Automated Vertebral Bone Quality Determination from T1-Weighted Lumbar Spine MRI Data Using a Hybrid Convolutional Neural Network–Transformer Neural Network

Stojšić, Kristian; Miletić Rigo, Dina; Jurković, Slaven

Download date / Datum preuzimanja: **2025-01-20**

Automated Vertebral Bone Quality Determination from T1-Weighted Lumbar Spine MRI Data Using a Hybrid Convolutional Neural Network–Transformer Neural Network

Kristian Stojšić [1], Dina Miletić Rigo [2,3] and Slaven Jurković [1,4,*]

[1] Medical Physics and Radiation Protection Department, University Hospital Rijeka, 51000 Rijeka, Croatia; kristian.stojsic@kbc-rijeka.hr

[2] Clinical Department of Diagnostic and Interventional Radiology, University Hospital Rijeka, 51000 Rijeka, Croatia; dina.miletic.rigo@medri.uniri.hr

[3] Department of Radiology, Faculty of Medicine, University of Rijeka, 51000 Rijeka, Croatia

[4] Department of Medical Physics and Biophysics, Faculty of Medicine, University of Rijeka, 51000 Rijeka, Croatia

[*] Correspondence: slaven.jurkovic@medri.uniri.hr

Abstract: Vertebral bone quality (VBQ) is a promising new method that can improve screening for osteoporosis. The drawback of the current method is that it requires manual determination of the regions of interest (ROIs) of vertebrae and cerebrospinal fluid (CSF) by a radiologist. In this work, an automatic method for determining the VBQ is proposed, in which the ROIs are obtained using a trained neural network model. A large, publicly available dataset of sagittal lumbar spine MRI images with ground truth segmentations was used to train a BRAU-Net++ hybrid CNN–transformer neural network. The performance of the trained model was evaluated using the dice similarity coefficient (DSC), accuracy, precision, recall and intersection-over-union (IoU) metrics. The trained model performed similarly to state-of-the-art lumbar spine segmentation models, with an average DSC value of 0.914 ± 0.007 for the vertebrae and 0.902 for the spinal canal. Four different methods of VBQ determination with automatic segmentation are presented and compared with one-way ANOVA. These methods use different algorithms for CSF extraction from the segmentation of the spinal canal using T1- and T2-weighted image data and applying erosion to the vertebral ROI to avoid a sharp change in SI at the edge of the vertebral body.

Keywords: MRI; CNN; transformer; spine imaging; VBQ; MRI segmentation

1. Introduction

Osteoporosis is a systemic bone disease that causes degeneration of bone tissue microstructures and low bone quality [1,2]. Therefore, the assessment of bone mineral density (BMD) in patients undergoing spinal surgery is important for clinical decision making [3,4]. BMD is determined using dual-energy X-ray absorptiometry (DEXA). This technique is recommended by the World Health Organization (WHO) as the diagnostic tool of choice for osteoporosis screening [5]. The main problem with DEXA and BMD assessment is the relatively low screening rate in the majority of patient groups. Some surveys found that only 44% of spine surgeons routinely obtained DEXA results prior to instrumental fusion, although DEXA screening prior to surgical instrumentation is generally advocated for [6,7].

The importance of assessing bone quality in the preoperative treatment of patients is recognized as a prerequisite for a positive outcome of vertebrosynthesis procedures and the avoidance of postoperative complications. The gold standard for BMD assessment is DEXA. However, it overestimates BMD in patients with degenerative changes in the spine and in patients with excess body weight [8]. In view of this, there are alternative radiological methods to measure BMD using MRI as a non-invasive routine examination prior to surgical procedures. Recent research has shown that it is possible to calculate BMD values

obtained through MRI examination. However, the correlation of BMD obtained by MRI examinations with the values obtained by DEXA needs to be confirmed so that this method can be included in the preoperative treatment of patients undergoing vertebrosynthesis. Deformities of the spine, atherosclerotic changes in the aorta and previous compression fractures of the vertebrae lead to increased X-ray absorption and falsely increase the T value [9,10]. Recent studies show that measuring the fat content in the cancellous bone on a T1-weighted image using MRI is a more sensitive method of assessing BMD and can predict the occurrence of vertebral fracture much more accurately than DEXA [6,11].

The vertebral bone quality (VBQ) score is a new scoring method in magnetic resonance imaging (MRI) that can be used to improve screening for osteoporosis [6,12]. The VBQ score can be determined from standard clinical, non-contrast T1-weighted MR images without additional software. The advantage of this method is that it does not require ionizing radiation, is easy to use, and does not require additional equipment or special clinical protocols. The proposed method measures the signal intensity (SI) of vertebrae and cerebrospinal fluid (CSF) in specific regions of interest (ROIs). The ROIs are usually drawn manually by the radiologist, which can jeopardize the consistency of the results obtained. This can be improved by automating the ROI determination. Therefore, automating the VBQ determination may drastically reduce the time required to yield plausible results. Furthermore, it does not require any additional work from a radiologist, except for final monitoring.

Although the automatic segmentation of anatomical features can be performed using analytical algorithms such as connected threshold or region growing, most of these methods require user input in addition to the medical images. Usually, this input consists of the coordinates of initial seed points from which the algorithm starts the calculation. An advanced method is the use of a trained neural network model for the semantic segmentation of medical images. The use of machine learning and deep learning in conjunction with computer vision has improved the interpretation of medical images and made these techniques an important tool in healthcare [13]. Currently, the segmentation model-based convolutional neural network (CNN) is used for most medical image segmentation applications, with U-Net being one of the most widely used models. The encoder–decoder design of U-Net combines information at multiple scales and improves the performance of the segmentation model. CNN excels at capturing local features but is unable to capture long distance dependencies. After the successful application of transformer-only architecture in the field of medical image segmentation, more and more transformer-based models have been developed to optimize semantic segmentation approaches [14]. MRI segmentation of the lumbar spine has been the subject of intensive research in recent years, with numerous approaches being proposed. The reliable detection of vertebrae from a variety of different MRI protocols and sequences remains a challenge, but at the same time, is crucial for the localization of spinal disease and subsequent treatment planning [15]. The challenge arises from the fact that different vertebrae are similar to each other, while the same vertebrae can look drastically different due to pathological deformities. Recently, reviews of different state-of-the-art methods have become available, allowing researchers to compare different neural network architectures and solutions for the segmentation of lumbar spine images, even across different imaging modalities such as CT, X-ray and MRI [16].

The use of deep learning in spinal imaging has led to many practical solutions and could have a potential future impact on research in the field of spinal imaging. These include image acquisition and reconstruction, image quality enhancement, quantitative image analysis and image interpretation [17]. Quantitative image analysis often employs the localization, labeling and segmentation of spinal structures. Image interpretation and pattern recognition have been used to detect and classify a variety of disorders, namely degenerative diseases of the lumbar spine, scoliosis, spinal tumors, spinal cord compression, cervical spondylosis and osteoporosis [17]. Recent examples of developments in spine image segmentation, detection and classification focus on lumbar spine stenosis

using a CNN-based deep learning model or a CNN integrated with multiple attention mechanisms [18,19].

To follow recent advances in neural network architecture and to test a transformer-based neural network for a medical image segmentation task, a hybrid CNN–transformer network BRAU-Net++ was used for this research [20]. Training a neural network model for medical image segmentation requires a large amount of image data with corresponding semantic segmentation masks. A large, publicly available dataset was used to train the BRAU-Net++ network for vertebral and spinal canal segmentation from sagittal MR images of the lumbar spine [21]. To evaluate the statistical significance of the different VBQ determination methods, 166 image series from 83 studies were randomly selected from the database of the resident hospital.

The main objective of this study is to provide a method for automatic VBQ determination from T1-weighted sagittal MRI images of the lumbar spine. The second aim is to compare different methods of post-processing vertebral and CSF ROI and to determine whether changing the shape of the measured ROI has a statistically significant effect on the overall VBQ score. The advantage of an automated method for VBQ scoring is its ease of application to a large amount of data, making validation and comparison studies more accessible. This is particularly important as the diagnostic value of the VBQ is not well documented and the complementary role of the VBQ score in the assessment of osteoporosis remains to be evaluated [22].

The use of the VBQ score may provide exceptional health benefits by reducing patient disability and shortening the rehabilitation time of patients following surgery. It is well known that osteoporosis leads to poor surgical outcomes and that fragility fractures in orthopedic surgery are more incident in osteoporotic bone. Postoperatively, osteoporosis is associated with an increased risk of implementation failures, extended hospitalization, a higher rate of revision surgery and higher patient mortality. Many patients with undiagnosed osteoporosis have been deprived of treatment, and in this sense, early diagnosis and timely treatment can prevent fragility fractures [23].

2. Materials and Methods

In order to achieve an automatic VBQ determination, semantic segmentation of the required anatomical regions (vertebrae and spinal canal) must first be performed. This can be achieved by training a neural network for the segmentation task using a large dataset of annotated MRI data. First, a large, publicly available dataset is presented along with the required image preprocessing steps. Then, BRAU-Net++, a hybrid CNN–transformer architecture selected for this task, is briefly described and the relevant metrics to evaluate the performance of the neural network are explained. Finally, the original manual method for determining the VBQ and four new automatic methods are presented. Each of these methods uses the segmentation obtained with the trained neural network model and an additional algorithm that separates the CSF from the spinal canal, together with the vertebral body erosion segmentation algorithm.

2.1. Dataset and Preprocessing

A large publicly available dataset was selected to train the BRAU-Net++ neural network for the segmentation task required for automatic VBQ determination. The dataset consists of 447 sagittal T1 and T2 MRI series from 218 studies of 218 patients acquired at four different hospitals. All patients had a history of low back pain. The quality and dimensions of these images vary between series, with voxel sizes ranging from $3.15 \times 0.24 \times 0.24$ mm to $9.63 \times 1.06 \times 1.23$ mm. Segmentation of the visible vertebrae (except the sacrum), intervertebral disks (IVDs) and spinal canal was performed using a semi-automated segmentation method. First, 20 randomly selected high-resolution series were chosen for manual segmentation. For the other image series, the segmentations were performed using an iterative automatic segmentation method. The first segmentations were performed using an automatic method that was trained with data from manually

segmented image series. New segmentations (annotations) were then manually corrected and added to the training dataset. This process was repeated four times. Figure 1 shows a diagram of the annotation process [21].
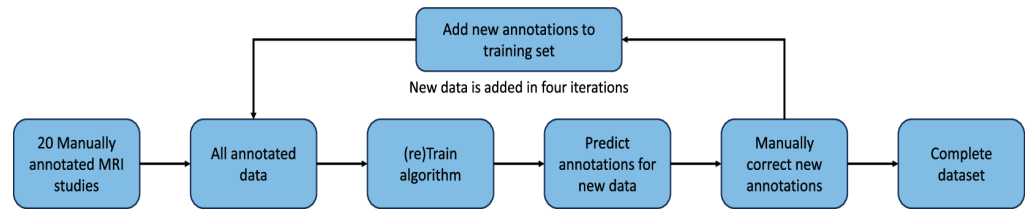

Figure 1. Diagram of the iterative annotation process [21].

Before this dataset was used to train the BRAU-Net++ neural network, several preprocessing steps were carried out. Since the signal intensity of the IVDs was not required for VBQ determination, all segmentations of the IVDs were removed from the segmentation masks in the dataset to improve the training performance of the network and shorten the training time. The segmentation masks were left with 9 classes: the L1–L5 vertebrae (numbered 1 to 5), the Th10–Th12 vertebrae (numbered 6 to 8) and the spinal canal (numbered 9). Figure 2 shows an example of a mid-sagittal slice image from a series and the corresponding segmentation mask.
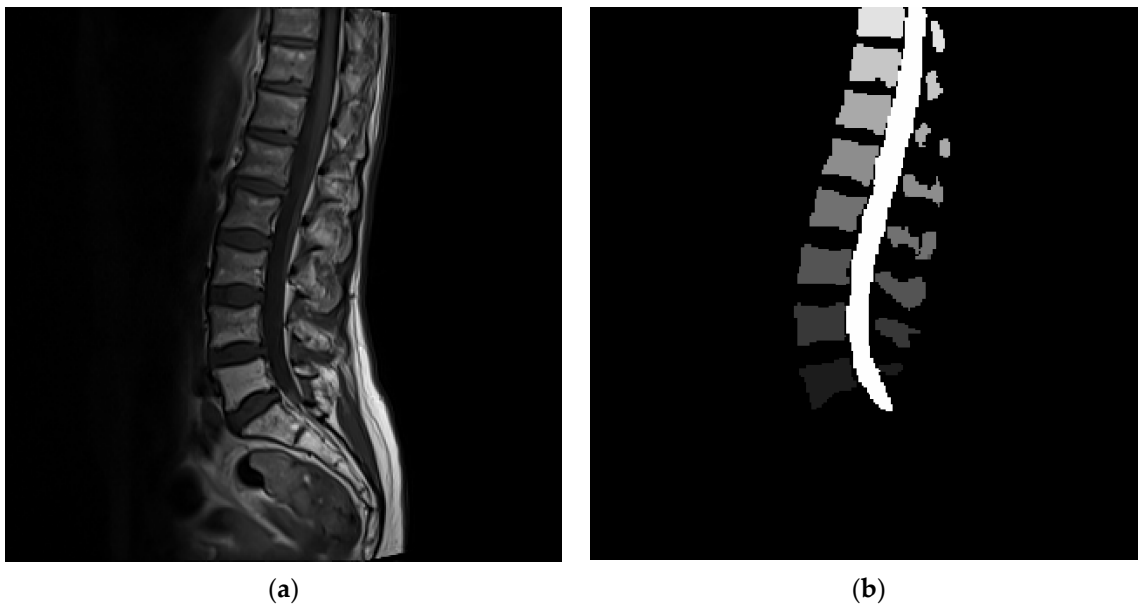

(**a**)   (**b**)

Figure 2. (**a**) Mid-sagittal slice of T1-weighted lumbar spine image series; (**b**) corresponding segmentation mask with pixel values ranging from 0 to 9.

The intensity values of the MR images were then rescaled to values between 0 and 1 in order to obtain uniform value ranges for the different series. All images and segmentation masks were rescaled to a size of 224×224 pixels. From the dataset, 20% of the randomly selected series were used as test data and 10% were randomly selected as validation data. The data used for training were saved in .npz format from the Numpy package for scientific computing in Python. The test and validation datasets were saved in HDF5 format from the h5py Python package. This was carried out to adapt the dataset to the form of the Synapse dataset used in the original BRAU-Net++ training, which facilitates the process of neural network training [20]. All preprocessing steps were coded in the Python (version 3.9.12.) programming language. Although only T1-weighted image series are required for VBQ determination, the neural network was trained with both T2- and T1-weighted data. T2-weighted images of the spine have a hyperintense signal in the spinal canal where the CSF

is located in relation to the neural tissue. In contrast, T1-weighted images of the spine tend to have lower signal intensity differences between the CSF and neural tissue in the spinal canal. For this reason, T2-weighted images may be useful for segmenting the CSF from the spinal canal masks. Another MRI dataset was randomly selected from the database of the University Hospital Rijeka, Croatia. It consists of 166 image series, 83 T1-weighted and 83 T2-weighted lumbar spine MRI data from 83 studies of 83 patients, all acquired using a 1.5T MRI device (Siemens Magnetom Aera, Siemens Healthineers, Erlangen, Germany). The T1- and T2-weighted data are available in pairs for each study (each patient). This dataset was selected to determine whether four different methods of VBQ determination showed statistically significant differences from each other. To be usable for the trained model, this dataset was subjected to the same preprocessing steps as the public dataset used for neural network training.

2.2. Neural Network and Training

The BRAU-Net++ neural network architecture was chosen for the segmentation task required for automatic VBQ determination. This u-shaped architecture is a hybrid CNN–transformer network that uses bi-level routing attention (BRA) as the core building block for its encoder–decoder structure. It also uses skip connection channel–spatial attention (SCCSA) which adopts convolution operations, aiming to minimize local spatial information loss and amplify the global dimension interaction of multi-scale features [20]. The code for the BRAU-Net++ neural network can be downloaded at: https://github.com/Caipengzhou/BRAU-Netplusplus/tree/master (accessed on 10 June 2024).

The loss function used for training the dataset was the same that used by the authors of the original article for Synapse dataset training. It is a hybrid loss function that combines both dice loss and cross-entropy loss. Dice loss (L_{dice}), cross-entropy loss (L_{ce}) and hybrid loss (L) are defined as follows:

$$L_{dice} = 1 - \sum_k^K \frac{2\omega_k \sum_i^N p(k,i)g(k,i)}{\sum_i^N p^2(k,i) + \sum_i^N g^2(k,i)} \; , \tag{1}$$

$$L_{ce} = -\frac{1}{N}\sum_{i=1}^N G(k,i) \cdot log\left(P(k,i)\right) + (1 - G(k,i)) \cdot log\left(1 - p(k,i)\right), \tag{2}$$

$$L = \lambda L_{dice} + (1 - \lambda)L_{ce} \; , \tag{3}$$

where N is the number of pixels; *G(k, i)* ∈ (0, 1) and *P(k, i)* ∈ (0, 1) indicate the ground truth label and the produced probability for class k, respectively; K is the number of classes; $\sum_k \omega_k = 1$ is the weight sum of all classes; and λ is a weighted factor that balances the impact of L_{dice} and L_{ce}. In this study, ω_k and λ are empirically set at 1 K and 0.6, respectively [20]. Figure 3 shows the BRAU-Net++ architecture and SCCSA module.

To evaluate the performance of the neural network, the average dice similarity coefficient (DSC) is used, along with the precision, accuracy, recall and Jaccard index, also known as intersection-over-union (IoU). These evaluation metrics are calculated as follows:

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \; , \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \; , \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \; , \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \; , \tag{7}$$

$$IoU = \frac{TP}{TP + FP + FN} \; , \tag{8}$$

where *TP*, *FP*, *TN* and *FN* are true positive, false positive, true negative and false negative. As can be seen from Equations (4) and (8), the IoU can be expressed as *IoU = DSC/(2-DSC)*. Nevertheless, the *IoU* value was calculated for the sake of completeness. Neural network predictions can formally be separated into these categories using a confusion matrix algorithm implemented in the Python programming language.
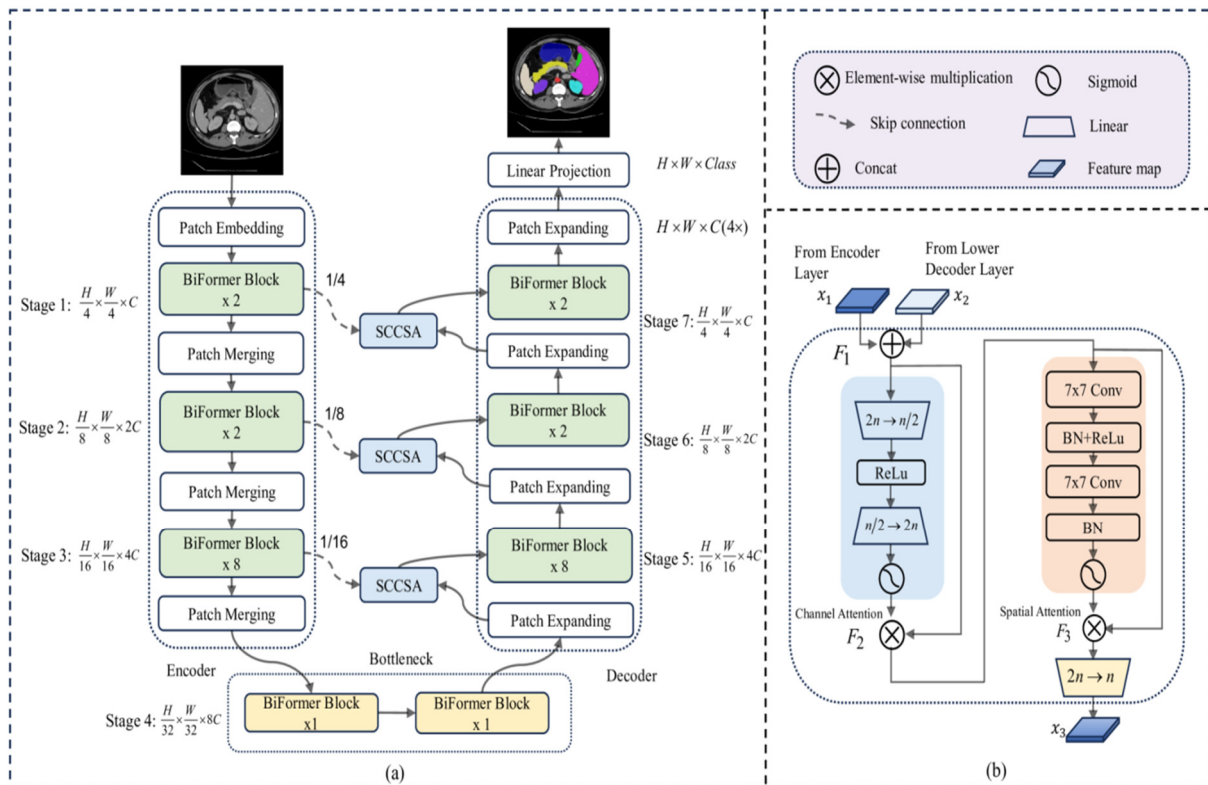

Figure 3. The architecture of the BRAU-Net++, which is constructed based on the BiFormer block. (**a**) U-shaped hybrid CNN-Transformer network architecture that uses sparse attention mechanism: bi-level routing attention as core building idea to hierarchically design the encoder-decoder structure. (**b**): The skip connection channel–spatial attention (SCCSA) module, which enhances the ability of cross-dimension interactions from both channel and spatial aspects and compensates for the spatial information loss caused by down-sampling [20].

These evaluation metrics can be used to compare the performance of BRAU-Net++ and other modern neural network architectures. nnU-Net, a self-configuring deep learning-based framework for medical image segmentation, has gained acceptance in the medical image analysis community as a state-of-the-art approach for 3D image segmentation tasks [21,24]. Since nnU-Net was trained on the dataset used for this work, its performance can be directly compared to the performance of the BRAU-Net++ network. The BRAU-Net++ neural network was trained over 400 epochs with the public dataset described above. The training was performed using an HP Z8 G4 workstation with an NVIDIA Quadro P6000 graphics card. The trained neural network model was then used to segment the dataset of 166 series acquired from the resident hospital.

2.3. VBQ Determination

The *VBQ* score is determined from pixel intensity values of T1-weighted non-contrast MR images using the following expression:

$$VBQ = \frac{SI_{L1-L5}}{SI_{CSF}} \;,$$

(9)

where SI_{L1-L5} is the mean pixel signal intensity (SI) value of the *L1*, *L2*, *L3*, *L4* and *L5* vertebrae, and SI_{CSF} is the mean signal intensity value of *CSF*. When drawing ROIs manually, a circular ROI was placed on the L1–L5 vertebral body, avoiding the lesion and posterior venous plexus. *CSF SI* was measured near the L3 vertebra. If there was a complete obstruction of the posterior CSF region, the adjacent segment CSF was chosen [12].

Four different methods of automatic VBQ determination were investigated. For each method, the CSF region SI had to be extracted from the segmentation provided by the neural network using an additional algorithm, since the trained neural network model generates segmentation of the entire spinal canal and not just the CSF region. For this purpose, a connected threshold algorithm was chosen.

In the first method (T1_VBQ), only T1-weighted data and the corresponding segmentation were used. The SI value of the CSF region was extracted in the following way: First, the point of local minimum was found in the segmentation of the spinal canal, with coordinates restricted to the vertical axis between L1 and L5. This local minimum was used as the starting point for the connected threshold algorithm. The lower limit of the threshold was chosen as the local minimum SI value and the upper limit was calculated using the following expression:

$$Upper\ boundary = SI_{min} + \frac{SI_{max} - SI_{min}}{3} \;,$$

(10)

where SI_{min} and SI_{max} are the minimum and maximum SI of the constrained spinal canal region, respectively. The connectivity value was set to 1. This was carried out for each slice of the segmented series where spinal canal segmentation was present.

The second method (T1_VBQ_eroded) uses the same CSF extraction algorithm, but changes the volume of vertebral segmentation and thus the ROI used for VBQ determination. This was achieved using the erosion algorithm. In erosion, pixels that are non-zero and have a neighboring pixel with a value of 0 are removed from the segmentation masks, essentially removing the outer layer of pixels from each segmented region. This method of changing the vertebral ROI volume was chosen to avoid sharp changes in SI at the edge of the segmented vertebral body, which can negatively affect the VBQ value.

In the third method (T2_VBQ), T2-weighted images and segmentation masks are used together with T1-weighted images. Since the T2-weighted MRI data of the spine have hyperintense SI in the CSF region, the CSF can be easily extracted using a segmented threshold algorithm. Similar to the first method, the local maximum point in the segmented region of the spinal canal in the T2-weighted image was used as the initial value for the connected threshold algorithm. The upper limit was set as the local maximum SI value and the lower limit as the maximum SI/1.5. The connectivity value was set to 1. The coordinates of this segmented region were then used to determine the SI in the CSF from the T1-weighted image of the same study (patient). T1- and T2-weighted images belonging to the same study do not necessarily match spatially, even if they were acquired sequentially as part of the same MRI protocol. This may be due to the movement of the patient during imaging or to the different parameter values of the T1 and T2 MR sequences, such as the number of slices or the spatial resolution. To address this, a co-registration algorithm implemented in the FMIRB Software Library (FSL) [25] was used together with a Python script to resize T1- and T2-weighted images together with their segmentation masks to the same size.

The fourth method (T2_VBQ_eroded) uses the same CSF SI extraction algorithm as the third method, together with erosion of the vertebral segmentation region, as used in the second method.

Each different method of CSF extraction and vertebral ROI determination was applied to each of the 83 study datasets from the resident hospital database. The average signal intensities for vertebrae and CSF were calculated from the original image pixel values without preprocessing steps using segmentation masks as ROI selections. The VBQ score was then determined using Equation (9). For six randomly selected series from the database, the VBQ was determined using manually drawn ROIs in addition to the first and second methods. All methods were compared using one-way analysis of variance (or one-way ANOVA) to determine if there were statistically significant differences between the four methods of VBQ determination.

3. Results

3.1. Trained Network Model Evaluation

The BRAU-Net++ neural network was trained with 447 sagittal T1- and T2-weighted MRI series with segmentation masks for nine classes over 400 epochs. The neural network achieved convergence and produced a trained model that was later used to generate segmentations of the vertebrae and spinal canal. The performance of the network was evaluated using DSC, accuracy, precision, recall and IoU for each of the nine classes for which the prediction masks were created. Table 1 shows the result of the evaluation metrics.

Table 1. The evaluation metrics of the trained model.

Class	DSC	Accuracy	Precision	Recall	IoU
1–L1	0.907	0.999	0.922	0.892	0.829
2–L2	0.917	0.999	0.927	0.907	0.846
3–L3	0.917	0.999	0.928	0.906	0.847
4–L4	0.912	0.999	0.923	0.902	0.838
5–L5	0.913	0.999	0.917	0.908	0.839
6–Th10	0.915	0.999	0.907	0.923	0.843
7–Th11	0.926	0.999	0.927	0.926	0.862
8–Th12	0.904	0.999	0.905	0.904	0.825
9–Spinal canal	0.902	0.998	0.897	0.907	0.821

The performance of the BRAU-Net++ network can be directly compared with that of the nnU-Net neural network, as both networks were trained with the same dataset. The results of the DSC evaluation metric for nnU-Net are 0.92 ± 0.05 for the vertebrae and 0.92 for the spinal canal. The average DSC value for BRAU-Net++ for the vertebrae is 0.914 ± 0.007, which is in the range of the values reported by nnU-Net. The DSC value for the spinal canal is 0.902 and thus below the result of the nnU-Net DSC score. Figure 4 shows examples of original images, ground truth segmentation and predictions for individual images from T1- and T2-weighted series.

Even though the trained model performed well according to the evaluation metrics presented in Table 1, for some of the 166 series data chosen for VBQ determination, a mixing of segmentation labels is present in the prediction provided by the trained model. An example of segmentation label mixing is shown in Figure 5.
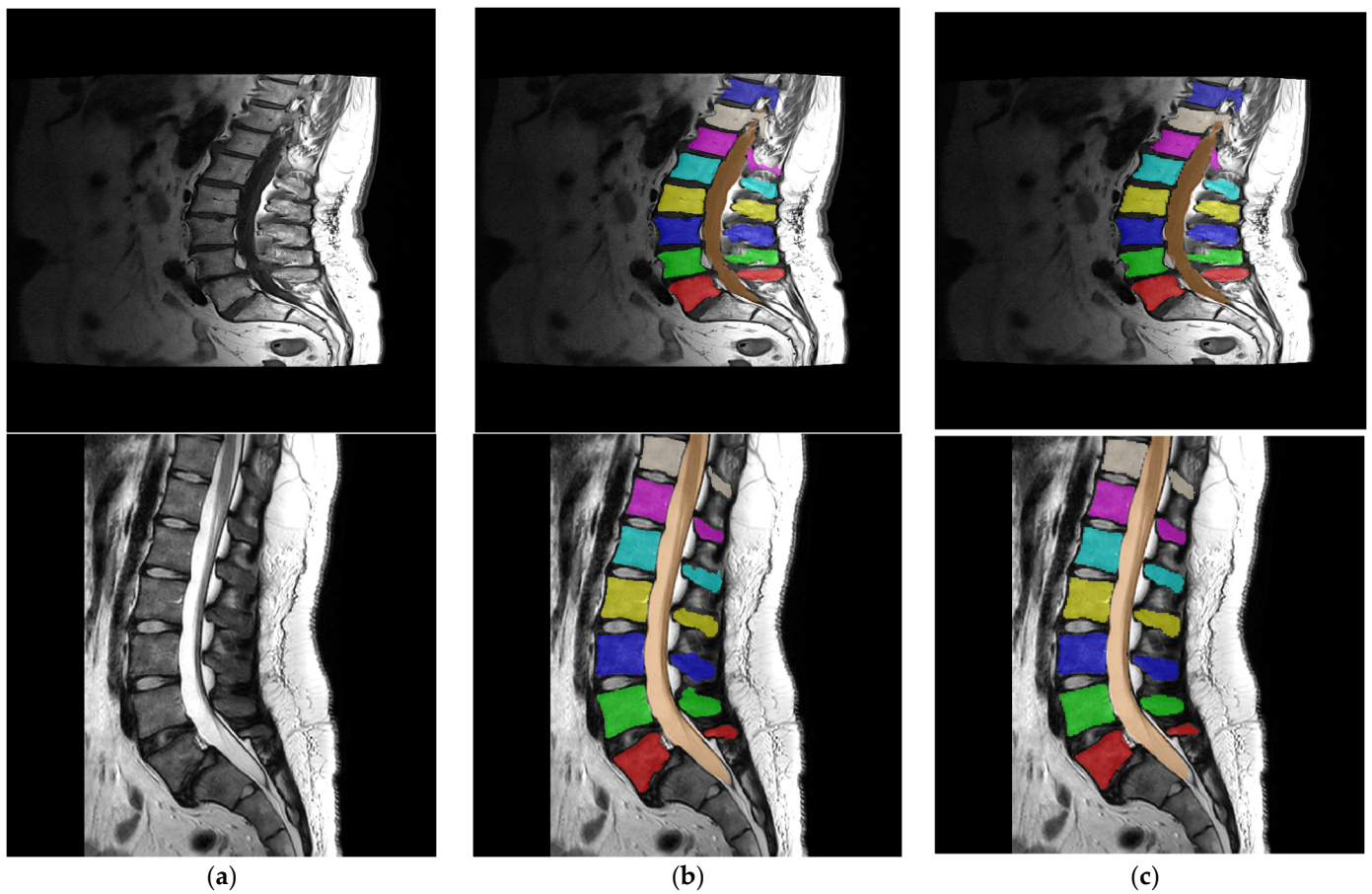
Figure 4. (**a**) Single mid-sagittal T1-weighted (**upper**) and T2-weighted (**lower**) image from a randomly chosen series. (**b**) Ground truth segmentations overlayed on the corresponding image from (**a**). (**c**) Prediction segmentation images of the trained neural network model. Different colors represent each of the 9 segmentation classes, which are labeled 1–9 in the segmentation masks.
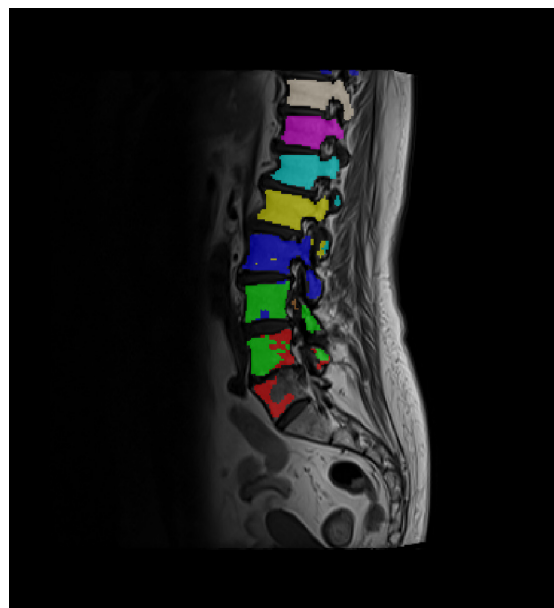

Figure 5. An image slice example showing mixing of labels (green/red and blue/yellow). Different colors represent each of the 9 segmentation classes, which are labeled 1–9 in the segmentation masks.

3.2. VBQ Determination Methods' Results

VBQ was determined for a mid-sagittal slice of six randomly selected series using three different methods: the hand-drawn ROI method, automated segmentation from T1-weighted images, and automated segmentation with an applied erosion algorithm. Figure 6 shows an example of the three VBQ determination methods on the same image.
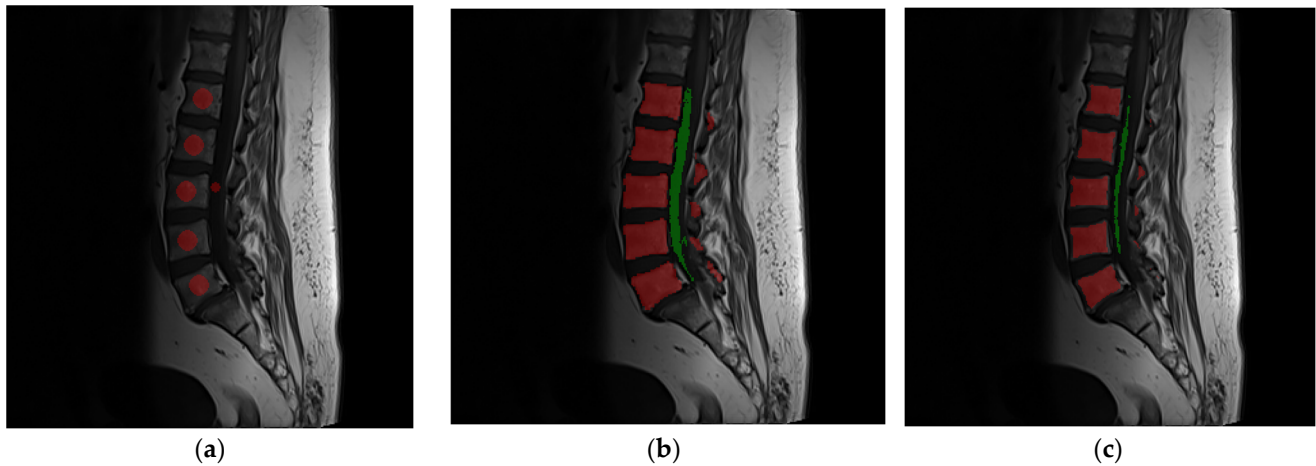

(**a**) (**b**) (**c**)

Figure 6. (**a**) Hand-drawn ROIs for vertebra and CSF, shown as red circles. (**b**) ROIs generated by automatic segmentation from trained neural network model and CSF extraction algorithm (T1_VBQ). (**c**) Same as (**b**), with erosion algorithm applied to vertebrae and CSF ROIs (T1_VBQ_eroded). In (**b**,**c**), red color represents vertebrae segmentation, while green color represents CSF segmentation.

Table 2 shows the result of VBQ score determination using Equation (9).

Table 2. Results of VBQ determination.

Series	Manually Drawn ROI VBQ	T1_VBQ	T1_VBQ_eroded
1	3.37	2.80	3.01
2	2.60	2.69	2.86
3	3.52	3.29	3.46
4	3.58	3.09	3.27
5	4.22	3.59	3.84
6	3.61	2.67	3.02

Four different methods were compared for VBQ determination (T1_VBQ, T1_VBQ_eroded, T2_VBQ and T2_VBQ_eroded). Examples of the segmentation with each method and the segmentation with the trained neural network model are shown in Figure 7 as a 3D rendering. The 3D rendering was performed with the ITK-SNAP software application (version 4.2.0.) [26].

Table 3 shows the results of the four methods of VBQ determination, with average values and standard deviations for VBQ score, vertebra SI, and CSF signal intensity. These values were determined across the 83 studies randomly selected from the resident hospital database, with 166 series of images.
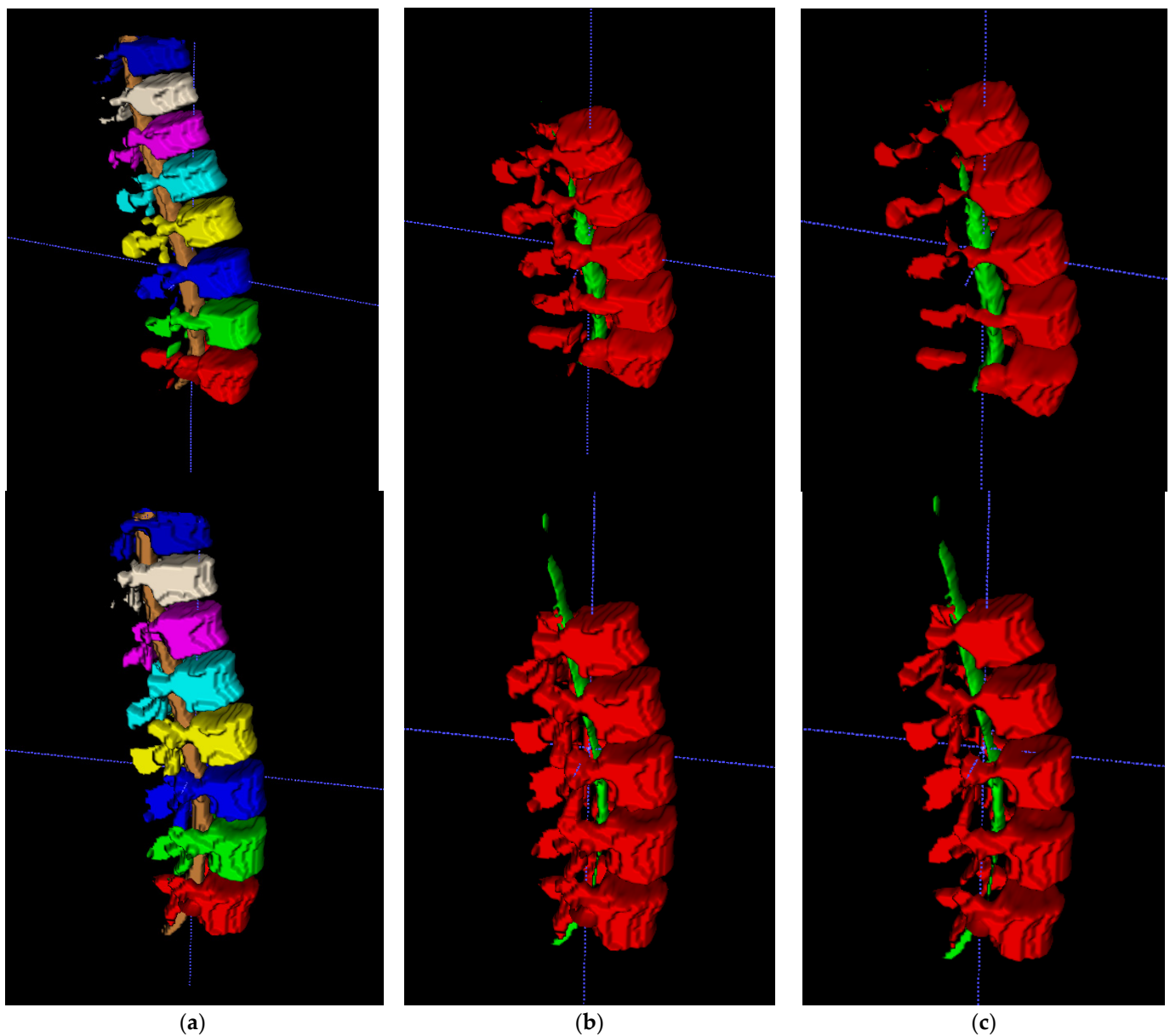
Figure 7. (**a**) Upper: T1-weighted segmentation mask provided by trained neural network model (vertebrae and spinal canal). Lower: T2-weighted segmentation mask provided by trained neural network model. Different colors represent each of the 9 segmentation classes, which are labeled 1 – 9 in the segmentation masks. (**b**) Upper: T1_VBQ method. Lower: T2_VBQ method. (**c**) Upper: T1_VBQ_eroded method. Lower: T2_VBQ_eroded method. In (**b**,**c**) green regions represent segmented CSF, while red colored regions represent segmented vertebrae.

Table 3. Results of VBQ determination.

	T1_VBQ	**T1_VBQ_eroded**	**T2_VBQ**	**T2_VBQ_eroded**
VBQ score	2.60 ± 0.56	2.73 ± 0.58	2.21 ± 0.54	2.33 ± 0.57
Vertebra SI	127.50 ± 31.84	133.94 ± 34.48	124.68 ± 30.52	131.61 ± 33.19
CSF SI	50.39 ± 14.48	50.39 ± 14.48	59.05 ± 22.39	59.05 ± 22.39

Each set of VBQ scores obtained using one of the four different methods was compared using a one-way ANOVA statistic to determine if there were statistically significant differences between the four methods. The resulting *p*-value is $p = 0.00001 < 0.05$, which

means that there is a significant statistical difference between the datasets. A post hoc Tukey's Honest Significant Difference (HSD) test was applied to the same data to assess the significance of differences between pairs of VBQ scores obtained by the different methods. The results show that there is a statistically significant difference between the T1_VBQ and T2_VBQ methods ($p = 0.00006$), the T1_VBQ and T2_VBQ_eroded methods ($p = 0.00907$), the T1_VBQ_eroded and T2_VBQ methods ($p = 0.000001$), and the T1_VBQ_eroded and T2_VBQ_eroded methods ($p = 0.00004$).

4. Discussion

Automatic segmentation of vertebrae and the spinal canal from T1- and T2-weighted MRI data was achieved by training a hybrid CNN–transformer neural network, BRAU-Net++, using a publicly available dataset. The resulting model achieved good accuracy, precision and DSC values. Compared to the DSC scores of the state-of-the-art nnU-Net neural network, BRAU-Net++ achieves comparable results for vertebral segmentation and poorer results for spinal canal segmentation. This is also in line with a review of modern deep learning architectures for lumbar spine imaging, which reports DSC values between 84.9 and 93 for different U-Net-based architectures [27]. The accuracy values approach 99% for each class, which can be explained by the relative number of TP and TN pixels in the segmentation images [27]. As is common in medical image segmentation, the segmented area comprises only a small part of the image, while the background contains all other pixels. For this reason, the number of TP pixels is much smaller than the number of TN pixels. Consequently, the accuracy of Equation (5) yields large values for each segmentation class [27].

The spinal canal class has the lowest DSC and accuracy values, which can be explained by the different pixel intensities of the CSF in T1-weighted and T2-weighted images. In T1-weighted images, the CSF is dark and has low intensity values. Conversely, the CSF in T2-weighted images is bright, i.e., it contains high pixel intensity values. Since the network was trained simultaneously with T1- and T2-weighted image data, the difference in pixel values in only one part of the segmented region (spinal canal) could influence the training and the resulting neural network model.

Although the evaluation metrics obtain good results, in some of the 166 data series selected for VBQ determination, label mixing is present in the resulting segmentation. The cause of this is not apparent, but should be investigated further. Fortunately, the mixing of labels does not affect the VBQ determination, as this only requires a mean vertebral ROI SI across all lumbar vertebrae.

Automatic VBQ determination was successfully achieved by using the segmentations from the trained neural network model with the addition of analytical CSF segmentation using a connected threshold algorithm. Four different VBQ determination methods were presented. Interestingly, the mean and standard deviation of the VBQ score for each of the four methods appear to be relatively close to each other. However, the result of the one-way ANOVA shows that there is a significant statistical difference between the datasets, which means that the choice of VBQ determination method is not straightforward. The results of the post hoc Tukey's HSD test show statistically significant differences between the methods using T1-weighted images and those using T2-weighted images for segmentation. The average VBQ scores of all four methods are in the range reported in [12]. The results of the manually drawn ROI method are also in the range cited by [12], except for the VBQ score for series 5, as shown in Table 2. The automatic VBQ determination method obtained plausible results for the same series, which, although only a single data point is involved, illustrates how useful a consistent automatic method can be. The main difficulty in the practical implementation of this method is the co-registration of T1- and T2-weighted images, which is required for two of the four proposed methods as it uses an additional program, FSL. The challenges of neural network training, data curation and preparation do not affect the practical implementation, as once a neural network model is trained, it can be easily shared and used for the segmentation task.

5. Conclusions

Four different methods of automatic VBQ determination using a trained neural network model for semantic segmentation were created and tested on a dataset of 83 MRI studies of the lumbar spine. A one-way ANOVA statistic with a post hoc Tukey's HSD test showed that there is a significant statistical difference between the results of the different VBQ determination methods. The main limitation of this study is the lack of BMD data for the 83 patients whose lumbar spine MRI data were used to test and compare the different methods. BMD data would allow a direct comparison of the VBQ score with the gold standard used in clinical settings. Nevertheless, this work is novel in its way of automating VBQ score determination, which could prove useful for further testing and validation of the VBQ score as a diagnostic tool for osteoporosis. For future work, the proposed automated VBQ determination methods could be used with MRI data from patients for whom BMD values are also available to evaluate the correlation coefficient and determine the best of the four methods. While the original hand-drawn VBQ score determination method was designed to be simple and easy to use, automating the segmentation process opens up new possibilities for assessing bone quality. For example, hyperintense signal intensity regions of vertebrae in T1-weighted images, often referred to as "bone islands", can be isolated from the final VBQ determination, potentially improving the consistency of the results. One could also use other methods to determine the properties of vertebral bone tissue, such as the heterogeneity equation in combination with proton density, perhaps creating a new radiomic parameter.

Author Contributions: Conceptualization, S.J. and K.S.; methodology, S.J., K.S. and D.M.R.; software, K.S.; validation, S.J. and D.M.R.; formal analysis, S.J.; investigation, K.S. and D.M.R.; resources, S.J.; data curation, K.S.; writing—original draft preparation, K.S.; writing—review and editing, S.J., K.S. and D.M.R.; visualization, K.S.; supervision, S.J.; project administration, S.J. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The publicly available sagittal spinal MRI dataset with ground truth segmentation masks can be downloaded at: https://zenodo.org/records/10159290 (accessed on 5 November 2024). The MRI dataset consisting of the 166 series of images from 83 anonymized studies which were randomly chosen from the resident hospital database is available upon request from the main author. All Python preprocessing, post-processing, and VBQ determination codes are available upon request from the main author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

VBQ	Vertebral bone quality
MRI	Magnetic resonance imaging
CNN	Convolutional neural network
ROI	Region of interest
CSF	Cerebrospinal fluid
DSC	Dice similarity coefficient
IoU	Intersection-over-union
ANOVA	Analysis of variance
BMD	Bone mineral density
DEXA	Dual-energy X-ray absorptiometry

SI	Signal intensity
IVD	Intervertebral disk
SCSSA	Skip connection channel–spatial attention
TP	True positive
FP	False positive
TN	True negative
FN	False negative

References

1. NIH Consensus Development Panel on Osteoporosis Prevention, Diagnosis, and Therapy. Osteoporosis prevention, diagnosis, and therapy. *JAMA* **2001**, *285*, 785–795. [CrossRef] [PubMed]
2. Seeman, E.; Delmas, P.D. Bone quality–the material and structural basis of bone strength and fragility. *N. Engl. J. Med.* **2006**, *354*, 2250–2261. [CrossRef] [PubMed]
3. Dimar, J.; Bisson, E.F.; Dhall, S.; Harrop, J.S.; Hoh, D.J.; Mohamed, B.; Wang, M.C.; Mummaneni, P.V. Congress of Neurological Surgeons Systematic Review and Evidence-Based Guidelines for Perioperative Spine: Preoperative Osteoporosis Assessment. *Neurosurgery* **2021**, *89* (Suppl. S1), S19–S25. [CrossRef] [PubMed]
4. Lin, W.; He, C.; Xie, F.; Chen, T.; Zheng, G.; Yin, H.; Chen, H.; Wang, Z. Assessment of bone density using the 1.5 T or 3.0 T MRI-based vertebral bone quality score in older patients undergoing spine surgery: Does field strength matter? *Spine J.* **2023**, *23*, 1172–1181. [CrossRef] [PubMed]
5. Celi, M.; Rao, C.; Scialdoni, A.; Tempesta, V.; Gasbarra, E.; Pistillo, P.; Tarantino, U. Bone mineral density evaluation in osteoporosis: Why yes and why not? *Aging Clin. Exp. Res.* **2013**, *25* (Suppl. S1), S47–S49. [CrossRef]
6. Ehresman, J.; Pennington, Z.; Schilling, A.; Lubelski, D.; Ahmed, A.K.; Cottrill, E.; Khan, M.; Sciubba, D.M. Novel MRI-based score for assessment of bone density in operative spine patients. *Spine J.* **2020**, *20*, 556–562. [CrossRef]
7. Dipaola, C.P.; Bible, J.E.; Biswas, D.; Dipaola, M.; Grauer, J.N.; Rechtine, G.R. Survey of spine surgeons on attitudes regarding osteoporosis and osteomalacia screening and treatment for fractures, fusion surgery, and pseudoarthrosis. *Spine J.* **2009**, *9*, 537–544. [CrossRef]
8. Cosman, F.; de Beur, S.J.; LeBoff, M.S.; Lewiecki, E.M.; Tanner, B.; Randall, S.; Lindsay, R.; National Osteoporosis Foundation. Clinician's guide to prevention and treatment of osteoporosis. *Osteoporos. Int.* **2014**, *25*, 2359–2381. [CrossRef]
9. Girardi, F.P.; Parvataneni, H.K.; Sandhu, H.S.; Cammisa, F.P.; Grewal, H.; Schneider, R.; Lane, J.M. Correlation between vertebral body rotation and two-dimensional vertebral bone density measurement. *Osteoporos. Int.* **2001**, *12*, 738–740. [CrossRef]
10. Rumancik, S.; Routh, R.H.; Pathak, R.D.; Burshell, A.L.; Nauman, E.A. Assessment of bone quantity and distribution in adult lumbar scoliosis: New dual-energy x-ray absorptiometry methodology and analysis. *Spine* **2005**, *30*, 434–439. [CrossRef]
11. Xie, Y.; Zhang, L.; Xiong, Q.; Gao, Y.; Ge, W.; Tang, P. Bench-to-bedside strategies for osteoporotic fracture: From osteoimmunology to mechanosensation. *Bone Res.* **2019**, *7*, 25. [CrossRef] [PubMed]
12. Chen, Z.; Lei, F.; Ye, F.; Yuan, H.; Li, S.; Feng, D. MRI-based vertebral bone quality score for the assessment of osteoporosis in patients undergoing surgery for lumbar degenerative diseases. *J. Orthop. Surg. Res.* **2023**, *18*, 257. [CrossRef] [PubMed]
13. Biswas, A.; Banik, R. Advancements in medical image analysis: A comprehensive method of AI-based classification and segmentation technique. *Artif. Intell. Appl.* **2024**. [CrossRef]
14. Pu, Q.; Xi, Z.; Yin, S.; Zhao, Z.; Zhao, L. Advantages of transformer and its application for medical image segmentation: A survey. *BioMed. Eng. OnLine* **2024**, *23*, 14. [CrossRef]
15. Zhao, S.; Wu, X.; Chen, B.; Li, S. Automatic vertebrae recognition from arbitrary spine MRI images by a category-Consistent self-calibration detection framework. *Med. Image Anal.* **2021**, *67*, 101826. [CrossRef]
16. Zhang, R. A state-of-the-art survey of deep learning for lumbar spine image analysis: X-ray, CT, and MRI. *AIM* **2024**, *1*, 3. [CrossRef]
17. Cui, Y.; Zhu, J.; Duan, Z.; Liao, Z.; Wang, S.; Liu, W. Artificial intelligence in spinal imaging: Current status and future directions. *Int. J. Environ. Res. Public Health* **2022**, *19*, 11708. [CrossRef]
18. Lin, J.; Zhang, H.; Shang, H. Convolutional neural network incorporating multiple attention mechanisms for MRI classification of lumbar spinal stenosis. *Bioengineering* **2024**, *11*, 1021. [CrossRef]
19. Tumko, V.; Kim, J.; Uspenskaia, N.; Honig, S.; Abel, F.; Lebl, D.R.; Hotalen, I.; Kolisnyk, S.; Kochnev, M.; Rusakov, A.; et al. A neural network model for detection and classification of lumbar spinal stenosis on MRI. *Eur. Spine J.* **2024**, *33*, 941–948. [CrossRef]
20. Lan, L.; Cai, P.; Jiang, L.; Liu, X.; Li, Y.; Zhang, Y. BRAU-Net++: U-Shaped Hybrid CNN-Transformer Network for Medical Image Segmentation. *arXiv* **2024**, arXiv:2401.00722.
21. van der Graaf, J.W.; van Hooff, M.L.; Buckens, C.F.M.; Rutten, M.; van Susante, J.L.C.; Kroeze, R.J.; de Kleuver, M.; van Ginneken, B.; Lessmann, N. Lumbar spine segmentation in MR images: A dataset and a public benchmark. *Sci. Data* **2024**, *11*, 264. [CrossRef] [PubMed]
22. Hu, F.; Li, X.; Zhao, D.; Chen, C.; Liu, G.; Yang, Q. The diagnostic value of MRI-based vertebral bone quality score for osteoporosis or osteopenia in patients undergoing lumbar surgery: A meta-analysis. *Osteoporos. Int.* **2024**, *35*, 1881–1895. [CrossRef] [PubMed]

23. Najafi, A.; Baradaran Bagheri, A.B.; Hadavi, D.; Mobedi, A.; Azarsina, S.; Chaghamirzayi, P. Vertebral bone quality score as a new tool for osteoporosis diagnosis in patients undergoing lumbosacral fusion surgery: A single center cohort study. *Eur. J. Transl. Myol.* **2024**. [CrossRef] [PubMed]

24. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **2021**, *18*, 203–211. [CrossRef]

25. Smith, S.M.; Jenkinson, M.; Woolrich, M.W.; Beckmann, C.F.; Behrens, T.E.J.; Johansen-Berg, H.; Bannister, P.R.; De Luca, M.; Drobnjak, I.; Flitney, D.E.; et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **2004**, *23* (Suppl. S1), S208–S219. [CrossRef]

26. Yushkevich, P.A.; Piven, J.; Hazlett, H.C.; Smith, R.G.; Ho, S.; Gee, J.C.; Gerig, G. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* **2006**, *31*, 1116–1128. [CrossRef]

27. Rainio, O.; Teuho, J.; Klén, R. Evaluation metrics and statistical tests for machine learning. *Sci. Rep.* **2024**, *14*, 6086. [CrossRef]