Traumatology and Orthopaedics of Kazakhstan
ISSN - e2789-9640

Original article

Comparative Analysis of Deep Learning Architectures for Osteoporosis Detection in Knee Radiographs: From Custom CNN to Transfer Learning

Assiya Boltaboyeva [1], Zhanel Baigarayeva [2], Baglan Imanbek [3*], Kassymbek Ozhikenov [4], Zhadyra Alimbayeva [5], Chingiz Alimbayev [6]

[1] PhD student at the Satbayev University; Alfa Center (Al-Farabi AI Center), Al-Farabi Kazakh National University, Almaty, Kazakhstan. E-mail: asiya322m@gmail.com
[2] Researcher at the Satbayev University; Alfa Center (Al-Farabi AI Center), Farabi University, Almaty, Kazakhstan. E-mail: zhanel.baigarayeva@gmail.com
[3] Research Professor at the Alfa Center (Al-Farabi AI Center), Al-Farabi Kazakh National University, Almaty, Kazakhstan. E-mail: imanbek.baglan@kaznu.kz
[4] Professor at the Department of Robotics and Automation Engineering, Satbayev University, Almaty, Kazakhstan. E-mail: k.ozhikenov@satbayev.university
[5] Lecturer at the Department of Information Technology and Library Science, Kazakh National Women's Pedagogical University, Almaty, Kazakhstan. E-mail: zhadyralimbay@gmail.com
[6] Associate Professor at the Department of Robotics and technical means of automation, Satbayev University, Almaty, Kazakhstan. E-mail: c.alimbayev@satbayev.university

***Corresponding author**: imanbek.baglan@kaznu.kz

Citation: Assiya Boltaboyeva, Zhanel Baigarayeva, Baglan Imanbek, Kassymbek Ozhikenov, Zhadyra Alimbayeva, Chingiz Alimbaev. Comparative Analysis of Deep Learning Architectures for Osteoporosis Detection in Knee Radiographs:From Custom CNN to Transfer Learning. Trauma & Ortho Kaz, 2026, 77 (3), jto053. https://doi.org/10.52889/1684-9280-2026-77-3-jto053

Abstract

Osteoporosis is a systemic skeletal disorder affecting millions worldwide, characterized by decreased bone mineral density and increased fracture risk. Early and accurate detection from knee radiographs remains clinically challenging due to the subtlety of trabecular changes and limited availability of annotated datasets.

This study presents a comparative analysis of three deep learning architectures — a custom Convolutional Neural Network (CNN), ResNet50 with partial fine-tuning, and EfficientNet-B3 with full fine-tuning — for binary classification of knee X-ray images into Normal and Osteoporosis categories.

Methods. The Osteoporosis Knee X-ray Dataset (372 images) was used as the benchmark. All models were trained with Contrast-Limited Adaptive Histogram Equalization (CLAHE)-based contrast enhancement, geometric and occlusion augmentations, and a class-weighted cross-entropy loss function to address class imbalance. Early stopping and L2 regularization were applied to improve generalization.

Results. Experimental results demonstrate that EfficientNet-B3 achieved the highest overall performance with an accuracy of 80.00%, recall of 89.74%, and F1-score of 82.35%, making it the most suitable model for clinical screening where sensitivity is prioritized. Custom CNN attained the highest Receiver Operating Characteristic – Area Under the Curve of 87.75%, indicating strong discriminative capacity despite its compact architecture.

Conclusions. ResNet50 with partial fine-tuning showed competitive precision (76.32%) and Area Under the Curve (86.89%). These findings provide insights into the trade-offs between model complexity, transfer learning strategy, and diagnostic performance for small-scale medical imaging tasks.

Keywords: osteoporosis detection, deep learning, transfer learning, EfficientNet, ResNet50, Contrast-Limited Adaptive Histogram Equalization, knee radiograph, convolutional neural network, class imbalance, medical image classification.

1. Introduction

Osteoporosis is a chronic metabolic bone disease defined by reduced bone mass and deterioration of bone microarchitecture, leading to enhanced skeletal fragility and susceptibility to fractures. According to the International Osteoporosis Foundation, osteoporosis affects an estimated 200 million people globally, with fragility fractures occurring every three seconds [1]. The knee joint, in particular, is one of the most frequently affected anatomical regions, and plain radiography remains the most widely accessible modality for initial assessment in low-resource clinical settings.

The clinical gold standard for osteoporosis diagnosis is dual-energy X-ray absorptiometry (DXA), which measures Bone Mineral Density (BMD) and assigns T-scores to classify patients as normal, osteopenic, or osteoporotic. However, DXA scanners are expensive, not universally available, and expose patients to additional radiation doses. This has motivated the development of computer-aided detection systems capable of inferring bone health status directly from conventional radiographs, which are already routinely acquired in orthopedic practice.

Deep learning, and Convolutional Neural Networks (CNNs) in particular, have demonstrated remarkable performance in medical image analysis tasks including chest X-ray interpretation, diabetic retinopathy screening, and histopathology classification [2,3]. Transfer learning from large-scale datasets such as ImageNet has emerged as a critical technique for adapting pre-trained representations to small medical datasets, where collecting and annotating thousands of samples is often impractical [4].

Despite growing interest in automated osteoporosis detection, the literature reveals a lack of systematic comparison between custom architectures and pre-trained models under identical experimental conditions, particularly with respect to class imbalance handling and contrast enhancement preprocessing. Most existing studies either focus on a single architecture [5,6] or apply augmentation without controlled ablations [7].

To address these gaps, this paper makes the following contributions: (1) a fair head-to-head comparison of three architectures — a purpose-built CNN, ResNet50 with partial fine-tuning of its final convolutional block (layer4), and EfficientNet-B3 with full fine-tuning — under identical training protocols; (2) systematic application of Contrast-Limited Adaptive Histogram Equalization (CLAHE) preprocessing to enhance trabecular contrast prior to training; (3) class-weighted cross-entropy loss to handle the imbalanced Normal/Osteoporosis distribution; and (4) evaluation across five complementary metrics: accuracy, precision, recall, F1-score, and Receiver Operating Characteristic – Area Under the Curve (ROC-AUC).

Significant progress has been made in applying deep learning to osteoporosis and bone disease detection across various anatomical sites and imaging modalities.

Initial efforts directly addressed knee-radiograph-based osteoporosis classification using transfer learning, where evaluating AlexNet, VGG-16/19, and ResNet on 381 Quantitative Ultrasound (QUS)-validated knee X-rays demonstrated that a pretrained AlexNet achieved 91.1% accuracy, substantially outperforming the same architecture trained from scratch (79%), underscoring the importance of transfer learning on small radiographic datasets, as shown by Wani and Arora [5].

A multi-architecture benchmark using 1,947 augmented knee X-rays reported VGG-19 as the top-performing model with 97.5% binary accuracy and 92.0% three-class accuracy, notably highlighting the disparity between results on augmented versus original-scale datasets as a critical consideration when interpreting published benchmark figures, according to Sarhan et al. [6].

A custom 54-layer "Superfluity DL" multi-branch architecture was developed to achieve 85.42% on the Mendeley dataset and 79.39% on the Kaggle stevepython benchmark, surpassing AlexNet and ResNet-50 baselines and emphasizing that task-specific architectural decisions can be more beneficial than depth alone on this problem, as introduced by Naguib et al. [7].

KONet, a weighted ensemble of DenseNet-121, EfficientNetB0, ResNet50, VGG-19, MobileNet, and InceptionV3, was shown to demonstrate precision, recall, and F1-scores of approximately 97% under five-fold cross-validation on the same 372-image Kaggle binary set; this indicates the ceiling achievable with ensemble fusion, although its computational complexity makes it unsuitable for resource-constrained clinical deployment (Rasool et al.) [8]).

https://doi.org/10.52889/1684-9280-2026-77-3-jto053

Beyond knee radiographs, EfficientNet-b3 ensembles were applied to hip radiographs (1,131 images), achieving an AUC of 0.937 and demonstrating the benefit of multimodal clinical-covariate fusion, as researched by Yamamoto et al. [9]. Furthermore, combining VGG-16 with non-local neural network modules for hip-radiograph osteoporosis prediction yielded an internal AUC of 0.867 and one of the few external validation results (AUC 0.700) in this literature, as reported by Jang et al. [10].

The use of CLAHE as a preprocessing step for bone radiographs was validated by demonstrating that it significantly improved fracture detection performance in wrist X-rays, as established by Hardalaç et al. [11]. Finally, class imbalance handling through weighted loss functions and Synthetic Minority Over-sampling Technique (SMOTE) variants has been proven as critical for medical classification, as detailed by Chawla et al. [12] and Lin et al. [13].

This study presents a comparative analysis of three deep learning architectures — a custom CNN, ResNet50 with partial fine-tuning, and EfficientNet-B3 with full fine-tuning — for binary classification of knee X-ray images into Normal and Osteoporosis categories.

2. Materials and methods

Dataset Description

This study uses the Osteoporosis Knee X-ray Dataset [14], publicly available on Kaggle, which contains 372 knee radiograph images distributed across two classes: Normal and Osteoporosis. The dataset originates from a clinical camp in India where patients underwent simultaneous knee X-rays and Quantitative Ultrasound System (QUS) T-score measurements, providing ground-truth labels validated by orthopedic specialists [5]. The inherent class imbalance in the binary formulation — with Osteoporosis cases constituting the minority class — is a key challenge that our methodology explicitly addresses. The dataset was split into training (80%, 298 images) and validation (20%, 74 images) subsets using stratified random sampling to preserve class proportions.

Preprocessing: CLAHE Enhancement

All images were resized to 224 × 224 pixels to conform with ImageNet pretraining conventions. Contrast Limited Adaptive Histogram Equalization (CLAHE) [15] was applied uniformly to both training and validation sets. CLAHE operates by dividing the image into a grid of contextual tiles (tile size 8 × 8 in this work), performing histogram equalization locally within each tile, and applying a clip limit (set to 4.0) to prevent noise amplification inherent to standard adaptive HE. For bone radiographs, CLAHE has been shown to significantly enhance the visibility of trabecular microstructure — the primary visual cue for osteoporosis assessment [11].

Data Augmentation

To reduce overfitting on the small training set, the following augmentation pipeline was applied exclusively to the training data using the Albumentations library: (1) Horizontal flip (p=0.5); (2) ShiftScaleRotate with shift limit 0.05, scale limit 0.05, and rotation limit ±15° (p=0.5); (3) CoarseDropout with 1–3 rectangular patches of 10–30 pixels (p=0.5). All images were normalized with ImageNet mean (0.485, 0.456, 0.406) and standard deviation (0.229, 0.224, 0.225).

Methodology

This study evaluates three deep learning architectures under identical experimental conditions to ensure a fair comparison. The first model is a purpose-built Custom CNN trained from random initialization, serving as a non-pretrained baseline. It comprises three sequential convolutional blocks — each consisting of a Conv2D layer (3×3 kernel, padding 1), Batch Normalization, ReLU activation, and 2×2 Max Pooling — with channel dimensions expanding progressively from 3 to 32, 64, and 128. The resulting 128×28×28 feature maps are flattened and passed through a two-layer fully connected head (512 hidden units, ReLU, Dropout p=0.6, 2-class output), with a learning rate of 1 × {10}^{-4}. The second model, ResNet50 [16], was initialized with ImageNet weights and trained using a selective fine-tuning strategy: the entire backbone was first frozen, after which only the final residual stage (layer4) was unfrozen to enable domain adaptation of high-level feature detectors while preserving low-level pretrained representations. Its classifier was replaced by Dropout (p = 0.6) followed by a Linear layer (2048 → 2), trained at a conservative learning rate of 1 × {10}^{-5}. to prevent catastrophic forgetting. The third model, EfficientNet-B3 [17], was also initialized with ImageNet weights but subjected to full fine-tuning across all layers, allowing complete domain adaptation. Its compound scaling strategy — simultaneously optimizing network depth (1.4×), width (1.2×), and input resolution (1.3) provides a superior accuracy-to-parameter ratio compared to single-dimension scaling approaches. The classifier head was replaced with Dropout (p = 0.5) and a Linear layer (1536 → 2), trained at 1×10^{-5}.

All three models shared an identical training protocol. Class-weighted cross-entropy loss was applied, with per-class weights computed via scikit-learn's 'balanced' weighting scheme to penalize misclassification of the minority osteoporosis class. The Adam optimizer was used with L2 weight decay (λ = 1

× 10⁻⁴) for regularization. Training was capped at 40 epochs with early stopping triggered after 7 consecutive epochs without validation loss improvement, and the best checkpoint — defined by minimum validation loss — was retained for final evaluation. All experiments were conducted on Graphics Processing Unit (GPU) (Compute Unified Device Architecture (CUDA)) with a batch size of 32.

3. Results

Quantitative Performance Comparison

Table 1 presents the evaluation metrics for all three models on the validation set. EfficientNet-B3 achieves the highest accuracy (80.00%), recall (89.74%), and F1-score (82.35%), establishing it as the best-performing model overall. Notably, the Custom CNN attains the highest ROC-AUC (87.75%), indicating stronger overall discriminative capacity across all classification thresholds.

Table 1 - Performance comparison of deep learning models on the validation set

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
EfficientNet-B3	0.8000	0.7609	0.8974	0.8235	0.8348
ResNet50	0.7467	0.7632	0.7436	0.7532	0.8689
Custom CNN	0.7867	0.7805	0.8205	0.8000	0.8775

Training Dynamics of EfficientNet-B3

Figure 1 illustrates the training and validation loss and accuracy curves for EfficientNet-B3 over 40 training epochs. The loss curves (left panel) show a consistent and synchronized decline for both training and validation loss, with no significant divergence between the two, indicating controlled overfitting under the applied regularization strategy (Dropout, weight decay, early stopping). Notably, the validation loss tracks slightly below the training loss in several epochs, which can be attributed to the use of class-weighted loss and the stochastic nature of augmentation applied only to training data.

The accuracy curves (right panel) demonstrate that validation accuracy reaches approximately 80% and stabilizes by epoch 25–30, while training accuracy continues to grow gradually. This pattern is consistent with well-regularized fine-tuning behavior on a small dataset. The absence of a sharp divergence between training and validation curves confirms that CLAHE preprocessing and data augmentation effectively prevented overfitting despite the limited dataset size of 298 training images.

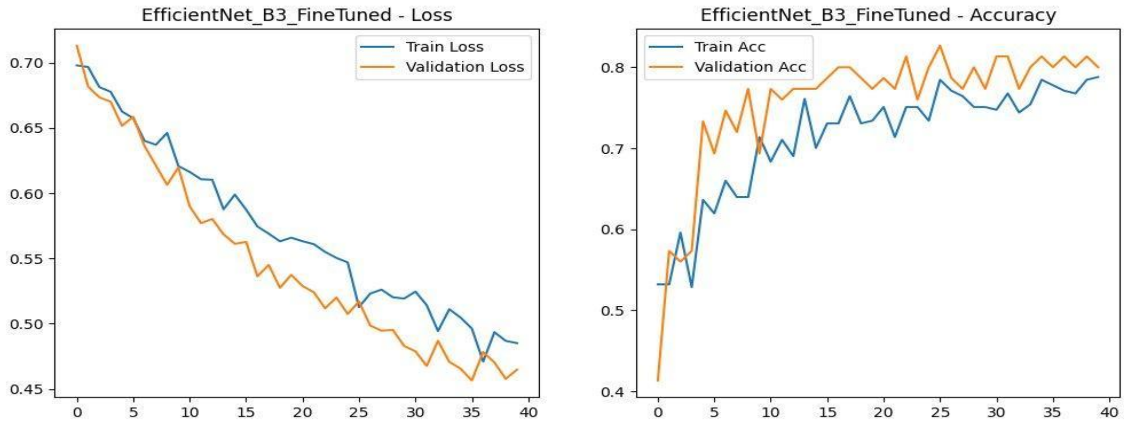

Figure 1 - Training and validation loss (left) and accuracy (right) curves for EfficientNet-B3 over 40 epochs

4. Discussion

Across all three models, a clear performance hierarchy emerges that reflects the interplay between architecture complexity, fine-tuning strategy, and dataset scale. EfficientNet-B3 achieves the strongest overall results with the highest accuracy (80.00%), recall (89.74%), and F1-score (82.35%), making it the most clinically suitable model for osteoporosis screening, where minimizing false negatives — missed diagnoses

— carries greater consequences than false positives. Despite this, its ROC-AUC (83.48%) is the lowest among the three models, suggesting that full fine-tuning on a dataset of only 372 images may introduce slight threshold-level confidence bias, causing the model to perform well at its default operating point but less consistently across the full range of classification thresholds.

The Custom CNN presents an instructive counterpoint: trained entirely from random initialization without any pretrained features, it nonetheless achieves 78.67% accuracy, 80.00% F1-score, and the highest ROC-AUC of 87.75%. The superior discriminative capacity across thresholds suggests that a compact, task-specific architecture — regularized with high Dropout (0.6) and Batch Normalization — can learn a more generalized decision boundary on this narrow domain than a heavily parameterized pretrained model forced to adapt from object-recognition features. This is consistent with the findings of Raghu et al. [18], who demonstrated that transfer learning on small medical datasets primarily accelerates convergence rather than delivering large accuracy gains. ResNet50, trained with partial fine-tuning limited to its final residual block (layer4), shows the weakest accuracy (74.67%) and F1-score (75.32%), though its AUC (86.89%) remains competitive. The constrained fine-tuning depth appears insufficient to adapt ResNet50's representations to the subtle trabecular texture changes that characterize osteoporotic knee radiographs; full fine-tuning of the backbone warrants investigation in future work.

Placed in the context of existing literature, the results obtained here are competitive given the experimental constraints. Studies reporting 90–98% accuracy on the same benchmark — including Sarhan et al. [6] at 97.5% and KONet [8] at approximately 97% — employed either substantially larger augmented training sets (up to 1,947 images) or weighted multi-model ensembles evaluated under cross-validation, both of which materially inflate the effective training distribution relative to our single-split protocol. The most directly comparable result in the literature is Naguib et al. [7], whose purpose-built Superfluity DL architecture achieved 79.39% on the same Kaggle dataset under similar conditions; our EfficientNet-B3 result of 80.00% is consistent with, and marginally above, that published baseline, providing independent validation of the reported performance range for this benchmark.

Several limitations must be acknowledged. The dataset of 372 images is small relative to clinical deployment requirements, and all reported metrics are from a single validation split without external validation. Future work should incorporate k-fold cross-validation, external test cohorts with independent Dual-energy X-ray absorptiometry ground truth, explainability tools such as Grad-CAM, and ensemble or hybrid CNN-Transformer architectures that have recently shown promise on this benchmark [8,19].

5. Conclusions

This study conducted a systematic comparative evaluation of three deep learning architectures for binary classification of knee radiographs into Normal and Osteoporosis categories. EfficientNet-B3 with full fine-tuning emerged as the best overall model, achieving 80.00% accuracy, 89.74% recall, and 82.35% F1-score. The Custom CNN demonstrated the strongest discriminative capacity by ROC-AUC (87.75%), challenging the assumption that transfer learning always yields superior generalization on small radiographic datasets. ResNet50's partial fine-tuning strategy was insufficient to fully exploit its pretrained representations, suggesting that more aggressive unfreezing may improve performance.

Funding. This research was funded by the Ministry of Science and Higher Education of the Republic of Kazakhstan under the scientific grant BR24992820, titled «Innovative medical technologies and devices aimed at improving surgical interventions for prosthetics and rehabilitation in the field of orthopedics and medical rehabilitation».

Conflict of Interest. The authors declare no conflict of interest.

Author Contributions. Conceptualization — A.B.; Data collection and analysis – Zh.B., B.I., Zh.A.; Writing-original draft – A.B.; Writing-review & editing – A.B., K.O., Ch.A.

AI Declaration. The authors declare that no AI or AI-assisted technologies were used in the writing or preparation of this manuscript.

References

1. Hernlund, E., Svedbom, A., Ivergard, M., Compston, J. E., Cooper, C., Stenmark, J., McCloskey, E. V., Jönsson, B., & Kanis, J. A. (2013). Osteoporosis in the European Union: Medical management, epidemiology and economic burden. Archives of Osteoporosis, 8, 136. https://doi.org/10.1007/s11657-013-0136-1

2. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005
3. Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., & Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. Journal of Medical Imaging and Radiation Oncology, 65(5), 545–563. https://doi.org/10.1111/1754-9485.13261
4. Tajbakhsh, N., Shin, J. Y., Gurumurthy, S. R., Hurst, A., Agrawal, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Transactions on Medical Imaging, 35(5), 1299–1312. https://doi.org/10.1109/TMI.2016.2535302
5. Xu, S., Yao, Q., Liu, G., Wang, M., Gu, J., & Guo, J. (2020). Combining DWI radiomics features with transurethral resection promotes the differentiation between muscle-invasive bladder cancer and non-muscle-invasive bladder cancer. European Radiology, 30, 1804–1812. https://doi.org/10.1007/s00330-019-06484-2
6. Sarhan, A. M., Gobara, M., Yasser, S., Elsayed, Z., Sherif, G., Moataz, N., ... & Ali, H. A. (2024). Knee osteoporosis diagnosis based on deep learning. International Journal of Computational Intelligence Systems, 17(1), 241. https://doi.org/10.1007/s44196-024-00615-4
7. Naguib, S. M., Saleh, M. K., Hamza, H. M., Hosny, K. M., & Kassem, M. A. (2024). A new superfluity deep learning model for detecting knee osteoporosis and osteopenia in X-ray images. Scientific Reports, 14(1), 25434. https://doi.org/10.1038/s41598-024-75549-0
8. Rasool, M. A., Ahmad, S., Sabina, U., & Whangbo, T. K. (2024). Konet: Toward a weighted ensemble learning model for knee osteoporosis classification. IEEE Access, 12, 5731-5742. https://doi.org/10.1109/ACCESS.2023.3348817
9. Yamamoto, N., Sukegawa, S., Kitamura, A., Goto, R., Noda, T., Nakano, K., ... & Ozaki, T. (2020). Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates. Biomolecules, 10(11), 1534. https://doi.org/10.3390/biom10111534
10. Jang, R., Choi, J. H., Kim, N., Chang, J. S., Yoon, P. W., & Kim, C. H. (2021). Prediction of osteoporosis from simple hip radiography using deep learning algorithm. Scientific reports, 11(1), 19997. https://doi.org/10.1038/s41598-021-99549-6
11. Hardalaç, F., Uysal, F., Peker, O., Çiçeklidağ, M., Tolunay, T., Tokgöz, N., ... & Mert, F. (2022). Fracture detection in wrist X-ray images using deep learning-based object detection models. Sensors, 22(3), 1285. https://doi.org/10.3390/s22031285
12. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953
13. Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2020). Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2), 318–327. https://doi.org/10.1109/TPAMI.2018.2858826
14. Halabi, S. S., Prevedello, L. M., Kalpathy-Cramer, J., Mamonov, A. B., Bilbily, A., Cicero, M., Pan, I., Pereira, L. A., Sousa, R. T., Abdala, N., Kitamura, F. C., Thodberg, H. H., Chen, L., Shih, G., Andriole, K., Kohli, M. D., Erickson, B. J., & Flanders, A. E. (2019). The RSNA pediatric bone age machine learning challenge. Radiology, 290(2), 498–503. https://doi.org/10.1148/radiol.2018180736
15. Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., Terzopoulos, D., Avery, M., & Zimmerman, J. B. (1987). Adaptive histogram equalization and its variations. Computer Vision, Graphics, and Image Processing, 39(3), 355–368. https://doi.org/10.1016/S0734-189X(87)80186-X
16. Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031
17. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. https://doi.org/10.1145/3065386
18. Ye, H. (2025). Artificial intelligence combined with computed tomography or X-ray radiography: Potential solution for opportunistic screening for osteoporosis. European Journal of Radiology Artificial Intelligence, 3, 100036. https://doi.org/10.1016/j.ejrai.2025.100036
19. Burghardt, A. J., Link, T. M., & Majumdar, S. (2011). High-resolution computed tomography for clinical imaging of bone microarchitecture. Clinical Orthopaedics and Related Research, 469(8), 2179–2193. https://doi.org/10.1007/s11999-010-1766-x

Тізе буынының рентгенограммалары бойынша остеопорозды анықтауға арналған терең оқыту архитектураларын салыстырмалы талдау: Пайдаланушылық тұйықталған нейрондық желілерден трансферлік оқытуға дейін

Болтабоева А.К. [1], Байғараева Ж.Е. [2], Иманбек Б.Т. [3*], Ожикенов К.А. [4], Алимбаева Ж.Н. [5], Алимбаев Ч.А. [6]

[1] PhD докторант, Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті; Alfa Center (Al-Farabi AI Center), Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан. E-mail: asiya322m@gmail.com

[2] Зерттеуші, Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті; Alfa Center (Al-Farabi AI Center), Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан. E-mail: zhanel.baigarayeva@gmail.com

[3] Зерттеуші профессор, Alfa Center (Al-Farabi AI Center), Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан. E-mail: imanbek.baglan@kaznu.kz

[4] Профессор, Робототехника және автоматика құралдары кафедрасы, Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті, Алматы, Қазақстан. E-mail: k.ozhikenov@satbayev.university

[5] Оқытушы, Ақпараттық технологиялар және кітапхана ісі кафедрасы, Қазақ ұлттық қыздар педагогикалық университеті, Алматы, Қазақстан. E-mail: zhadyralimbay@gmail.com

[6] Қауымдастырылған профессор, Робототехника және автоматика құралдары кафедрасы, Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті, Алматы, Қазақстан. E-mail: c.alimbayev@satbayev.university

Түйіндеме

Остеопороз — сүйек ұлпасының минералды тығыздығының төмендеуімен және сыну қаупінің артуымен сипатталатын, бүкіл әлем бойынша миллиондаған адамға әсер ететін жүйелі қаңқа ауруы. Трабекулярлық өзгерістердің өте нәзік болуына және таңбаланған деректер жиынтығының шектеулілігіне байланысты тізе буынының рентгенограммасы арқылы ауруды ерте әрі дәл анықтау клиникалық тұрғыдан күрделі мәселе болып қалуда.

Бұл зерттеуде тізе буынының рентген суреттерін «Қалыпты» және «Остеопороз» санаттарына бинарлы жіктеулерге арналған терең оқытудың үш архитектурасына — кастомды (пайдаланушылық) тұйықталған нейрондық желісіне (CNN), ішінара дәл баптаудан өткен ResNet50 және толық дәл баптаудан өткен EfficientNet-B3 үлгілеріне салыстырмалы талдау жүргізілді.

Әдістері. Эталондық деректер базасы ретінде остеопороз кезіндегі тізе рентгенінің деректер жиынтығы (372 кескін) пайдаланылды. Барлық модельдер контрастты арттыруға арналған контрасты шектелген адаптивті гистограмманы теңестіру әдісін, геометриялық және окклюзиялық аугментацияларды, сондай-ақ кластар теңгерімсіздігін жою үшін кластық-салмақталған кросс-энтропиялық шығын функциясын қолдану арқылы оқытылды. Модельдердің жалпылау қабілетін жақсарту үшін ерте тоқтату және L2-регуляризациясы қолданылды.

Нәтижелері. Эксперименттік нәтижелер EfficientNet-B3 моделінің 80,00% дәлдік, 89,74% толықтық және 82,35% F1-өлшемін көрсетіп, ең жоғары жалпы өнімділікке қол жеткізгенін көрсетті. Бұл оны сезімталдыққа басымдық берілетін клиникалық скрининг үшін ең қолайлы модель етеді. Пайдаланушылық CNN өзінің жинақы архитектурасына қарамастан, жоғары айыру қабілетін байқатып, ең жоғары қабылдағыштың жұмыс сипаттамасы – қисық астындағы аудан (ROC-қисығы) көрсеткішіне (87,75%) жетті. Ішінара дәл баптаудан өткен ResNet50 бәсекеге қабілетті дәлдік (76,32%) пен қисық астындағы аудан (86,89%) көрсеткіштерін байқатты.

Қорытынды. Алынған нәтижелер шағын ауқымды медициналық бейнелерді өңдеу тапсырмаларында модельдің күрделілігі, трансферлік оқыту стратегиясы және диагностикалық тиімділік арасындағы өзара байланыстар мен таңдаулар туралы құнды мәліметтер береді.

Түйін сөздер: остеопорозды анықтау, терең оқыту, трансферлі оқыту, EfficientNet, ResNet50, контрасты шектелген адаптивті гистограмма, тізе рентгенографиясы, пайдаланушылық нейрондық желі, кластар теңгерімсіздігі, медициналық кескіндерді жіктеу.

https://doi.org/10.52889/1684-9280-2026-77-3-jto053

Сравнительный анализ архитектур глубокого обучения для обнаружения остеопороза по рентгенограммам коленного сустава: От кастомных сверточных нейронных сетей до трансферного обучения

Болтабоева А.К. [1], Байғараева Ж.Е. [2], Иманбек Б.Т. [3*], Ожикенов К.А. [4], Алимбаева Ж.Н. [5], Алимбаев Ч.А. [6]

[1] PhD докторант, Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Alfa Center (Al-Farabi AI Center), Казахский национальный университет имени Аль-Фараби, Алматы, Казахстан. E-mail: asiya322m@gmail.com

[2] Исследователь, Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Alfa Center (Al-Farabi AI Center), Казахский национальный университет имени Аль-Фараби, Алматы, Казахстан. E-mail: zhanel.baigarayeva@gmail.com

[3] Профессор-исследователь, Alfa Center (Al-Farabi AI Center), Казахский национальный университет имени Аль-Фараби, Алматы, Казахстан. E-mail: imanbek.baglan@kaznu.kz

[4] Профессор кафедры робототехники и технических средств автоматизации, Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Алматы, Казахстан. E-mail: k.ozhikenov@satbayev.university

[5] Преподаватель кафедры информационных технологий и библиотечного дела, Казахский национальный женский педагогический университет, Алматы, Казахстан. E-mail: zhadyralimbay@gmail.com

[6] Ассоциированный профессор кафедры робототехники и технических средств автоматизации, Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Алматы, Казахстан. E-mail: c.alimbayev@satbayev.university

Резюме

Остеопороз — это системное заболевание скелета, поражающее миллионы людей по всему миру и характеризующееся снижением минеральной плотности костной ткани и повышенным риском переломов. Ранняя и точная диагностика по рентгенограммам коленного сустава остается сложной клинической задачей из-за едва заметных изменений трабекулярной структуры и ограниченной доступности размеченных наборов данных.

В данном исследовании представлен сравнительный анализ трех архитектур глубокого обучения — кастомной (пользовательской) сверточной нейронной сети (CNN), ResNet50 с частичной тонкой настройкой и EfficientNet-B3 с полной тонкой настройкой — для бинарной классификации рентгеновских снимков коленного сустава на категории «Норма» и «Остеопороз».

Методы. В качестве эталонного датасета использовался набор рентгеновских снимков колена при остеопорозе (372 изображения). Все модели обучались с применением метода контрастно-ограниченной адаптивной эквализации гистограммы для повышения контрастности, геометрической аугментации и аугментации перекрытия, а также взвешенной по классам функции потерь кросс-энтропии для устранения дисбаланса классов. Для повышения обобщающей способности моделей применялись раннее прекращение обучения и L2-регуляризация.

Результаты. Экспериментальные результаты показывают, что EfficientNet-B3 продемонстрировала наилучшую общую производительность с точностью 80,00%, полнотой 89,74% и F1-мерой 82,35%, что делает ее наиболее подходящей моделью для клинического скрининга, где приоритетом является чувствительность. Кастомная CNN достигла наивысшего значения рабочей характеристики приемника — площади под кривой (87,75%), что указывает на высокую способность к дискриминации, несмотря на компактную архитектуру. Модель ResNet50 с частичной тонкой настройкой показала конкурентоспособную точность (76,32%) и площадь под кривой (86,89%).

Выводы. Полученные результаты позволяют оценить компромиссы между сложностью модели, стратегией трансферного обучения и диагностической эффективностью при решении задач классификации медицинских изображений на малых выборках.

Ключевые слова: обнаружение остеопороза, глубокое обучение, трансферное обучение, EfficientNet, ResNet50, контрастно-ограниченная адаптивная эквализация гистограммы, рентгенография коленного сустава, сверточная нейронная сеть, дисбаланс классов, классификация медицинских изображений.

https://doi.org/10.52889/1684-9280-2026-77-3-jto053