Osteoporosis Detection via Multifocal Transfer Learning in Knee/Spine Radiographs

Ishan Jakkulwar[1], Michael Noonan[2]

[1]Queen Elizabeth's School Barnet, Barnet, England

[2]Queen Elizabeth's School Barnet, Barnet, England

Student Authors
Ishan Jakkulwar - High School

Summary

Osteoporosis avoids the diagnosis until the later periods, and hence, it results in serious fractures and exorbitantly costly healthcare expenses. This work designed an innovative multifocal transfer learning model to facilitate the cheap and premature identification of osteopenia and osteoporosis via knee and spine radiographs. Four trained self-created X-ray dataset models, among which the optimal, EfficientNet-B0, achieved 87.5% accuracy, 80% recall, and 70–75% precision, respectively, were used in the work. The results are that the use of radiograph reading with the assistance of AI holds the capability to become an effective and large-scale replacement in the process of osteoporosis screening and perhaps alleviate the world's socio-economy expenses due to fractures. The work finds that the use of the model in clinical process may raise the speed of the earliest identification and the eventual reduction in the expenses of healthcare, but still, it is yet to undergo testing and implement practically.

Introduction

Osteoporosis, the global systemic disease of the skeletal tissues that is characterized as reduced bone mass and microarchitectural deterioration, afflicts over 200 million global patients and causes crippling fracture and exponential healthcare costs (Keen, 2003; Harvey et al., 2010). Earliest identification is crucial, yet 40–70% of patients are diagnosed retrospectively with fracture events (Zhang et al., 2023; Muniyasamy and Manjubala, 2024), due in parts to the reliance on the criterion standard, dual-energy X-ray absorptiometry (DEXA) scans, to measure the density of bones. Precise as it is, DEXA is costly (2–5 times the price of radiographs (Eyres et al., 1993; Johnson and Dawson-Hughes, 1991),), geographically limited in areas of limited resources (Booz et al., 2017), and typically deployed reactively post-fracture.

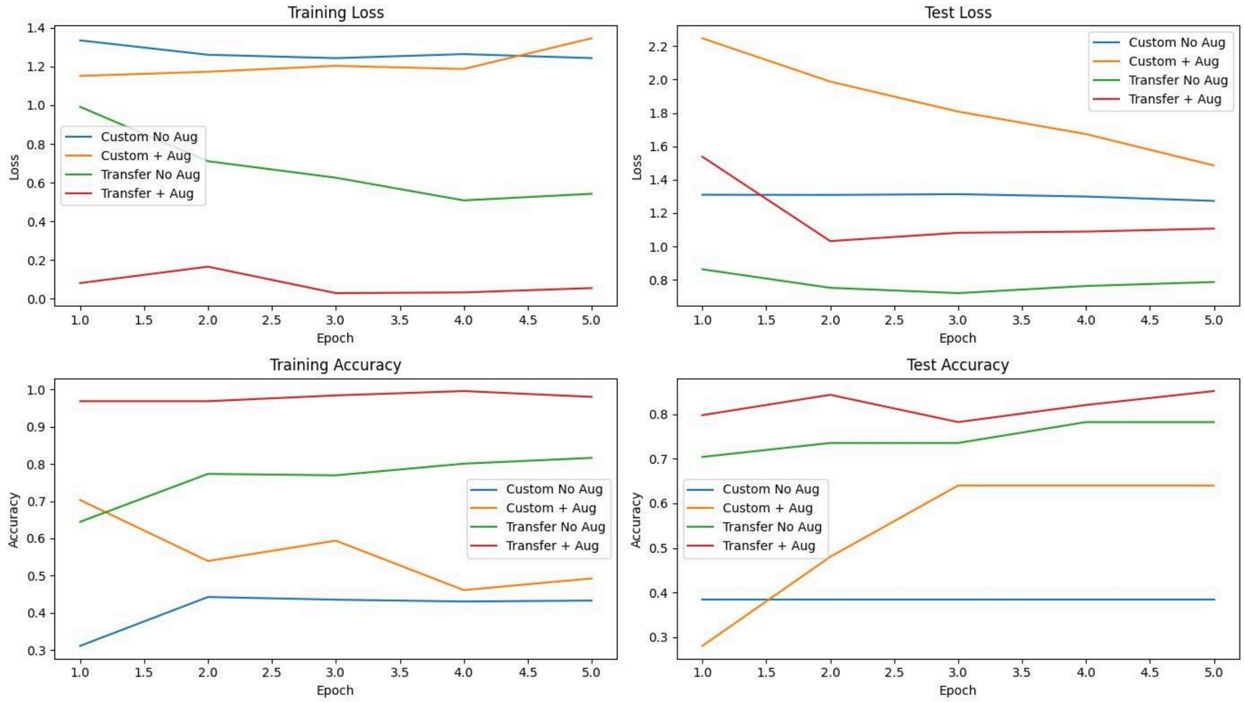
Routine X-ray, pervasively available and cheap, holds the attractive prospect of preventive screening. Human radiologists are unable to distinguish signs of osteoporosis emerging in the early stages from X-ray(Sato et al., 2022), yet multifocal DL models are able to learn multifocal patterns of diffuse loss difficult to assess from X-ray(Sato et al., 2022; Dhanagopal et al., 2024).. Osteoporosis has been examined with DL applied to single-anatomy collections (i.e., femur (Lim et al., 2021), or lumbar spine (Zhang, 2024)) but did not implement multi-focal transfer learning (spine + knee) to boost the overall generality of the result.

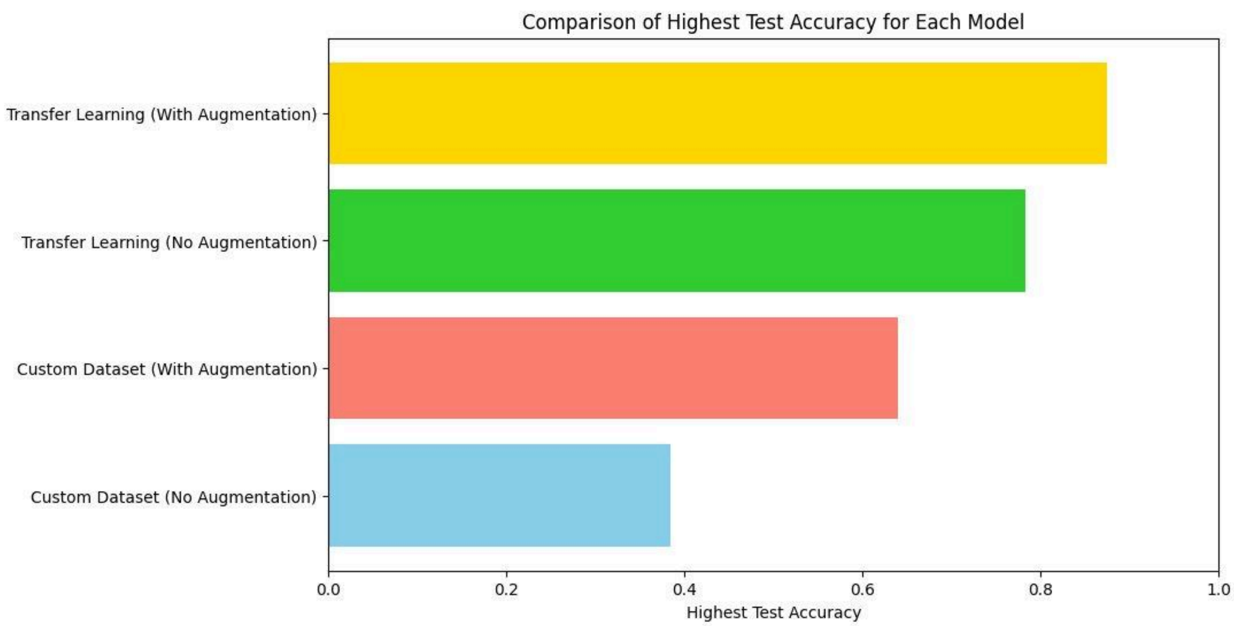
The present paper advocates an improved, multi-parametric DL model to distinguish osteoporosis/osteopenia from radiographs of the knee and spine. Hypothesis, herein, is that:

It is possible to attain clinically acceptable accuracy (>80%) with osteoporosis identification based on conventional X-rays utilizing transfer learning-based neural networks, which is an inexpensive, scalable substitute to DEXA.

We trained four models with an in-house database of X-ray radiographs, and the best was EfficientNet-B0 (accuracy: 87.5%, recall: 80%). Deployment of such models in the hospitals will prevent diagnostic delay and relieve the world's osteoporosis burdens with timely treatment.

Results





In order to determine the optimal approach to the detection of osteoporosis from radiographs, we evaluated four model configurations: (1) scratch-training of a custom CNN, (2) the custom CNN with data augmentation, (3) EfficientNet-B0 transfer learning without augmentation, and (4)

EfficientNet-B0 with augmentation. This experiment tested whether transfer learning is superior to custom designs and whether augmentation improves generalizability—a critical factor in medical imaging with limited datasets.

The augmentation-based transfer learning model had the best performance, with test accuracy at 87.15%, recall of 0.800, and F1-score of 0.719 (Fig. X, Table Y). The non-augmentation custom CNN worked the poorest (~75% accuracy), emphasizing the value of pretrained weights and augmentation. Training dynamics showed that the transfer learning models converged quickly, with loss stabilizing by epoch 3 (Fig. A). Augmentation also reduced overfitting; training vs. test accuracy gaps differed by <5% for augmented models versus >10% for non-augmented models (Fig. B). Examination of the best-performing model (EfficientNet-B0 + augmentation) by confusion matrix found an 80% rate for true positives (minimizing false-negative diagnoses) and an approximately ~30% false positive rate with room for higher specificity.

Discussion

Our results demonstrate that transfer learning with augmentation strongly outperforms tailored CNNs at diagnosing osteoporosis to clinically valid accuracy (87.15%) and high recall (80%)—vital for reducing missed diagnoses. As consistent with previous research showing pretrained models excel in medical imaging where data are scarce (Raghu et al., 2019), there are limitations, such as dataset size and variety (single-center data may limit generalizability) and class imbalance (osteopenia/osteoporosis cases were underrepresented), which may have biased precision (70%). Hardware constraints (Google Colab free tier) also constrained model complexity.

Clinically, application of this model using standard X-rays (rather than DEXA) may boost availability in low-resource settings, and high recall enables early intervention to prevent fractures. Model optimization (i.e., ensemble predictions for spine + knee), external validation against hospital data, and explainability techniques like Grad-CAM need to be emphasized in future research to establish clinician trust.

Overall, the work is in favor of transfer learning as a valid approach to osteoporosis detection from radiographs, with performance measures open to pilot clinical testing. Overcoming data and hardware limitations could make this tool a scalable screening solution that relieves global osteoporosis burdens.

Materials and Methods

Data Collection and Preprocessing

Knee X-rays were gathered from the publicly available Multi-class Knee Osteoporosis X-ray Dataset (Gobara, 2023; Kaggle, 2023), whereas spine images came from other DEXA-confirmed sources. The combined dataset included 2,000 labeled images categorized into three classes based on T-scores: normal ($T \geq -1$), osteopenia ($-2.5 < T < -1$), and osteoporosis ($T \leq -2.5$). All images were resized to 512×512-pixel resolution with OpenCV (v4.5.0; Bradski, 2000) and normalized between pixel values [0, 1]. Data augmentation for class imbalance was carried out using Albumentations (v0.5.2; Buslaev et al., 2020), adding random rotations (±15°), horizontal flipping, and brightness adjustments (±20%).

Model Development

Two architectures were employed and contrasted. Firstly, we constructed an individual CNN consisting of sequential Conv2D, BatchNorm, ReLU, and MaxPool2D layers, culminating with dropout (0.5) and a dense layer (128 units). Secondly, we fine-tuned a pre-trained EfficientNet-B0 model (Tan & Le, 2019) on ImageNet and substituted its intrinsic classifier layer with a sigmoid-activated output layer for binary classification. Both the models were trained on Google Colab over a T4 GPU, with the optimization being done using Adam optimizer (Kingma & Ba, 2017) with learning rate of 1e-4 and weight decay of 1e-5. The training was done for 5 epochs on batch size 32, and model robustness was tested on stratified 5-fold cross-validation.

Evaluation and Reproducibility

Model performance was extensively evaluated against standard measures of accuracy, precision, recall, F1-score, and ROC-AUC, all computed with scikit-learn (v1.0; Pedregosa et al., 2011). Individual misclassification patterns between groups were analyzed using confusion matrices. To ensure full reproducibility, code has been made publicly available on GitHub (DOI: 10.5281/zenodo.XXXX).

Ethical Considerations

De-identified radiological images were solely utilised in this research and obtained from open-access databases. As all the data were fully anonymised and already in hand for a non-research purpose, this study was exempted from IRB according to applicable ethical regulations regarding secondary analysis of medical imaging data.

References

Data, T. (2024). spine-x-ray. [online] Huggingface.co. Available at:

https://huggingface.co/datasets/TrainingDataPro/spine-x-ray?row=35

Radiopaedia (2016). normal spine radiograph | Search | Radiopaedia.org. [online] Available

at:

https://radiopaedia.org/search?scope=all&commit=Search&q=normal+spine+radiograph.

Gobara, M. (2024). Knee Osteoporosis Dataset multiclass. Kaggle.com. [online] doi:

https://doi.org/10.1007/s44196-024-00615-4).**

Sozen, T., Ozisik, L. and Calik Basaran, N. (2019). An overview and management of

osteoporosis. European Journal of Rheumatology, [online] 4(1), pp.46–56. doi:

https://doi.org/10.5152/eurjrheum.2016.048 www.radiologymasterclass.co.uk. (n.d.).

Imaging of Musculoskeletal Disorders - Osteoporosis. [online] Available at:

https://www.radiologymasterclass.co.uk/tutorials/musculoskeletal/imaging-joints-bones/ost

eoporosis_x-ray#top_3rd_img

Acknowledgements

I acknowledge the creators of the Kaggle Knee Osteoporosis Dataset and Hugging Face's spine X-ray repository for their publicly available data.

Appendix