller percentage than type A. The main reason for type C being misidentified is that in the coronal plane type C is so similar to type A that features characterizing type C osteochondral misalignment are only available when in the sagittal plane. Type C and type B can have similar
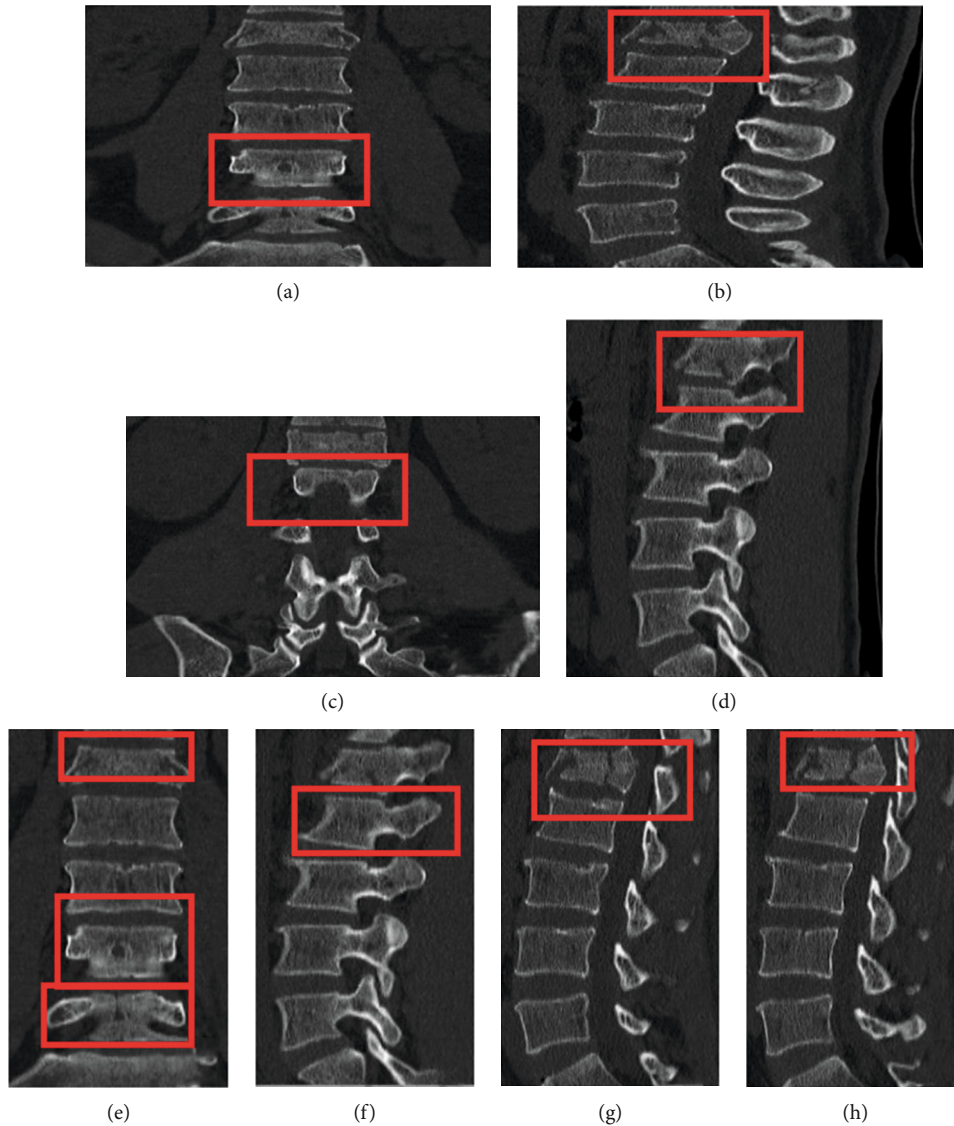
FIGURE 6: Example of ABC-type classification test results. (a) Correctly identified type A3 (coronal). (b) Correctly identified type A3 (sagittal). (c) Correctly identified type A4 (coronal). (d) Correctly identified type A4 (sagittal). (e) Correctly identified A1, A4 type (sagittal). (f) Misidentified A1 type as A4 type (sagittal). (g) Misidentified A3 type as A4 type (sagittal). (h) Misidentified A4 type as A3 type (sagittal).
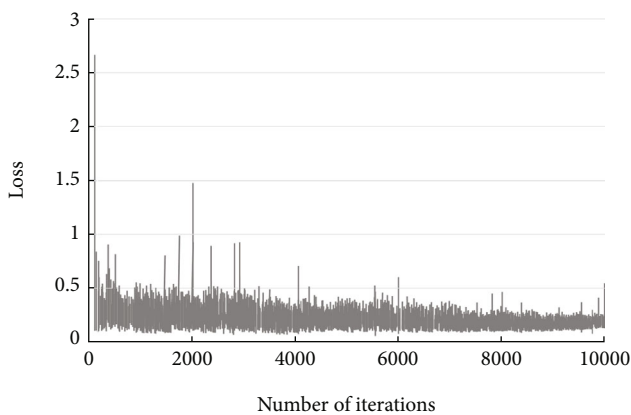

FIGURE 7: Type A subfractional loss.

problems in the coronal plane. In the case of similarity, the misidentification as type A is mainly due to the higher proportion of type A pictures, which obtains a more comprehensive characterization of the characterized type A fracture caused by the uneven proportion of data set types. Type C can have a higher correct rate because it is a displaced fracture morphologically more different from A and B and can be easily distinguished.

Type B fractures had the lowest Jorden index in the test, between 0.7 and 0.8, with a specificity of 100%, mainly because too few type B samples were used to train the network, even less than type C. There was an imbalance between the sample proportions, and the learning network model learned the features of type A and C fractures better; secondly, type B tends to be a fracture of the bone joint extending into the spinous portion when the fracture of the spinous is not apparent; it is easy to misidentify type B as type A.

For subtype A fracture classification, the overall correctness of the model was 85.3%, with a kappa coefficient of 0.815. The sensitivity was relatively high for types A3 and A4, exceeding 90%. The main problem with A3 and A4 identification was mutual misidentification. Type A3 fractures had only one endplate fracture, while A4 had both upper and lower endplates fractured. The distinction between A3 and A4 can only be made by whether 2 endplates are involved, and the difference between the two is slight, which is the main reason for their confusion. In addition, both A3 and A4 are burst fractures with diverse and variable fracture patterns. In a way, the amount of data in the training sample of 500 cases was insufficient, leading to unmarked A3 and A4 types.

The sensitivity of both A1 and A2 recognition is low, mainly due to the small training set of A1 and A2. Except for the unmarked cases, the false identifications are all types A3 and A4 due to the relatively large proportion of A3 and A4 in this dataset. In this study, the type ratio has been artificially balanced in the selection of images so that the number of A1 and A2 is similar and the number of A3 and A4 is similar, which will ensure a specific recognition correct rate with insufficient samples in the dataset.

This study's results show that the sensitivity of type B, type C, and two subtypes, A1 and A2, is lower than that of type A and types A3 and A4, respectively. Due to the small number of cases of these fracture types, the amount of data for these types is not enough compared to the amount of data for other types, and the learning network cannot fully extract the features of these fracture types, resulting in poorer recognition of these types. Deep learning has high requirements for dataset quality, and the acquisition of medical images is heavily dependent on clinical cases, making it difficult to establish a universal image dataset. This study found that the proportion of different types in the dataset also impacts the experimental results. Based on clinical experience, patients with A1 and A2 fracture types are themselves fewer, and some fracture cases are less distinctive, resulting in less easy identification by the machine; A3 and A4 types are more concentrated but have more fracture manifestations, resulting in the learning network not learning enough comprehensive features for correct identification. The imbalance in the proportion of dataset types can lead to the misidentification of the less represented categories as the more represented ones. The uneven distribution of the dataset due to insufficient clinical imaging data will impact the experimental results and require an increase in sample size to reduce bias in further studies.

The next issue is the number of training datasets and iterations. The ABC typing in this study used 935 training and validation sets and 10,000 iterations of learning. Before the final training, 10,000, 3,000, and 6,000 iterations were used successively, whose recognition results had a significant loss to achieve the expected results, and the number of iterations of 10,000 obtained better results. Therefore, it can be concluded that the size of the training set directly affects the classification and recognition results, which could be solved using other advanced machine learning techniques [20].

5. Conclusion

In summary, this study starts from the current problems faced by diagnosing spinal fractures caused by basketball injuries and introduces the latest deep learning methods for fracture diagnosis classification. The advanced thoracolumbar fracture classification method can realize the machine's autonomous fracture detection and fracture classification, which helps to shorten the time required for manual diagnosis, effectively promote high-quality medical resources, and improve the accuracy and consistency of early diagnosis of spinal fractures, thereby improving the early diagnosis of spinal fractures, efficacy of treatment of fractures.

There are still some limitations in this study: with a small sample data size, the classification accuracy of the method proposed in this research is not high, and it is easy to miss and misidentify. Future research will focus on solving the above small sample problem and achieving accurate learning for fracture patterns by changing the network structure and algorithm.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Consent

All human subjects in this study have given their written consent for the participation in our research.

Conflicts of Interest

The authors declare no conflict of interest for this paper.

References

[1] M. Havaei, A. Davy, D. Warde-Farley et al., "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.

[2] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[3] B. H. Menze, A. Jakab, S. Bauer et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[4] P. F. Christ, M. E. Elshaer, F. Ettlinger et al., "Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields," in *International Conference on Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016*, pp. 415–423, Cham, 2016.

[5] H. R. Roth, L. Lu, A. Farag et al., "DeepOrgan: multilevel deep convolutional networks for automated pancreas segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pp. 556–564, Cham, 2015.

[6] L. Yu, X. Yang, H. Chen, J. Qin, and P. A. Heng, "Volumetric convNets with mixed residual connections for automated prostate segmentation from 3D MR images," in *Proceedings*

of the Thirty-First AAAI Conference on Artificial Intelligence, vol. 31, pp. 66–72, San Francisco, 2017.

[7] X. Zhou, R. Takayama, S. Wang, T. Hara, and H. Fujita, "Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method," *Medical Physics*, vol. 44, no. 10, pp. 5221–5233, 2017.

[8] M. Adams, W. Chen, D. Holcdorf, M. W. McCusker, P. D. L. Howe, and F. Gaillard, "Computer vs human: deep learning versus perceptual training for the detection of neck of femur fractures," *Journal of Medical Imaging and Radiation Oncology*, vol. 63, no. 1, pp. 27–32, 2019.

[9] D. W. Stoller, Ed., *Magnetic Resonance Imaging in Orthopaedics and Sports Medicine*, vol. 1, Lippincott Williams & Wilkins, 2007.

[10] U. J. Muehlematter, M. Mannil, A. S. Becker et al., "Vertebral body insufficiency fractures: detection of vertebrae at risk on standard CT images using texture analysis and machine learning," *European Radiology*, vol. 29, no. 5, pp. 2207–2217, 2019.

[11] Y. C. Li, H. H. Chen, H. Horng-Shing Lu, H. T. Hondar Wu, M. C. Chang, and P. H. Chou, "Can a deep-learning model for the automated detection of vertebral fractures approach the performance level of human subspecialists?," *Clinical Orthopaedics and Related Research*, vol. 479, no. 7, pp. 1598–1612, 2021.

[12] C. F. Ciuşdel, A. Vizitiu, F. Moldoveanu, C. Suciu, and L. M. Itu, "Towards deep learning based estimation of fracture risk in osteoporosis patients," in *2017 40th International Conference on Telecommunications and Signal Processing (TSP)*, pp. 659–665, Barcelona, Spain, 2017.

[13] J. E. Burns, J. Yao, and R. M. Summers, "Vertebral body compression fractures and bone density: automated detection and classification on CT images," *Radiology*, vol. 284, no. 3, pp. 788–797, 2017.

[14] N. Lessmann, B. van Ginneken, P. A. de Jong, and I. Išgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification," *Medical Image Analysis*, vol. 53, pp. 142–155, 2019.

[15] C. K. Kepler, A. R. Vaccaro, J. D. Koerner et al., "Reliability analysis of the AOSpine thoracolumbar spine injury classification system by a worldwide group of naïve spinal surgeons," *European Spine Journal*, vol. 25, no. 4, pp. 1082–1086, 2016.

[16] A. Suri, B. C. Jones, G. Ng et al., "A deep learning system for automated, multi-modality 2D segmentation of vertebral bodies and intervertebral discs," *Bone*, vol. 149, p. 115972, 2021.

[17] J. W. Seo, S. H. Lim, J. G. Jeong, Y. J. Kim, K. G. Kim, and J. Y. Jeon, "A deep learning algorithm for automated measurement of vertebral body compression from X-ray images," *Scientific Reports*, vol. 11, no. 1, p. 13732, 2021.

[18] Y. Li, Y. Zhang, E. L. Zhang et al., "Differential diagnosis of benign and malignant vertebral fracture on CT using deep learning," *European Radiology*, vol. 31, no. 12, pp. 9612–9619, 2021.

[19] F. Rehman, S. I. Ali Shah, M. N. Riaz, and S. O. Gilani, "A region-based deep level set formulation for vertebral bone segmentation of osteoporotic fractures," *Journal of Digital Imaging*, vol. 33, no. 1, pp. 191–203, 2020.

[20] Z. Tang, S. Wang, X. Chai, S. Cao, T. Ouyang, and Y. Li, "Auto-encoder-extreme learning machine model for boiler NOx emission concentration prediction," *Energy*, vol. 256, p. 124552, 2022.