Vertebral Bone Segmentation and Detection of Non-Traumatic Vertebral Compression Fractures with CNN from Computed Tomography Images

Murat Türkmen[1*] , Zeynep Orman[2] , Rauf Hamid[3] , Serdar Arslan[3] , Osman Kızılkılıç[3]

[1] Vocational School of Technical Sciences, Department of Computer Technologies, Başkent University, Ankara 6490, Turkey
[2] Faculty of Engineering, Department of Computer Engineering, Istanbul University-Cerrahpasa, İstanbul 34315, Turkey
[3] Faculty of Medicine, Department of Internal Medicine, Istanbul University-Cerrahpasa, İstanbul 34098, Turkey

Corresponding Author Email: muratturkmen@baskent.edu.tr

ABSTRACT

Non-traumatic vertebral compression fractures are increasingly common due to longer life expectancies. Age-related bone mass loss significantly contributes to these fractures. Typically asymptomatic for extended periods, early detection of non-traumatic vertebral compression fractures can reduce associated health issues and enable more effective treatment. Deep learning methods have shown high accuracy and sensitivity in detecting, classifying, diagnosing, and segmenting various pathological conditions in healthcare. Recently, these methods have been applied more frequently in the detection of non-traumatic vertebral compression fractures and vertebral body segmentation research. This study introduces a unique dataset to apply deep learning techniques, using raw computed tomography (CT) images of patients. The dataset was compiled from retrospective CT images taken at Istanbul University-Cerrahpasa, Cerrahpasa Faculty of Medicine, Department of Radiology. It includes 197 individuals, with 100 diagnosed with non-traumatic vertebral compression fractures and 97 without. Radiological diagnoses of non-traumatic vertebral compression fractures were added based on CT reports. The dataset comprises a total of 118,200 cross-sectional images in DICOM format, which were enhanced using the Wiener filter. The U-Net network was used to segment 6,301 vertebrae, achieving a 100% dice overlap index score. Additionally, 593 features of vertebral fractures confirmed by reports were extracted using the radiomics method, and 537 features were selected via the logarithmic lambda method. The convolutional neural network (CNN) classification model was employed, achieving an accuracy of 86.7%. The classification results were evaluated through ROC-AUC, loss, and accuracy graphs.

1. INTRODUCTION

In recent years, there has been an increasing focus on the application of machine learning to radiological images [1]. Studies have demonstrated that deep learning can interpret complex medical data at a level comparable to healthcare professionals [2]. The vertebral column, consisting of 33 vertebrae organized into cervical, thoracic, lumbar, sacral, and coccyx groups, forms the main structure supporting the body's weight. A non-traumatic vertebral compression fracture is a type of fracture that occurs gradually without a sudden injury. Early detection of these fractures can facilitate effective treatment to prevent future fractures and osteoporosis. However, non-traumatic vertebral fractures are often asymptomatic before becoming noticeable, leading to underdiagnosis and insufficient treatment [3].

Multiple imaging techniques, such as radiographs, CT scans, MRI, and PET, are utilized to assess vertebral anatomy and diagnose spinal conditions. CT scans, with their current generation scanning techniques, offer a spatially accurate method for evaluating the three-dimensional structure of vertebrae [4]. Computational algorithms are employed to process these images, extracting valuable information, or enhancing image quality through techniques like enhancement, convolution, and filtering. Minimizing data loss during the enhancement of vertebral images is crucial. Segmentation and labeling of vertebrae are key for subsequent analysis tasks like identifying abnormalities, biomechanical modeling, or image-guided interventions, which often require high precision. Manual segmentation is labor-intensive and subjective, hence fully automated, or semi-automated methods are preferred for clinical applications.

For semi-automatic or automatic vertebral fracture detection, specific characteristics of fractures must be identified and selected using methods like radiomics, Bayesian approaches, and logarithmic lambda. Classification methods then help segregate patients into groups with or without fractures. Success rates of classification, accuracy, loss, and ROC-AUC evaluations provide quantitative evidence for these methods. Non-traumatic vertebral fracture detection uses various imaging techniques and tools, including MRI, DXA, QBT, PET, SPECT/CT, CT, and the fracture risk assessment tool (FRAX) [5]. However, there is a lack of studies applying deep learning to detect fractures in CT scans [6]. Radiologists,

orthopedic surgeons, and physical therapists diagnose diseases by analyzing biomedical images of the vertebrae. Given the limited number of specialists and the increasing number of patients with vertebral disorders, there is a shortfall in medical professionals [7].

This study created a CT image dataset of human vertebrae based on research conducted at Istanbul University-Cerrahpasa Faculty of Medicine's Radiology Department. Ethical approval was obtained from Istanbul University-Cerrahpasa for the dataset's creation and use in scientific studies. The dataset includes 197 patients, with a total of 111,8200 DICOM images and 6301 vertebral bone images. Among these, 100 patients were diagnosed with vertebral fractures, while 97 patients had no fractures.

In the proposed model, Wiener filtering was applied in the preprocessing phase. Vertebra segmentation was performed using the U-Net convolutional neural network method, and its success was measured with the Dice score. Accurate labeling is crucial for classification success, and the labels for each patient's vertebrae were created using connected-component labeling (CCL). Feature extraction and selection were conducted using the radiomics library, resulting in 593 fracture-related features. Logarithmic lambda selection reduced these to 537 features. The convolutional neural network classification method was used to classify the vertebral bone images, and results were evaluated using loss scores, accuracy, and AUC-ROC curves. The classification performance of the proposed model was compared with other methods such as MobileNetV3, EfficientNetB1, VGG16, and ResNet50V2.

The rest of the paper is structured as follows: Section 2 discusses related literature, Section 3 explains the materials and methods used, Section 4 details the proposed model, Section 5 presents the findings and results, and Section 6 provides the conclusions.

2. RELATED WORKS

In the literature, datasets of human vertebrae have been developed for various studies. The characteristics and applications of these datasets are as follows:

Yi et al. [8] collected T2-weighted MRIs from 804 patients with lumbar degenerative disease symptoms from three hospitals. They proposed a deep learning model integrating 3D ResNet18 and transformer architecture to detect lumbar degenerative disease, with performance evaluated via free-response receiver operating characteristic (fROC) curve, precision-recall (PR) curve, precision, recall, and F1 score [8]. Johnson et al. [9] utilized the COLONOG dataset to identify large adenomas and cancer cells, which includes 784 CT colonography trials in DICOM format. Bejarano et al. [10] used a 16-slice CT scanner to detect squamous cell carcinoma in 31 patients, employing the HNSCC-3DCT-RT dataset, which includes high-resolution 3D fan-beam CT scans collected before, during, and after treatment. Simpson et al. [11] focused on segmentation algorithm development and evaluation with the MSD T10 dataset, presented at the 10th Medical Segmentation Decathlon14, featuring images in NIfTI format.

With the rise of artificial intelligence methods, recent studies have employed machine learning to detect vertebral disorders from images. For example, Duan et al. [12] divided a dataset of 280 patients into training and validation sets,

creating three models: a deep learning model using CT and clinical data, a radiomics model, and a combined model. They evaluated these models using receiver operating characteristic (ROC) curves, area under the curve (AUC), and accuracy (ACC), also examining the correlation between Rad features and DCNN features [12]. Harmon et al. [13] created the COVID-19 sub-dataset, merging DenseNet and AH-Net convolutional neural networks in chest tomography to identify COVID-19 pneumonia, utilizing chest CT images from 632 COVID-19 patients. Pu et al. [14] conducted a retrospective study on patients from 2018 to 2020, assessing vertebral bone quality scores and CT-based Hounsfield unit values from MRI and CT, respectively, and evaluating predictive performance with ROC curves. Deng et al. [15] developed the VerSe'19,20 dataset, segmenting vertebral bones using the U-Net method, with over 11,100 labeled vertebral bones and 1,005 CT volumes.

Langerhuizen et al. [16] reviewed the applications of AI in orthopedic trauma imaging, systematically examining databases like PubMed, Embase, and Cochrane, comparing models such as VGG VNN, U-Net, DenseNet, kNN, LDA, and ResNet by pre-training. Roth et al. [17] used deep convolutional networks to automatically detect posterior element fractures in vertebral CT, achieving an AUC of 0.857 with sensitivity of 71% or 81% at 5 or 10 false positives per patient, respectively. Acheson et al. [18] identified cervical vertebral fractures in 49 out of 160 patients using high-resolution, thin-slice CT alongside conventional radiographs, highlighting the superior detection rates of CT. Muehlematter et al. [19] used tissue analysis and machine learning to identify vertebral bones at risk of compression fractures, classifying vertebrae as "stable" or "unstable" based on sequential scans. Ahammad et al. [20] developed a hybrid CNN-based method for segmenting and classifying vertebral injury data, achieving 96% classification success. Yoo et al. [21] analyzed data from Korean postmenopausal women, using SVM, RF, ANN, and LR models to predict osteoporosis risk, achieving an AUC of 0.827 with 76.7% accuracy. Fang et al. [22] used U-Net for automatic vertebral segmentation and DenseNet-121 for BMD calculation in a dataset of 1449 patients, correlating automatic and manual segmentation results. Bar et al. [23] presented an automated method for detecting vertebral compression fractures on CT, using CNN and RNN networks. Atherya et al. [24] developed a method for spinal CT image segmentation, demonstrating performance with reference standard comparisons and Dice coefficient evaluations. Murata et al. [25] tested a dataset of 300 images for vertebral fractures, resizing images to reduce complexity and evaluating segmentation accuracy.

From the literature, it is evident that disease detection from CT images is often conducted semi-automatically or automatically. Semi-automatic methods involve physician assistance, whereas automatic methods do not. This study employs an automatic method. The proposed method differs from existing studies in several ways:

•To improve image quality, the Wiener transform, previously unused for enhancing vertebral CT images, was applied to the dataset. An example image post-transformation is provided in the study.

•Segmentation and labeling were performed using CCL with the U-Net convolutional neural network model, achieving a Dice score of 100%. This high score indicates no data loss after enhancement, segmentation, and labeling.

•The study evaluated fractures in all vertebrae, focusing on

non-traumatic fracture types.

•Radiomics and logarithmic slide methods were used for feature extraction and selection, a first in detecting non-traumatic vertebral bone compression fractures.

•The CNN classification algorithm was adapted to the original dataset, classifying patients as with or without traumatic vertebral bone fractures.

•Performance measurements such as loss score, accuracy, and AUC-ROC curve were provided to ensure data analysis reliability in healthcare.

3. MATERIAL AND METHOD

After obtaining the approval of the ethics committee to collect retrospective patient data from Istanbul University-Cerrahpasa Faculty of Medicine, Department of Radiology, studies were initiated for the detection of vertebral fractures. In the detection of vertebral bone fractures, the raw data set was first preprocessed with the Wavelet transform. It was then segmented with the U-Net network and its labels were extracted with CCL. The location of the vertebral bone fractures in each patient was determined and reported by our physicians. Upon this determination, the features of the vertebral bone fractures were extracted with the radiomics library. The features suitable for the three-dimensional image were selected and classified with a special convolutional neural network.

3.1 Data set

The data set was created from 6085 vertebral bone images with the permission of the ethics committee. Vertebral series of patients in 118200 DICOM format were reviewed retrospectively. In the raw data set consisting of 197 patient data, each patient's report was taken as a reference. These reports were given by the physicians working in the Radiology Department of Istanbul University-Cerrahpasa Faculty of Medicine. In the reports, the severity of the fracture with the fractured vertebral region was stated in the vertebrae of 100 patients. In the reports of the remaining 97 patients, it was stated that there were no fractures. Sagittal, axial, and coronary images of the patients are displayed with auxiliary software.

The data set was created under ethical permissions, keeping patient information confidential and considering whether the patient's spine is fractured or not. Among the cases with fractures, 60 were adult females aged between 55 and 96, and 40 were adult males aged between 30 and 91. Among the patients without fractures, 40 were adult males aged between 25 and 89, and 57 were females aged between 23 and 94.

Folders consisting of DICOM image series of each patient were converted to NIfTI (Neuroimaging Informatics Technology Initiative) format. NIfTI files were saved as a file format from which the patient's personal information was extracted. Thus, it was possible to evaluate patient images anonymously. Patients are numbered up to 197 in a series with the patient's pre-name. A sequential and regular file structure has been created.

Figure 1 shows examples of sagittal, coronary, and axial images of patients 172, 197, and 1 taken from the data set.

It is seen that it is difficult to make a quantitative determination of the fracture from the sagittal vertebra images given in Figure 1. Trying to get the best angle in the data DICOM series can be considered as a separate challenge.

Although the applications used for manual analysis of images are advanced, capturing the desired angle and examining all series reduces the possibility of stable interpretation.

| Sagittal | Coroner | Axial |
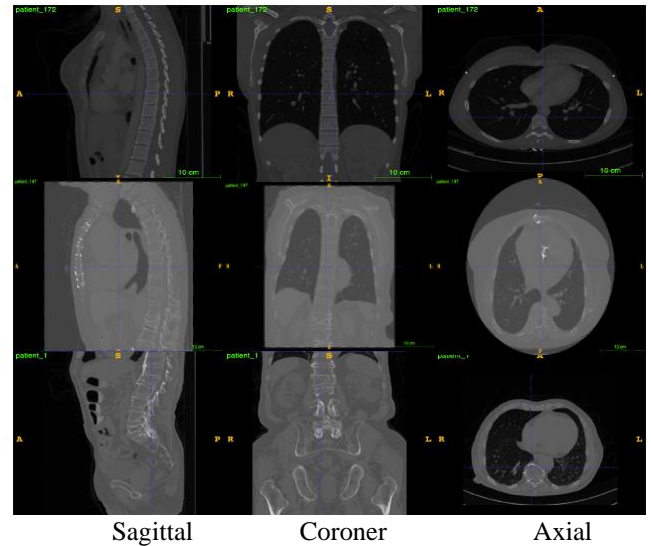
Figure 1. Patient_172, patient_197, patient_1 sagittal, coronary, and axial image examples

3.2 Preprocessing

Preprocessing is one of the important stages of image data that affects the classification result. In addition, improved images contribute to the diagnosis of diseases by physicians who evaluate vertebral disorders in hospitals. At this stage, wiener filtering and deconvolution methods were applied to the patient images.

The Wiener method is the optimal fixed linear enhancement method for images distorted by additional noise and blur. The calculation of the Wiener filter assumes that the signal and noise processes are quadratic stationary. For this, only zero-mean noise processes are based [26].

In Eq. (1), when a stationary signal is corrupted by some other additional and unrelated signal (referred to as noise), the pattern can be expressed as g; where g is a distorted signal, f is a pristine signal, and n is the total noise. The restoration filter that finds the optimal linear estimate can be represented as:

$$g = f + n \; RPD_{CV} = \; SD/RMSE_{cv} \qquad (1)$$

$$W(u) = \frac{|F(u)|^2}{|F(u)|^2 + |N(u)|^2} \qquad (2)$$

In Eq. (2), $|F(u)|^2$ is the three-dimensional power spectrum of the signal or pristine image volume, and $|N(u)|^2$ is the three-dimensional power spectrum of the noise, that is, the reverberation intensity. In Figure 2, sample u and output $W(u)$ for a Wiener image are shown.

$W(u)$ outputs were obtained for each patient by applying the Wiener filter to each u input. An example obtained by applying the Wiener filtering method on patient images is given in Figure 2. After image enhancement, the volume of the fractured vertebral bone increased, resulting in a better fractured vertebral bone for quantitative observation. This stage contributed to the semi-automatic evaluation.
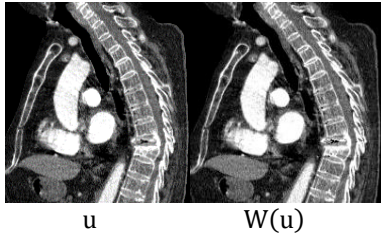
Figure 2. Wiener image input and output examples

3.3 Segmentation and labeling

Segmentation and labeling operations are generally used to classify any object or part in an image. In our study, the vertebral bones in the patient vertebral images were segmented. For this purpose, convolutional U-Net segmentation, and labeling method, which can be used effectively from segmentation and labeling methods, has been applied.
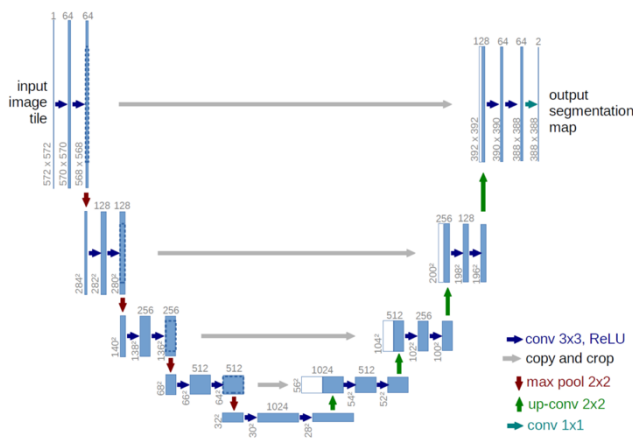

Figure 3. U-Net architecture [27]

The primary motivation behind U-Net was to address the challenges associated with semantic segmentation, where the goal is to classify each pixel in an image into one of several predefined classes or categories. Traditional CNN architectures for image classification, such as AlexNet or VGG, do not directly apply to segmentation tasks because they involve downsampling the spatial resolution of the input image, which can result in the loss of spatial information crucial for precise segmentation. U-Net was designed to overcome this limitation by preserving spatial information through a symmetric encoder-decoder architecture. U-Net architecture used in various applications might consist of 5 convolutional blocks in both the encoder and decoder, resulting in a total network 20 layers. Data for the U-Net model is divided into training and validation. In Figure 3, the U-Net network architecture with a U view is given [27].

The U-Net architecture in Figure 3 shows the jump links connecting the encoder feature map to the decoder, in the direction of the gray arrows to the right, after the biomedical image is input; this helps the inclines flow backward for enhanced training. The lilac arrows to the right represent the Rectifier (ReLu) function that trains the neural network. Downward red arrows indicate max pooling. In this layer, max pooling is applied to the feature matrices. Upward green arrows denote the up-conv 2×2 operation, which causes image pixels to fold using the nearest neighbor. Blue arrows to the right, 1×1 conv i.e. 1 to 1-dimensional conv maps an input

pixel with all its channels to an output pixel, not looking at anything around it. After these stages, a segmented predicted biomedical image is formed.

In Figure 4, the output is obtained after applying the biomedical image input and the U-Net model.
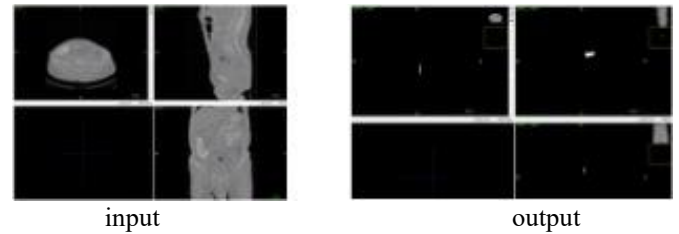
Figure 4. Wiener image input and output examples

Fractured vertebral bone labels of patient_1 data, whose sagittal, axial, and coronal images are shown in Figure 4, were created with CCL. Here, the sagittal view of the fractured vertebral bone is separated from the coronal and axial view. It is possible to use the sagittal image to detect the fracture of the vertebral bone in the image.

In the sagittal input image for patient_1 in Figure 5, the fractured vertebral bone was marked by our physicians.
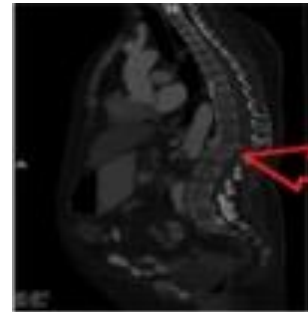
Figure 5. Sagittal view of the fractured vertebral bone

In Figure 5, it is seen that the vertebra L1 (Lumbar1) vertebral bone is fractured, and its structure has changed. In retrospective physician reports, fracture detection was made by specifying the vertebral bone numbers in this way. In the proposed model, the fracture information is known beforehand. This indicates that the model will be evaluated in the supervised learning group.

Figure 6 shows the sagittal, coronary, and axial angles of the labeled image of the fractured vertebral bone.

Figure 6. Labeling angles of patient 1's fractured vertebral bone

In Figure 6, the fractured vertebral bone label of patient_1 was evaluated together in 3 aspects. The evaluation was made for each patient. It is important that the removed vertebral bone labels match the input of the vertebral bone without loss.

Removing the labels of the fractured vertebral bone without loss has provided successful segmentation and classification results.

In Figure 7, the graph of the U-Net segmentation model including the training and validation accuracy and the losses in the same period is given.

In Figure 7, the accuracy of the segmentation model is over 94% for the validation set and over 90% for the training set. In addition, the segmentation model loss value was below 25% for the training set and below 20% for the validation set. Thus, labels were created with minimum loss, and segmentation was carried out.
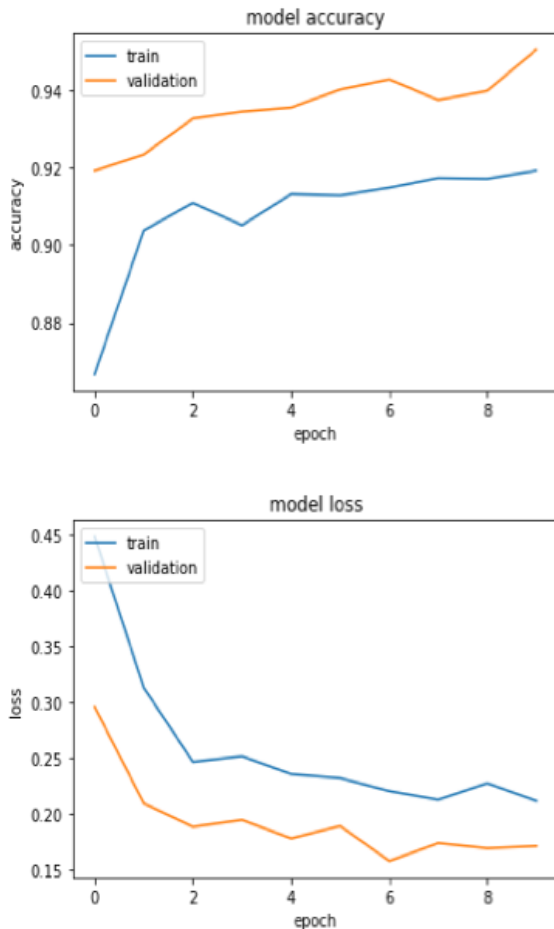

Figure 7. U-Net model accuracy and loss graphs

3.4 Feature extraction and selection

Feature extraction is used to determine the characteristics of the target fragment for disease detection in the biomedical image. Correct identification of these features contributes significantly to the success of disease detection [28].

In this study, feature extraction was performed using the radiomics method. 593 features of segmented and labeled fractures were extracted. Features: First-order statistics can be listed as 3-dimensional shape-based, 2-dimensional shape-based, gray-level co-occurrence matrix. In addition, the features of the radiomics method in the back level run length, gray level dimension region matrix, adjacent grayscale difference matrix, and gray level dependency matrix groups were also used.

In Figure 8, morphological feature extraction options are given for use in radiomics method computed tomography

images [29].

In Figure 8, the radimoics method offers 3 options based on histogram, shape, and texture for the computed tomography image. Labels created with CCL in vertebra CT images and features were extracted with the shape-based method.

The use of logarithmic scale for hyperparameter tuning, such as lambda in regularization techniques, is a common practice in feature selection and model selection. Logarithmic lambda, expressed in logarithmic scale, offers several advantages in the context of hyperparameter optimization. Hyperparameters like lambda often vary over a wide range of values. For instance, lambda values can range from 0.1 to 1000 [30].

In Figure 9, success rates with logarithmic lambda method [31] are given for the selection of radiomics properties. Using a logarithmic scale ensures that this wide range is represented more uniformly, facilitating comparison of regularization effects across different values [32]. Logarithmic scale helps in achieving numerical stability, especially during optimization procedures like gradient descent. Large differences between values can lead to numerical instability, which logarithmic scale mitigates by compressing large values. Logarithmic scale is scale-invariant, meaning that the importance of the starting value is diminished. For example, a lambda value of 0.1 and 10 represent the same level of regularization when transformed logarithmically.

In the graph given in Figure 9, the lambda value of the accuracy ratio reached the desired value at -3. In this case, for the -3 value, the coefficient (dice score) ratio in Figure 10 was summed up at -3 lambda values. There are 537 features with the -3-lambda value providing the best success rate. The selection of 537 features according to accuracy and their intersection with the Dice Coeggicient score gained importance.

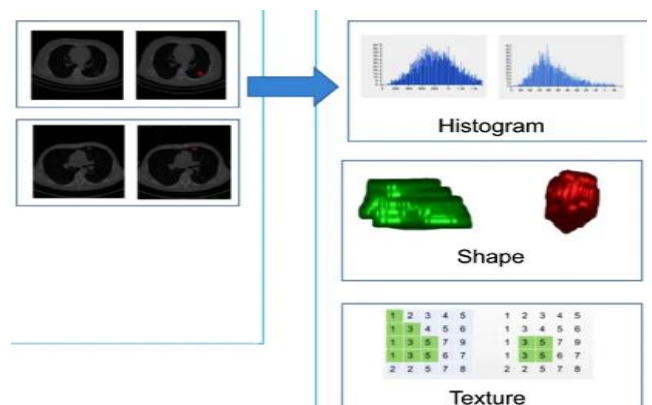Dice Coefficient and lambda value profiles of 593 radiomics features are given in Figure 10.

Figure 8. Radiomics image feature extraction options [29]
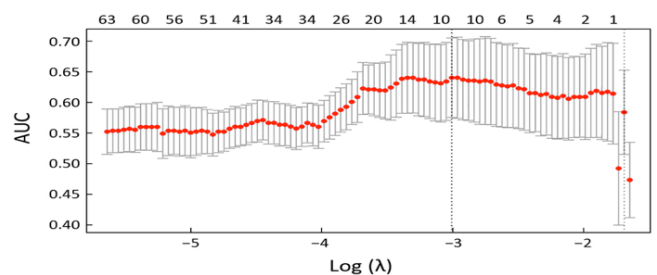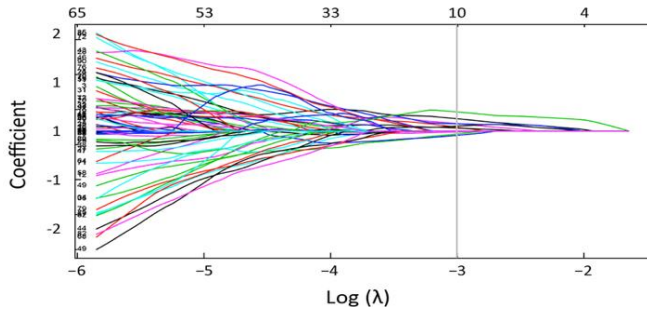

Figure 9. Selection and accuracy of radiomics properties by logarithmic lambda method [33]

Figure 10. Selection of radiomics properties by logarithmic lambda method and coefficient (dice score) ratio

The features at the intersections of Log(lambda)=-3, auc=0.65, and coefficient=1, concentrated from the graphs in Figure 9 and Figure 10, were used. Of these features, the number of lambda=-3 is selected as 537. Thus, before classification, 537 features were selected from 593 features, and the feature extraction and selection process was completed.

Segmentation results were evaluated with the Dice evaluation criterion given in Eq. (3).

$$Dice\ skoru(P,T) = \frac{2 * |P \cap T|}{|P| + |T|} \qquad (3)$$

In Eq. (3), for each vertebral bone region, *P* represents the model estimate and *T* represents the baseline truth labels. The Dice Coefficient score given in Formula 2 was calculated after subtracting the features in Figure 9 and Figure 10. And the Dice score is 100%. Performing the segmentation in an error-free and lossless manner has ensured that there is no loss of patient data.

3.5 Classification

Classification is the process of categorizing a disease on radiological images according to its characteristics [34]. Machine learning methods give effective results in disease detection from biomedical images. In this study, a special convolutional neural network, one of the classification methods, was used. The private convolutional neural network is compared with the MobileNetV3, EfficientNetB1, VGG16, and ResNet50V2 classification methods. The special convolutional neural network was chosen because of the success factor.

In Figure 11, the 3D image input of the custom convolutional neural network and the estimation diagram of this image are given.
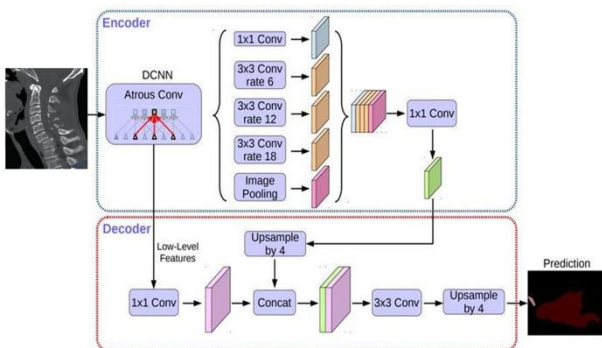
Figure 11. Custom convolutional neural network diagram [32] input and prediction output

Atrous Conv in Figure 11 enables the segmentation of objects at multiple scales. This allows for increased accuracy by considering different scales. The model consists of an encoder module that reduces feature maps incrementally and captures higher semantic information and a decoder that recovers spatial information. The overfitting preventive layers added to the model are added to the target class by combining them in the 3D image with the combination of the features after feature extraction. In this way, the images of patient data were classified as fractured and non-fractured.

CustomCNN has an architecture customized for a specific medical imaging task. This allows the model to be designed and optimized to better suit a particular problem. CustomCNN designed for a specific non-traumatic vertebral fracture detection task. This allows the model to identify and use the most important features from the input data.

CustomCNN is trained on a specific dataset and equipped with features specific to this dataset. This allows the model to be trained to better perform the target task. And CustomCNN can provide higher accuracy and performance compared to other models. This indicates that the model is more effective at a particular medical imaging task.

4. PROPOSED METHOD

In the proposed model, the folders containing the computed tomography DICOM image series of the patients taken from the Radiology Department of Istanbul University-Cerrahpaşa Faculty of Medicine were converted into NIfTI format. In the preprocessing stage, the images were filtered with the wiener method. Vertebral bone labeling was performed using CCL. Segmentation of the vertebral bones in the images was performed with the U-Net network. A special CNN network, which is a supervised learning method adapted to the data set, was developed and the patients were classified as with and without vertebral fracture. Classification performance was evaluated with Loss, ROC-AUC, and accuracy criteria. The rough code of the proposed model is given in Figure 12.

```
function main()
    input patientFolder
    loop i from 1 to 197
        patientNIfTIFolder = patientFolder
    end loop
    loop i from 1 to 197
        wienerTransform = patientNIfTIFolder
    end loop
    CclLabeling = patientNIfTIFolder
    loop i from 1 to 197
        UnetSegmentation = CclLabeling + patientNIfTIFolder
    end loop
    CustomCNN = patientSF + patientNSF
end function
```

Figure 12. Pseudocode of the proposed model main function

In Figure 12, the patients folder is entered as patientFolder. With a subconversion method, the patient DICOM series in the file were converted to NIfTI format and saved in the patientNIfTIFolder folder. Vertebra labels in patient NIfTI files from the patientNIfTIFolder folder were created and

saved in the CclLabeling folder. U-Net segmentation was performed using tags and patient NIfTI files together and saved in the UnetSegmentation folder. With the custom convolutional neural network (CustomCNN) classification method, patients were classified as patient with fractured vertebral bone (patientSF) and patient without fractured vertebral bone (patientNSF).

In Figure 13, the diagram containing all the phases of the model is given.
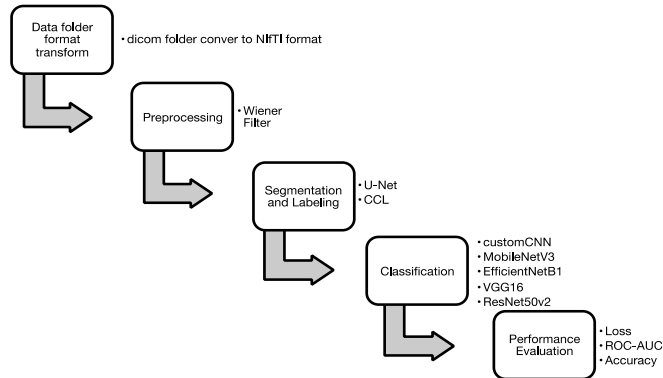

Figure 13. Suggested model diagram

In the first stage of Figure 13, 197 patient files in the data set were converted and saved by the transformation method. In the second stage, the preprocessing stage, the data images were improved by the winer filtering method. In step 3, segmentation and labeling were performed using U-Net convolutional neural network and associated component labeling. In the 4th step, the data set was classified using the MobileNetv3, EfficientB1, VGG16, ResnET50v2 methods, and the recommended customCNN method. Loss, ROC-AUC, and Accuracy performance criteria results of the proposed model and other classification methods were obtained in the 5th step.

5. FINDINGS

At the stage of segmentation of the vertebral bones in the data set, segmentation of the vertebral bones with a 100% dice overlap index score was performed using the U-Net segmentation method. An example of a vertebra with segmented vertebral bones is given in Figure 14.

In Figure 14, the color images of the vertebral bone can be clearly read because of the segmentation of the vertebral bone of a sample patient through Un-Net. The segmentation output given with the vertebral bone to the marked fracture, and the detection of the fractured T1 vertebral bone without wasting time could also be achieved quantitatively. In the segmentation performed with U-Net without data loss, the Dice Coefficient

score was obtained as 100%. All vertebral bones in the patients' vertebral images were segmented.
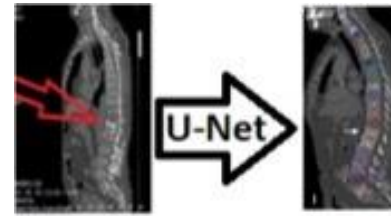
Figure 14. Fractured vertebral bone image input and U-Net segmentation output of the patient

Creating a hold-out test set is used to evaluate whether a model performs successfully on training data. Typically, it consists of data not previously seen by the model, often different from the training set [35]. The initial step involves partitioning the available dataset, reserving a portion for training and allocating the rest for the extension test set. Typically, data is used for training 80%, leaving a smaller portion for the extension test set 20%.

After splitting the data, samples for the extension test set are randomly selected. This random selection ensures that the extension test set has a different distribution from the training set, facilitating better evaluation of the model's generalization ability [36].

In classification, it's essential to create an extension test set containing a balanced number of samples from each class. This enables evaluating the performance of the model on each class separately, mitigating the impact of class imbalances on performance evaluation [37].

The segmented fractured vertebral bones were classified by a special convolutional neural network, MobileNetV3, EfficientNetB1, VGG16 and ResNet50V2 classification methods. In Table 1 and Table 2, the results of the vertebral fracture classification evaluation criteria with a special convolutional neural network, MobileNetV3, EfficientNetB1, VGG16, and ResNet50V2 models are given.

As seen in Table 1, in addition to the low loss value and high ROC-AUC value in the EfficientNetB1 model, the accuracy value of over 86.7% in the special convolutional neural network model stands out.

The Confidence Interval (CI) values in Table 2 indicate the level of precision for a specific metric for each model. The lower and upper bounds of the CI show the range of values within which the model's performance could lie with 95% confidence. These CI values specify how reliable the performance of each model is on a particular metric, playing a crucial role in evaluating the reliability of the model in the decision-making process.

Figure 15 shows the evaluation criteria graph of the classification models. In the chart, sub-graphs are combined. Ease of comparison and evaluation is provided.

Table 1. Evaluation results of classification models

Evaluation Type	CustomCNN	Mobile NetV3	EfficientNetB1	VGG16	ResNet50V2
Loss	0.409174	0.511587	0.396508	0.431550	0.515297
ROC-AUC	0.836110	0.765697	0.868855	0.865474	0.836928
Accuracy	0.867072	0.771503	0.842745	0.826238	0.764553
Sensitivity	0.894054	0.750346	0.848	0.802532	0.750395
specificity	0.848235	0.789194	0.842176	0.842634	0.787106

**Table 2.** Confidence intervals (CI) of classification models

Model	Metric	CI (95%)
CustomCNN	Loss	(0.396, 0.421)
	ROC-AUC	(0.825, 0.847)
	Accuracy	(0.840, 0.894)
	Sensitivity	(0.883, 0.905)
	Specificity	(0.837, 0.859)
MobileNetV3	Loss	(0.495, 0.527)
	ROC-AUC	(0.751, 0.779)
	Accuracy	(0.742, 0.801)
	Sensitivity	(0.735, 0.765)
	Specificity	(0.774, 0.804)
EfficientNetB1	Loss	(0.382, 0.410)
	ROC-AUC	(0.857, 0.879)
	Accuracy	(0.815, 0.871)
	Sensitivity	(0.833, 0.863)
	Specificity	(0.827, 0.857)
VGG16	Loss	(0.416, 0.446)
	ROC-AUC	(0.853, 0.877)
	Accuracy	(0.797, 0.855)
	Sensitivity	(0.787, 0.817)
	Specificity	(0.827, 0.857)
ResNet50V2	Loss	(0.499, 0.531)
	ROC-AUC	(0.825, 0.847)
	Accuracy	(0.733, 0.796)
	Sensitivity	(0.735, 0.765)
	Specificity	(0.772, 0.802)

Figure 15 shows the ROC-AUC, accuracy and loss curves of the training set and the test set. When the accuracy rate is considered, the custom convolutional neural network (CustomCNN) fracture classification method stands out with the highest success rate of over 86%. When ROC-AUC success evaluation is considered, the EfficientNetB1 classification model stands out with the highest success rate above 86% and the lowest loss rate below 40%.

Integrating assessment methods into clinical decision support systems can provide healthcare professionals with more comprehensive information about treatment options [38]. For example, an algorithm used for diagnosing or planning treatment for a disease can produce more accurate results by taking into account data obtained from different assessment methods.

Developing customized Treatment approaches, assessment methods can be used to better understand patients' unique needs and risk profiles [39]. This information can be used to create customized treatment plans, which can have a positive impact on the effectiveness of treatment and patient outcomes.

Utilization in patient tracking and management, it can also be used for patient tracking and management [40]. For example, assessment methods can be used regularly to monitor a patient's response to treatment, and this information can lead to the reorganization or improvement of treatment plans.
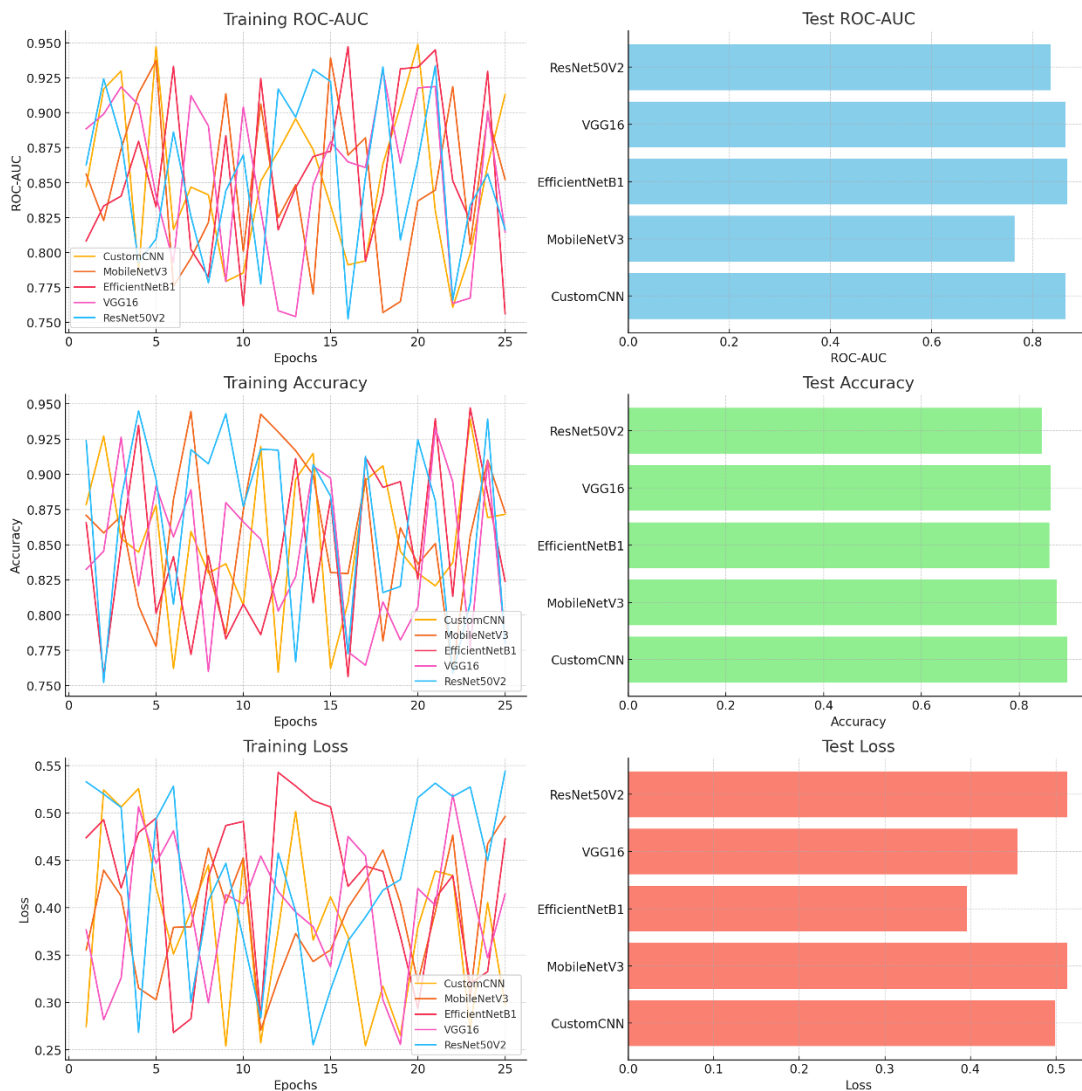
Figure 15. Evaluation graph of classification models

6. CONCLUSIONS

The use of retrospective CT image data of patients with the permission of the ethics committee contributes to patients, the health system, and scientific studies. Since it is not possible to directly use the first images taken from the CT device, file conversion has been applied for processing in the computer environment. Care was taken to find the fracture reports of the patients who made up the data set. Radiology, orthopedics, general surgery etc. To detect vertebral fracture, which is one of the leading vertebral disorders inwards, data from both patients with and without fractures were needed. For this reason, the data set was composed of two types as fractured vertebrae and non-fractured vertebrae.

Better images were presented to physician reviews by using the improvement method in the dataset preprocessing phase. The vertebral bones in the patient images were segmented by processing a raw data set, labeling, and coloring. In feature extraction, feature selection, and classification, radiomics was applied to a custom model dataset by adapting logarithmic lambda and CNN. Thus, it was seen that efficient results were obtained in the detection of vertebral fracture of the model by using 5 different methods in improvement, segmentation, feature selection, feature extraction and classification.

Future research should investigate how the model performs in different geographical regions and in multicenter studies. This exploration can provide insights into the generalizability of the model. Additionally, enhancing data integrity and quality through further preprocessing and scrutiny is essential. Moreover, integrating data from diverse sources can enhance the model's performance. Further research can delve into the clinical applications of the model, particularly its integration into clinical decision support systems and its utility for healthcare professionals. Furthermore, there is a need for more studies examining the impact of the model on patient outcomes and conducting cost-effectiveness analyses on its potential effects on healthcare systems.

Take into account these limitations and potential avenues, future research presents a significant opportunity to advance in this field. Such studies can contribute to a better understanding of the role of artificial intelligence-based models in clinical practice and help improve access to treatment and outcomes for patients.

Considering classification and segmentation criteria results, our physicians will be able to benefit from segmentation images and provide decision support for estimation. In the datasets to be created in future studies, osteoporosis disease evaluation can be made by adding the osteoporosis values of the patients. The data set can be expanded with data enlargement methods. The proposed data set and vertebral fracture detection model will lead to new studies in this field.

REFERENCES

[1] Adams, M., Chen, W., Holcdorf, D., McCusker, M.W., Howe, P.D., Gaillard, F. (2019). Computer vs human: Deep learning versus perceptual training for the detection of neck of femur fractures. Journal of Medical Imaging and Radiation Oncology, 63(1): 27-32. https://doi.org/10.1111/1754-9485.12828

[2] Tomita, N., Cheung, Y.Y., Hassanpour, S. (2018). Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. Computers in Biology and Medicine, 98: 8-15. https://doi.org/10.1016/j.compbiomed.2018.05.011

[3] Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Webster, D.R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22): 2402-2410. https://doi.org/10.1001/jama.2016.17216

[4] Brett, A., Miller, C., Hayes, C.W., Krasnow, J., Ozanian, T., Abrams, K., Block, J.E., van Kuijk, C. (2009). Development of a clinical workflow tool to enhance the detection of vertebral fractures: Accuracy and precision evaluation. Spine, 34(22): 2437-2443. http://doi.org/10.1097/BRS.0b013e3181b2eb69

[5] Aggarwal, V., Maslen, C., Abel, R.L., Bhattacharya, P., Bromiley, P.A., Clark, E.M., Poole, K.E. (2021). Opportunistic diagnosis of osteoporosis, fragile bone strength and vertebral fractures from routine CT scans; a review of approved technology systems and pathways to implementation. Therapeutic Advances in Musculoskeletal Disease, 13: 1759720X211024029. https://doi.org/10.1177/1759720x211024029

[6] Tompson, J.J., Jain, A., LeCun, Y., Bregler, C. (2014). Joint training of a convolutional network and a graphical model for human pose estimation. Advances in Neural Information Processing Systems. https://doi.org/10.48550/arXiv.1406.2984

[7] Kelly, M.A., McCabe, E., Bergin, D., Kearns, S.R., McCabe, J.P., Armstrong, C., Carey, J.J. (2021). Osteoporotic vertebral fractures are common in hip fracture patients and are under-recognized. Journal of Clinical Densitometry, 24(2): 183-189. https://doi.org/10.1016/j.jocd.2020.05.007

[8] Yi, W., Zhao, J., Tang, W., Yin, H., Yu, L., Wang, Y., Tian, W. (2023). Deep learning-based high-accuracy detection for lumbar and cervical degenerative disease on T2-weighted MR images. European Spine Journal, 2(11): 3807-3814. https://doi.org/10.1007/s00586-023-07641-4

[9] Johnson, C.D., Chen, M.H., Toledano, A.Y., Heiken, J.P., Dachman, A., Kuo, M.D., Limburg, P.J. (2008). Accuracy of CT colonography for detection of large adenomas and cancers. New England Journal of Medicine, 359(12): 1207-1217. https://doi.org/10.1056/NEJMoa0800996

[10] Bejarano, T., De Ornelas Couto, M., Mihaylov, I. (2018). Head-and-neck squamous cell carcinoma patients with CT taken during pre-treatment, mid-treatment, and post-treatment Dataset. The Cancer Imaging Archive, 10: K9. https://doi.org/10.7937/K9%2FTCIA.2018.13UPR2XF

[11] Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Cardoso, M.J. (2019). A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063. https://doi.org/10.48550/arXiv.1902.09063

[12] Duan, S., Hua, Y., Cao, G., Hu, J., Cui, W., Zhang, D., Liu, B. (2023). 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. European Journal of Radiology, 110899. https://doi.org/10.1016/j.ejrad.2023.110899

[13] Harmon, S.A., Sanford, T.H., Xu, S., Turkbey, E.B., Roth, H., Xu, Z., Turkbey, B. (2020). Artificial

intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. Nature Communications, 11(1): 4080. https://doi.org/10.1038/s41467-020-17971-2

[14] Pu, X., Wang, X., Ran, L., Xie, T., Li, Z., Yang, Z., Zeng, J. (2023). Comparison of predictive performance for cage subsidence between CT-based Hounsfield units and MRI-based vertebral bone quality score following oblique lumbar interbody fusion. European Radiology, 3(12): 8637-8644. https://doi.org/10.1007/s00330-023-09929-x

[15] Deng, Y., Wang, C., Hui, Y., Li, Q., Li, J., Luo, S., Zhou, S.K. (2021). Ctspine1k: A large-scale dataset for spinal vertebrae segmentation in computed tomography. ArXiv, abs/2105:14711. https://doi.org/10.48550/arXiv

[16] Langerhuizen, D.W., Janssen, S.J., Mallee, W.H., Van Den Bekerom, M.P., Ring, D., Kerkhoffs, G.M., Doornberg, J.N. (2019). What are the applications and limitations of artificial intelligence for fracture detection and classification in orthopaedic trauma imaging? A systematic review. Clinical Orthopaedics and Related Research, 477(11): 2482-2491. https://doi.org/10.1097/corr.0000000000000848

[17] Roth, H.R., Wang, Y., Yao, J., Lu, L., Burns, J.E., Summers, R.M. (2016). Deep convolutional networks for automated detection of posterior-element fractures on spine CT. In SPIE Medical Imaging 2016: Computer-Aided Diagnosis, San Diego, California, US, pp. 165-171. https://doi.org/10.1117/12.2217146

[18] Acheson, M.B., Livingston, R.R., Richardson, M.L., Stimac, G.K. (1987). High-resolution CT scanning in the evaluation of cervical spine fractures: Comparison with plain film examinations. American Journal of Roentgenology, 148(6): 1179-1185. https://doi.org/10.2214/ajr.148.6.1179

[19] Muehlematter, U.J., Mannil, M., Becker, A.S., Vokinger, K.N., Finkenstaedt, T., Osterhoff, G., Guggenberger, R. (2018). Vertebral body insufficiency fractures: Detection of vertebrae at risk on standard CT images using texture analysis and machine learning. European Radiology, 29: 2207-2217. https://doi.org/10.1007/s00330-018-5846-8

[20] Ahammad, S.H., Rajesh, V., Rahman, M.Z.U., Ekuakille, A. (2020). A hybrid CNN-based segmentation and boosting classifier for real time sensor spinal cord injury data. IEEE Sensors Journal, 20(17): 10092-10101. https://doi.org/10.1109/JSEN.2020.2992879

[21] Yoo, T.K., Kim, S.K., Kim, D.W., Choi, J.Y., Lee, W.H., Park, E.C. (2013). Osteoporosis risk prediction for bone mineral density assessment of postmenopausal women using machine learning. Yonsei Medical Journal, 54(6): 1321-1330. https://doi.org/10.3349/ymj.2013.54.6.1321

[22] Fang, Y., Li, W., Chen, X., Chen, K., Kang, H., Yu, P., Li, S. (2021). Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks. European Radiology, 31(4): 1831-1842. https://doi.org/10.1007/s00330-020-07312-8

[23] Bar, A., Wolf, L., Amitai, O.B., Toledano, E., Elnekave, E. (2017). Compression fractures detection on CT. In Medical Imaging 2017: Computer-Aided Diagnosis. Orlando, United States, pp, 1013440. https://doi.org/10.1117/12.2249635

[24] Athertya, J.S., Kumar, G.S. (2016). Automatic segmentation of vertebral contours from CT images using fuzzy corners. Computers in Biology and Medicine,

72: 75-89. https://doi.org/10.1016/j.compbiomed.2016.03.009

[25] Murata, K., Endo, K., Aihara, T., Suzuki, H., Sawaji, Y., Matsuoka, Y., Yamamoto, K. (2020). Artificial intelligence for the detection of vertebral fractures on plain spinal radiography. Scientific Reports, 10(1): 1-8. https://doi.org/10.1038/s41598-020-76866-w

[26] Chen, J., Benesty, J., Huang, Y., Doclo, S. (2006). New insights into the noise reduction Wiener filter. IEEE Transactions on Audio, Speech, and Language Processing, 14(4): 1218-1234. https://doi.org/10.1109/TSA.2005.860851

[27] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, pp, 3-11. https://doi.org/10.48550/arXiv.1807.10165

[28] Caby, B., Kieffer, S., De Saint Hubert, M., Cremer, G., Macq, B. (2011). Feature extraction and selection for objective gait analysis and fall risk assessment by accelerometry. Biomedical Engineering Online, 10(1): 1-19. http://doi.org/10.1186/1475-925X-10-1

[29] Liu, Q., Jiang, P., Jiang, Y., Ge, H., Li, S., Jin, H., Li, Y. (2019). Prediction of aneurysm stability using a machine learning model based on PyRadiomics-derived morphological features. Stroke, 50(9): 2314-2321. https://doi.org/10.1161/strokeaha.119.025777

[30] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. Nature, 521(7553): 436-444. https://doi.org/10.1038/nature14539

[31] Wu, L., Wang, C., Tan, X., Cheng, Z., Zhao, K., Yan, L., Liang, C. (2018). Radiomics approach for preoperative identification of stages I− II and III− IV of esophageal cancer. Chinese Journal of Cancer Research, 30(4): 396. https://doi.org/10.21147/j.issn.1000-9604.2018.04.02

[32] Bergstra, J., Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13: 281-305. https://doi.org/10.5555/2503308.2188395

[33] Türkmen, M., Orman, Z. (2023). Improvement and repair methods of biomedical images used in the diagnosis of osteoporosis. In Investigations in Pattern Recognition and Computer Vision for Industry 4.0, IGI Global, pp. 68-80.

[34] Khan, S.H., Abbas, Z., Rizvi, S.D. (2019). Classification of diabetic retinopathy images based on customised CNN architecture. In 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, pp. 244-248. https://doi.org/10.1109/AICAI.2019.8701231

[35] Brownlee, J. (2016). XGBoost with python: Gradient boosted trees with XGBoost and scikit-learn. Machine Learning Mastery, 115.

[36] Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, 1-758.

[37] Molnar, C. (2020). Interpretable Machine Learning. Lulu. Press, Inc, p. 318.

[38] Castaneda, C., Nalley, K., Mannion, C., Bhattacharyya,

P., Blake, P., Pecora, A., Suh, K.S. (2015). Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. Journal of Clinical Bioinformatics, 5(1): 1-16. https://doi.org/10.1186/s13336-015-0019-3

[39] Spring, B. (2007). Evidence-based practice in clinical psychology: What it is, why it matters; what you need to know. Journal of Clinical Psychology, 63(7): 611-631. https://doi.org/10.1002/jclp.20373

[40] York, A.S., McCarthy, K.A. (2011). Patient, staff and physician satisfaction: A new model, instrument and their implications. International Journal of Health Care Quality Assurance, 24(2): 178-191. https://doi.org/10.1108/09526861111105121