# Development of Technology for Analysing X-Ray Images of Bone Tissue for Computer Diagnosis of Osteoporosis

**Daria V. Nekrasova[1*], Nataly Yu. Ilyasova[1,2], Nikita S. Demin[1,2], and Sergey S. Pervushkin[3]**
[1] Samara National Research University, 34 Moskovskoye Shosse, Samara 443086, Russian Federation
[2] IPSI, NRC "Kurchatov Institute", 151 Molodogvardeyskaya str., Samara 443001, Russian Federation
[3] FSBEI HE "SamSMU" of the Ministry of Health of the Russian Federation, 89 Chapaevskaya str., Samara 443099, Russian Federation
*e-mail: dashanek007@gmail.com

**Abstract.** The interpretation of lumbar spine radiographs, the most accessible diagnostic modality for osteoporosis, is often subjective and hampered by limited accuracy. To address this, we developed a neural network-based algorithm for computer-aided diagnosis. This study provides a comparative analysis of key neural network architectures for image segmentation and classification. The proposed system, with a classification accuracy of 87.42%, offers a valuable adjunct for radiologists, mitigating diagnostic risks and enhancing the overall efficacy of osteoporosis screening.

**Keywords:** osteoporosis; radiography; machine learning; resnet; deeplab.

Paper #9383 received 20 Oct 2025; revised manuscript received 19 Jan 2026; accepted for publication 20 Jan 2026; published online 7 May 2026. doi: 10.18287/JBPE26.12.020303.

## 1 Introduction

In contemporary clinical practice, the diagnosis of osteoporosis remains a significant challenge, given the disease's widespread prevalence, particularly among the elderly [1]. The gold standard for detecting changes in bone mineral density is dual-energy X-ray absorptiometry (DXA), which enables a quantitative assessment of bone health. However, this technique requires specialized and costly equipment, limiting its availability in many healthcare settings, especially within primary care.

Consequently, radiography persists as the most widely available and accessible imaging modality for visualizing bone structures. Although it does not provide a direct quantitative measure of mineral density, radiographs contain numerous visual indicators that allow an experienced clinician to infer the presence of osteoporosis. These indicators include alterations in the shape and architecture of vertebral bodies, as well as a disruption of the trabecular pattern [2].

Radiographic analysis can reveal signs of vertebral body deformity, such as wedging or biconcavity, suggestive of compression fractures. According to study [3], 77% of osteoporotic patients exhibit signs of vertebral remodeling resulting from multiple, often subclinical fractures. This evidence establishes radiographic evaluation as not only feasible but also a clinically valuable tool in diagnosing the disease.

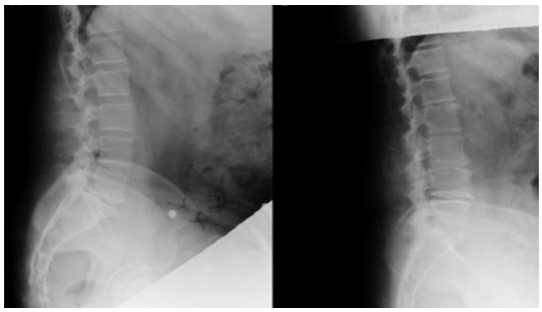
Recent years have witnessed substantial research efforts directed toward the automation of spinal morphometric analysis. For instance, a study [4] introduced a method for assessing the degree of wedge deformity utilizing a convolutional neural network (CNN). The developed model achieved a high degree of accuracy, with an average measurement error of 0.9 mm. However, it is crucial to recognize that vertebral body deformation represents a consequence of osteoporosis rather than its underlying etiology. Furthermore, the visual detection of alterations in vertebral shape is a comparatively simpler task than the quantitative assessment of bone density.

The classic radiological indicator of osteoporosis is a 20–30% reduction in bone mass [5]. Nevertheless, visual identification of such a loss on a radiograph is exceptionally challenging. This limitation is particularly significant in the early stages of the disease, when structural degradation of the trabeculae is not yet pronounced.

A contemporary approach to this problem involves the automated analysis of radiographs to evaluate bone density and identify disruptions in the trabecular architecture. This methodology has the potential to significantly enhance diagnostic accuracy and reproducibility while mitigating the impact of subjective interpretation.

Figure 1 presents radiographic comparisons of healthy and osteoporotic bone tissue, illustrating the distinct visual differences in their respective trabecular structures.

This paper was presented at the XI International Conference "Information Technology and Nanotechnology" (ITNT-2025), Samarkand, Uzbekistan, October 7 – 9, 2025.



(a) (b)

Fig. 1 Lumbar radiographs (a) without disease, (b) with disease.

The study [6] presents a diagnostic technology for osteoporosis based on the analysis of femoral neck radiographs, a site particularly susceptible to osteoporotic fractures. The authors addressed the extraction and statistical processing of textural features. A classification model constructed upon the selected features achieved an accuracy of 84%, thereby substantiating the potential of textural analysis for diagnostic applications. Nevertheless, the proposed methodology exhibits inherent limitations: its efficacy demonstrates a critical dependence on image brightness and requires subjective selection of the region of interest, which constrains its broader applicability.

The present study aims to develop a comprehensive system, grounded in machine learning methods, capable of automatically segmenting radiographic images, selecting regions of interest, and performing binary classification (osteoporosis/normal), thereby minimizing human factor intervention.

## 2 Materials and Methods

### *2.1 Data Description and Preprocessing*

The study utilized radiographic images of the lumbar spine as the primary data source. This anatomical region was selected for its optimal clarity and high diagnostic value in the context of osteoporosis. In contrast to the thoracic spine, where the superimposition of rib structures frequently obscures the vertebral bodies, the lumbar region affords a less obstructed view. The anatomical superimposition present in thoracic radiographs introduces significant visual noise, which complicates automated analysis and accurate computational estimation of bone density. Consequently, a focus on the lumbar spine enables more reliable feature extraction and model training.

A dataset of 252 anonymized lumbar spine radiographs, acquired with a Siemens X-ray apparatus, was procured from Samara State Medical University. The images were stored in the Digital Imaging and Communications in Medicine (DICOM) format, ensuring high fidelity and the preservation of essential metadata required for subsequent analysis.

Prior to the experiment, all data were manually annotated by a qualified radiology technician and categorized into two cohorts: 149 images from patients with a confirmed diagnosis of osteoporosis, and 103 images from patients exhibiting no evidence of the pathology.

During the preparation of the training set, each vertebra was systematically categorized into one of two classes: vertebrae obscured by superimposition of adjacent bony structures (e.g., the iliac bone or sacrum), and fully visible vertebrae devoid of such anatomical overlap. This classification served to enhance the quality of the training data and minimize potential errors during model training.

Figure 2 provides an example of an annotated lumbar spine image with delineated regions of interest.

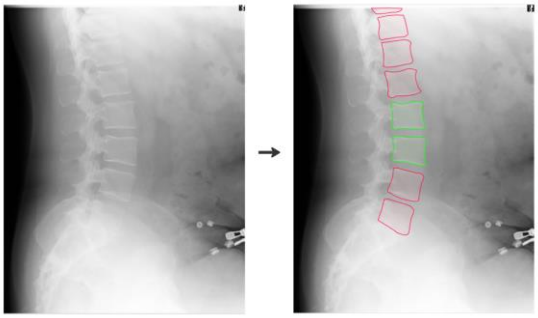


Fig. 2 Labeled image for model training.

For the construction and evaluation of machine learning models, the original dataset was partitioned into a training set and a validation set. The training set, comprising 157 images, was utilized to train the neural network in the recognition of pathological features. The validation set, consisting of 40 images, was designated for assessing the model's generalizability and mitigating overfitting.

Prior to being input into the segmentation model, the radiographic images underwent a preprocessing pipeline consisting of resizing and normalization. The original DICOM images, which exhibit heterogeneous dimensions, were standardized to a uniform size of 1024×1024 pixels. This standardization was achieved via interpolation to preserve maximal informational integrity.

Subsequently, the images were normalized to a consistent numerical scale. To compensate for variations in illumination and contrast, a second normalization step was applied using predetermined mean and standard deviation values derived from the ImageNet dataset. This alignment is particularly crucial due to the limited volume of annotated medical images available for this study, as it enables effective utilization of transfer learning. All models were initialized with weights pre-trained on ImageNet, a standard practice that provides robust feature extractors and significantly enhances training stability on small datasets. This approach mitigates the risk of overfitting and improves generalization capability, thereby compensating for the constrained dataset size while maintaining model reliability.

## 2.2 Review of Segmentation Model Architectures

This section compares three segmentation architectures employed in this study: U-Net, U-Net++, and DeepLabV3+. These models were selected based on their proven efficacy in medical image segmentation, particularly for X-ray analysis [7–9].

U-Net establishes a foundational encoder-decoder framework for biomedical image segmentation. The encoder extracts hierarchical features through convolutional and downsampling operations, while the decoder reconstructs spatial dimensions to generate segmentation masks [10]. Its distinctive skip connections concatenate high-resolution encoder features with upsampled decoder outputs, enabling precise feature localization and enhanced accuracy for fine anatomical structures. U-Net++ constitutes an architectural refinement of the standard U-Net, engineered to improve performance through a more complex and densely connected topology. It incorporates a series of nested, dense skip pathways between the encoder and decoder, creating an ensemble of U-Net-like architectures of varying depths. This design enhances feature propagation, mitigates the vanishing gradient problem, and enables more effective learning on complex datasets. The result is a model capable of producing more precise and detailed segmentation masks.

DeepLabV3+ constitutes a sophisticated segmentation architecture founded on deep convolutional neural networks with an advanced encoder-decoder structure [11]. A principal feature of this model is the implementation of atrous (dilated) convolutions, which enables the expansion of the receptive field without a proportional increase in computational parameters. This facilitates the integration of multi-scale contextual information, thereby enhancing segmentation precision while maintaining computational efficiency. Furthermore, this approach confers robustness to variations in object scale within the input imagery.

Model performance was quantified using the Dice coefficient, a similarity metric for segmentation tasks. The coefficient is defined by the following Eq.:

$$DC = \frac{2|A \cap B|}{|A|+|B|}, \quad (1)$$

where $A$ is the segmented region of the image, and $B$ is the actual area.

To optimize the model, the Focal Loss function was employed as the primary loss metric. Although originally formulated for object detection, this function is effectively adapted to segmentation by addressing class imbalance. Focal Loss modifies the standard cross-entropy loss such that the model's attention is redirected towards hard-to-classify pixels, while diminishing the contribution of easily classified examples during training. This mechanism enhances the model's capacity for accurate identification of underrepresented classes. The loss function is defined as follows:

$$FL(p_t) = -(1-p_t)^{\gamma} log(p_t), \quad (2)$$

where $p_t$ is the probability of belonging to the correct class, γ is a parameter that allows focus tuning [8].

## 2.3 Review of Classification Model Architectures

To address the classification objective, an evaluation of five prominent convolutional neural network architectures was conducted: ResNet-18, ResNet-34, ResNet-50, EfficientNet-B0, and EfficientNet-B1. This selection was predicated on their well-documented efficacy in image analysis tasks, prevalence in the scientific literature, and systematically varying architectural depth, thereby enabling an assessment of the relationship between model complexity and classification performance [12, 13].

The ResNet architecture family (Residual Networks) is distinguished by its use of residual connections. These skip connections mitigate the problem of gradient degradation, thereby facilitating the stable training of substantially deeper networks. Among the selected variants, ResNet-18 and ResNet-34 represent computationally efficient models with 18 and 34 layers, respectively. In contrast, ResNet-50 employs a more complex structure utilizing bottleneck blocks, which increases computational cost while offering a potential gain in representational capacity and accuracy [14].

The EfficientNet models (B0 and B1) represent a subsequent architectural paradigm based on a compound scaling method that uniformly optimizes network depth, width, and input resolution. This principled approach to scaling yields architectures that achieve an optimal trade-off between accuracy and computational efficiency, rendering them particularly suitable for scenarios demanding a balance between performance and resource constraints [15].

Model performance was quantified using a comprehensive set of evaluation metrics to facilitate a multi-faceted analysis. These included Accuracy, Recall (Sensitivity), Precision, the $F_1$-score (the harmonic mean of precision and recall), and the Cross-Entropy Loss function. The collective use of these metrics provides a robust assessment of classification quality from complementary perspectives. The metrics were calculated as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (3)$$

$$Precision = \frac{TP}{TP+FP}, \quad (4)$$

$$Recall = \frac{TP}{TP+FP}, \quad (5)$$

$$F_1 = 2\frac{Precision \times Recall}{Precision+Recall}, \quad (6)$$

$$CE(p,y) = -\sum_{i=1}^{C} y_i \cdot log(p_i), \quad (7)$$

where *TP* are true positives, *TN* are true negatives, *FP* are false positives, *FN* are false negatives, *C* is the number of classes, $y_i$ is the true label (one-hot), $p_i$ – predicted probability of class membership.

## 3 Results and Discussion

### *3.1 Quantitative Evaluation of Segmentation Models*

This section provides a detailed analysis of the efficacy of three modern architectures for semantic segmentation: the foundational U-Net, its enhanced variant U-Net++, and the advanced DeepLabV3+ model. The models were evaluated on a single dataset under identical training conditions to ensure a valid comparative analysis. The evaluation focused on two principal criteria: segmentation accuracy, quantified by the Dice coefficient, and training stability, assessed through the dynamics of the loss function.

The experimental results, summarized in Table 1, reveal notable differences between the architectures. DeepLabV3+ achieved the highest Dice coefficient of 0.8439, exceeding the baseline U-Net by 2% and U-Net++ by 3.5%. While this absolute difference may appear modest, the improvement is significant in the context of medical segmentation, where precise boundary delineation is critical. It should be noted that the higher Focal Loss value observed for DeepLabV3+ (0.0466 versus 0.0324 for U-Net) is attributable to its greater architectural complexity. This does not, however, preclude its application, as the primary objective remains the maximization of the Dice coefficient.

Table 1 Segmentation model training results.

| Metric | Architecture | | |
|---|---|---|---|
| | **U-Net** | **U-Net++** | **DeepLabV3+** |
| Dice | 0.8272 | 0.8089 | 0.8439 |
| Focal loss | 0.0324 | 0.0423 | 0.0466 |

The training dynamics, illustrated in Figs. 3–5, further distinguish the models. U-Net demonstrates stable growth of the Dice metric for the first 15 epochs, followed by a plateau, indicating a favorable balance between learning rate and final accuracy. U-Net++ is characterized by rapid convergence, attaining its peak accuracy by epoch 7; however, a subsequent slight regression in metrics suggests an onset of overfitting. In contrast, DeepLabV3+ requires a substantially longer training period (62 epochs) but exhibits a consistent, monotonic improvement in performance without indications of overfitting, a characteristic of complex architectures with extensive parameter spaces.

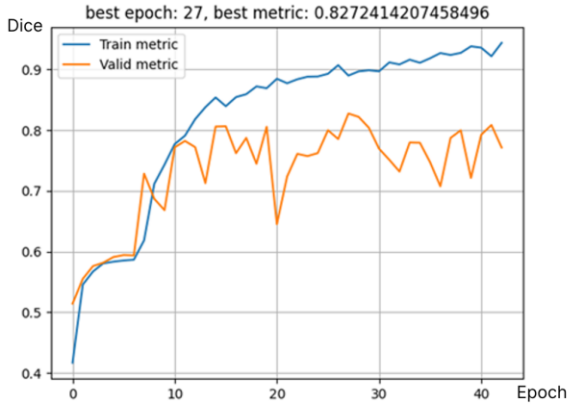



Fig. 3 Training and validation metrics across epochs for U-Net model.

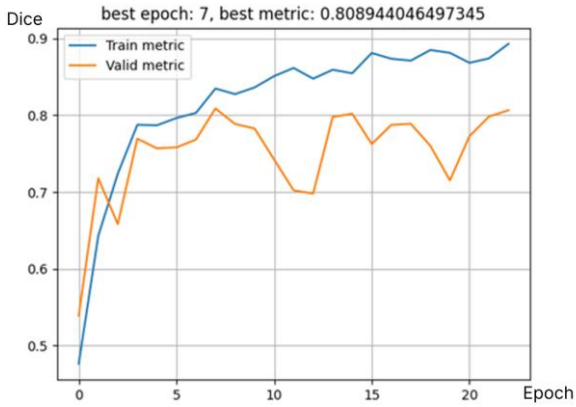



Fig. 4 Training and validation metrics across epochs for U-Net++ model.

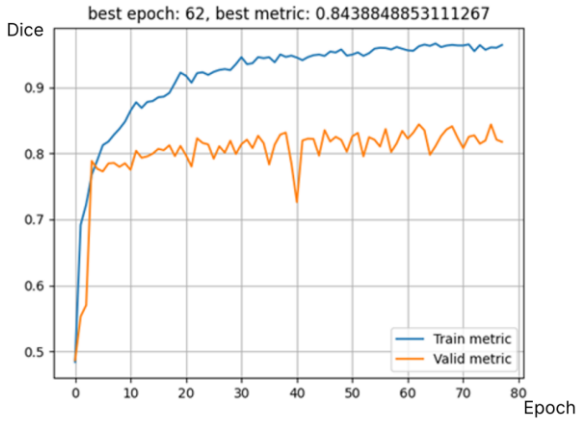



Fig. 5 Training and validation metrics across epochs for DeepLabV3+ model.

The outputs of each model are juxtaposed in Fig. 6. The DeepLabV3+ exemplar demonstrates superior performance in delineating object boundaries, particularly in complex regions characterized by heterogeneous texture. This architecture exhibits

enhanced capability in segmenting fine anatomical details and generates fewer spurious artifacts compared to its counterparts. A critical advantage is its consistent maintenance of topological integrity in challenging anatomical domains, where U-Net and U-Net++ occasionally produce topological errors.

The selection of DeepLabV3+ as the final model is justified by a confluence of factors. Its statistically significant superiority on the primary accuracy metric is coupled with robust training dynamics and a demonstrated capacity for resolving complex segmentation challenges. Although the model demands greater computational resources for training, this expenditure is offset by the substantial improvement in final output quality. Furthermore, the architecture presents significant potential for future performance gains through hyperparameter optimization, the use of pre-trained encoders, and the application of advanced data augmentation techniques. Within the context of medical image analysis, where segmentation precision is paramount, these collective advantages establish DeepLabV3+ as the optimal choice.

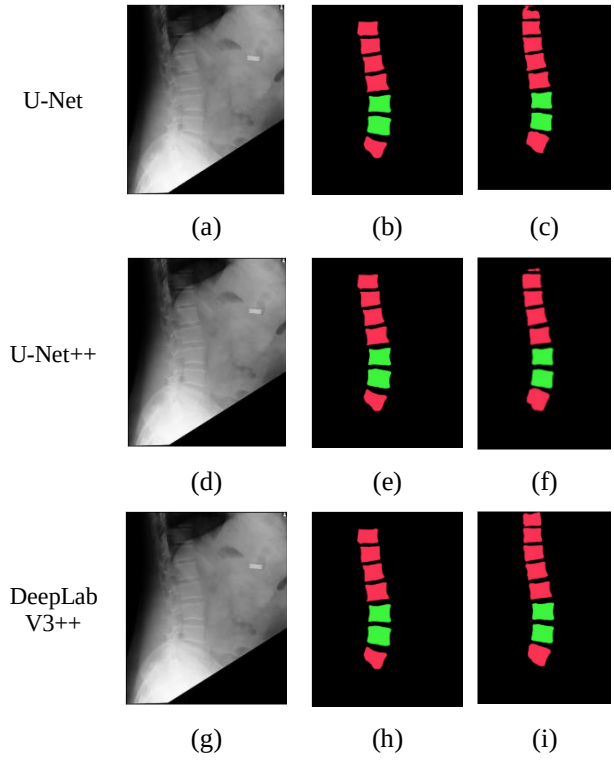



Fig. 6 Experimental outputs of the segmentation models. (a, d, g) Images; (b, e, h) Ground Truth Masks; (c, f, i) Predicted masks.

## *3.2 Quantitative Evaluation of Classification Models*

This study conducts a comparative analysis of deep learning architectures for classifying vertebral pathology. The proposed methodology processes individual vertebrae as separate images with subsequent aggregation of results, preserving diagnostically critical features while maintaining clinical applicability. The training dataset comprised exclusively non-overlapping vertebrae, retaining minimal background tissue to provide essential anatomical context without introducing substantial noise.

Training dynamics, illustrated in Figs. 7–11, reveal substantial performance disparities among architectures. ResNet-34 emerged as the most promising model, attaining the highest quantitative evaluation score (0.8857 at epoch 65) while demonstrating consistent improvement throughout training. Conversely, the deeper ResNet-50 architecture, despite achieving a comparable performance peak (0.8786 at epoch 46), exhibited substantial overfitting, as indicated by significantly elevated cross-entropy values.

The EfficientNet architectures produced divergent outcomes. While EfficientNet-B0 achieved competitive recall rates, its precision marginally trailed ResNet-34, potentially increasing false positive instances in clinical deployment. EfficientNet-B1 demonstrated the least effectiveness among all evaluated models, suggesting limited suitability for this specific diagnostic application.

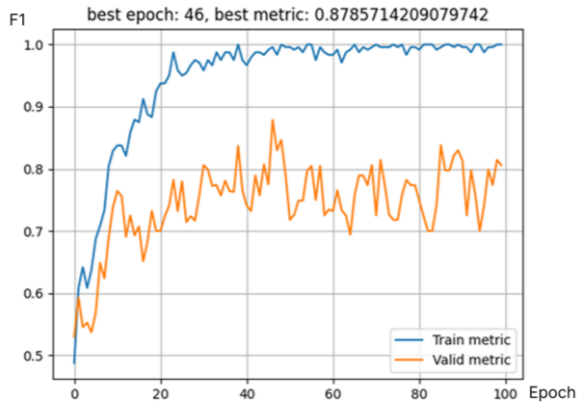



Fig. 7 Training and validation metrics across epochs for ResNet-18 model.

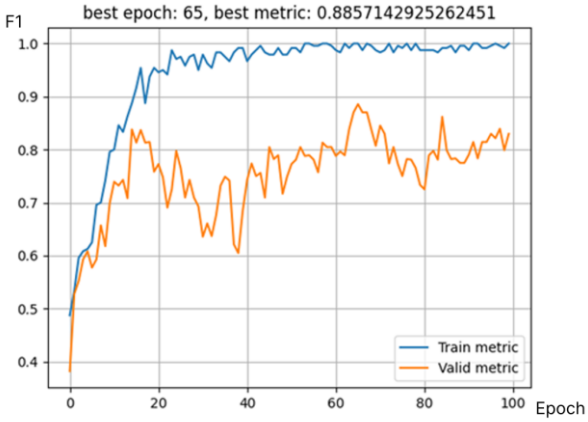



Fig. 8 Training and validation metrics across epochs for ResNet-34 model.

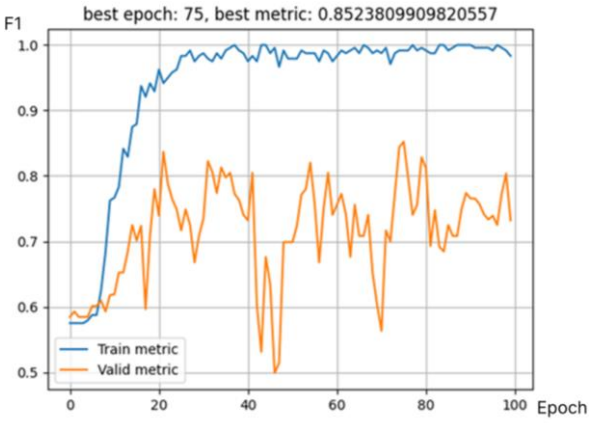



Fig. 9 Training and validation metrics across epochs for ResNet-50 model.

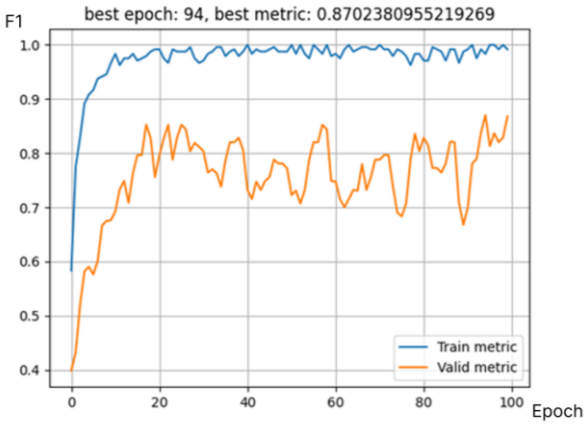



Fig. 10 Training and validation metrics across epochs for EfficientNetB0 model.

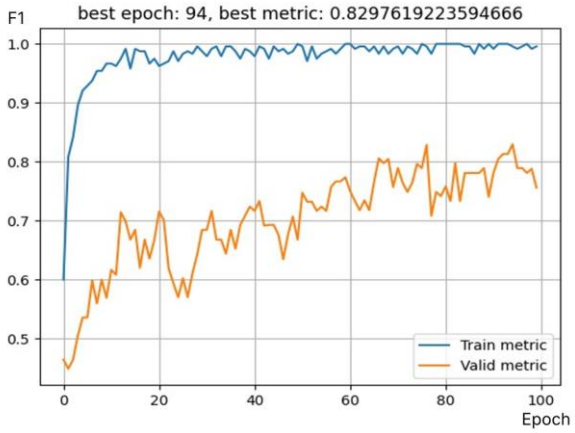



Fig. 11 Training and validation metrics across epochs for EfficientNetB1 model.

Comprehensive metric analysis presented in Table 2 confirms ResNet-34's superior performance across all evaluation criteria. This model achieves an optimal balance between precision and recall while maintaining minimal prediction uncertainty – a crucial consideration in medical diagnostics. The consistently elevated recall values observed across all models reflect an essential clinical principle: the preference for cautious diagnosis over missed pathology.

Table 2 Classification model training results.

| Metric | Architecture | | | | |
|---|---|---|---|---|---|
| | **RN-18** | **RN-34** | **RN-50** | **ENB0** | **ENB1** |
| $F_1$ | 0.8523 | 0.8857 | 0.8524 | 0.8702 | 0.8298 |
| Cross-entropy | 0.3611 | 0.3568 | 0.6736 | 0.8515 | 0.8096 |
| Recall | 0.8273 | 0.8742 | 0.8273 | 0.8811 | 0.8450 |
| Precision | 0.8237 | 0.8742 | 0.8237 | 0.8745 | 0.8387 |
| Accuracy | 0.8273 | 0.8742 | 0.8273 | 0.8811 | 0.8450 |

## *3.3 Complete System Execution Results*

The final system architecture was derived from a comprehensive comparative analysis, integrating the highest-performing models identified in our evaluation. Figure 12 illustrates a representative output of the fully integrated system.

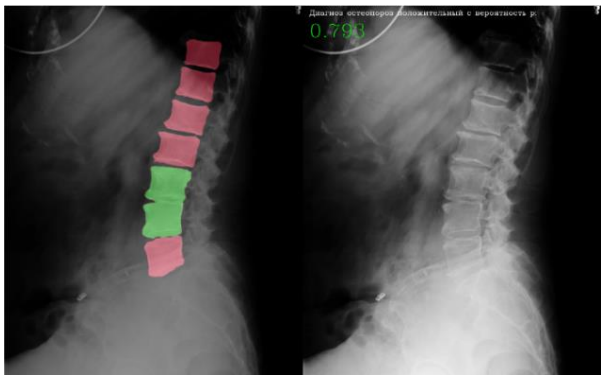


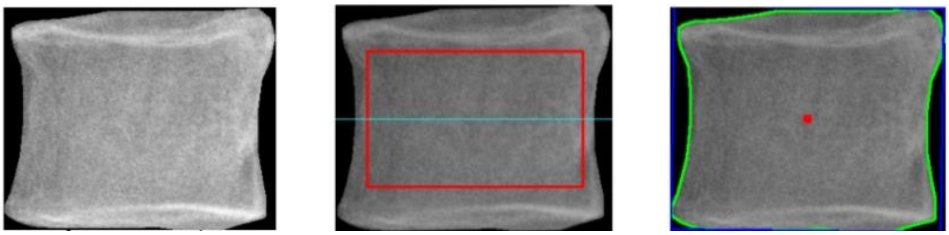
Fig. 12 Output of the resulting system.

## *3.4 Model Optimization Strategy*

The developed neural network for osteoporosis diagnosis from spinal X-rays exhibits a critical limitation: low interpretability. This opacity in decision-making hinders clinical trust and complicates implementation. To address this, we introduce a complementary image processing approach designed to enhance both model transparency and accuracy.

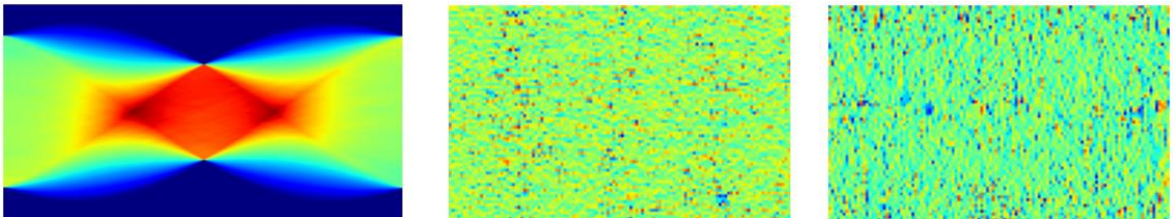
The pipeline utilizes isolated vertebral images, with background removed by the segmentation model, to focus analysis exclusively on bone architecture. Preliminary processing involves contrast enhancement using either adaptive CLAHE or linear stretching to accentuate structural details.

A crucial preprocessing step is precise vertebral orientation. The minimum bounding rectangle is calculated to determine angular deviation from the vertical axis. The vertebral center of mass is subsequently identified through image moments to ensure positional accuracy.



(a) (b) (c)

Fig. 13 Example of preprocessing. (a) After preprocessing, (b) cropping, (c) 0.8° angle.



(a) (b) (c)

Fig. 14 Radon and Wavelet transform visualization. (a) Sinogram, (b) Wavelet (cH1 – Horizontal Details), (c) Wavelet (cV1 – Vertical Details).

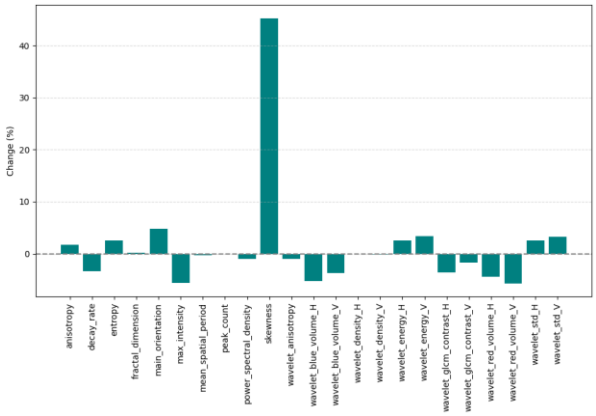



Fig. 15 Percentage difference between healthy and affected vertebrae.

Following alignment, cropping is performed based on anatomical landmarks. The vertical axis of symmetry is established through the center of mass. The width is determined by vertebral boundaries at the center mass level with a 5% margin to exclude edge artifacts. The height is set at 60% of the total image height, centered vertically on the mass coordinate. This method ensures inclusion of the most diagnostically relevant central structures while eliminating peripheral noise. The processed result is demonstrated in Fig. 13.

To extract the most informative features, we implemented two complementary mathematical methods: Radon transform and wavelet analysis. The Radon transform characterizes spatial density distribution patterns within bone tissue, while wavelet analysis facilitates multi-scale texture examination. The implemented code calculates 18 distinct metrics, including principal orientation, anisotropy coefficient, entropy, fractal dimension, and other parameters that quantitatively describe bone tissue status. These metrics constitute a set of transparent and interpretable features for the neural network model, enabling clinicians to identify precisely which structural characteristics of the vertebra contributed to the diagnostic outcome. An example of the applied transformations is presented in Fig. 14.

This methodology was applied to 241 vertebral images. Figure 15 presents a comparative analysis via a bar chart, illustrating the percentage difference in extracted feature values between the two diagnostic classes, thereby highlighting the most discriminative parameters.

## 4 Conclusion

The system developed in this study enables automated osteoporosis diagnosis through the analysis of lumbar spine radiographs. The implementation of the DeepLabV3+ architecture, which achieved a Dice coefficient of 0.8439 for segmentation, coupled with the ResNet-34 model, attaining a classification accuracy of 87.42%, substantiates the system's high efficacy. This integrated approach mitigates radiologist workload and reduces diagnostic subjectivity, offering particular utility in clinical settings with constrained access to advanced modalities such as densitometry.

A principal limitation of the system, however, is the limited interpretability of its results for clinical practitioners. The deployed deep learning models operate as functional "black boxes", obscuring the specific features that drive diagnostic decisions. This opacity may impede trust and hinder seamless integration into routine clinical workflows.

Future work will focus on enhancing the system's transparency by integrating advanced image processing

techniques, including wavelet analysis and the Radon transform. These methods are anticipated to elucidate the structural features of vertebrae that underpin the model's decisions, thereby fostering clinical trust and adoption. Furthermore, we plan to extend the system's analytical capabilities by incorporating additional quantifiable biomarkers of bone tissue integrity. These developments are directed toward augmenting the system's diagnostic robustness and versatility, with the ultimate objective of improving patient care outcomes.

Ethical approval for this study was ensured through the informed voluntary consent obtained from all patients, in accordance with the Order of the Ministry of Health of the Russian Federation (No. 1051n, November 12, 2021).

## Acknowledgments

This work was carried out within the state assignment of NRC "Kurchatov Institute".

## Disclosures

All authors declare that there is no conflict of interest in this paper.

## References

1. J. A. Kanis, C. Cooper, R. Rizzoli, and J. Y. Reginster, "European guidance for the diagnosis and management of osteoporosis in postmenopausal women," Osteoporosis International 30(1), 3–44 (2019).
2. G. Guglielmi (Ed.), Osteoporosis and Bone Densitometry Measurements, Springer Berlin, Heidelberg (2013). ISBN: 978-3-642-27884-6 (eBook).
3. E. M. Eidlyna, G. V. Diachkova, and K. A. Diachkov, "Modern radiation diagnosing the spine pathological fractures thro ugh osteoporosis," Genij Ortopedii 2, 38–43 (2012). [in Russian]
4. A. V. Petraikin, Zh. E. Belaya, A. N. Kiseleva, Z. R. Artyukova, M. G. Belyaev, V. A. Kondratenko, M. E. Pisov, A. V. Solovev, A. K. Smorchkova, L. R. Abuladze, I. N. Kieva, V. A. Fedanov, L. R. Iassin, D. S. Semenov, N. D. Kudryavtsev, S. P. Shchelykalina, V. V. Zinchenko, E. S. Akhmad, K. A. Sergunova, V. A. Gombolevsky, L. A. Nisovstova, A. V. Vladzymyrskyy, and S. P. Morozov, "Artificial intelligence technology for recognition of vertebral compression fractures using a morphometric analysis model based on convolutional neural networks," Problems of Endocrinology 66(5), 48–60 (2020). [in Russian]
5. Zh. A. Klimova, A. A. Zaft, and V. B. Zaft, "Modern laboratory diagnosis of osteoporosis," International Journal of Endocrinology 7.63, 75–84 (2014).
6. A. Gaidel, S. Pervushkin, "Research of the textural features for the bony tissue diseases diagnostics using the roentgenograms," Computer Optics 37(1), 113–119 (2013).
7. I. A. Belozerov, V.A. Sudakov, "Investigation of machine learning models for medical image segmentation," Keldysh Institute Preprints 37, 1–15 (2022).
8. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: a nested U-Net architecture for medical image segmentation," Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support 11045, 3–11 (2018).
9. I. A. Lozhkin, M. E. Dunaev, K. S. Zaytsev, and A. A. Garmash, "Augmentation of image datasets for training neural networks in semantic segmentation tasks," International Journal of Open Information Technologies 11(1), 109–117 (2023).
10. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," Medical Image Computing and Computer-Assisted Intervention 9351, 234–241 (2015).
11. L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," European Conference on Computer Vision 11211, 801–818 (2018).
12. G. Amiya, K. Ramaraj, P.R. Murugan, V. Govindaraj, M. Vasudevan, and A. Thiyagarajan, "A review on automated algorithms used for osteoporosis diagnosis," Lecture Notes in Networks and Systems. Inventive Systems and Control 436, 247–262 (2022).
13. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," IEEE Conference on Computer Vision and Pattern Recognition, 770–778 (2016).
14. D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, "Deep learning in image classification using residual network (resnet) variants for detection of colorectal cancer," Procedia Computer Science 179, 423–431 (2021).
15. M. Tan, Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," International Conference on Machine Learning 97, 6105–6114 (2019).