Bone Quality Classification of Dual Energy X-ray Absorptiometry Images Using Convolutional Neural Network Models

Mailen Gonzalez[1], José M. Fuertes García[2], Manuel J. Lucena López[3],
Rubén Abdala[4], José M. Massa[5]
INTIA, Universidad Nacional Del Centro De La Provincia De Buenos Aires (UNCPBA), Tandil, Buenos Aires, Argentina[1,5]
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina[1]
Instituto de Diagnóstico e Investigaciones Metabólicas (IDIM), CABA, Buenos Aires, Argentina[3]
Departamento de Informática, Escuela Politécnica Superior, Universidad de Jaén, Jaén, Spain[2,3]

***Abstract*—The assessment of bone trabecular quality degradation is important for the detection of diseases such as osteoporosis. The gold standard for its diagnosis is the Dual Energy X-ray Absorptiometry (DXA) image modality. The analysis of these images is a topic of growing interest, especially with artificial intelligence techniques. This work proposes the detection of a degraded bone structure from DXA images using some approaches based on the learning of Trabecular Bone Score (TBS) ranges. The proposed models are supported by intelligent systems based on convolutional neural networks using two kinds of approaches: ad hoc architectures and knowledge transfer systems in deep network architectures, such as AlexNet, ResNet, VGG, SqueezeNet, and DenseNet retrained with DXA images. For both approaches, experimental studies were made comparing the proposed models in terms of effectiveness and training time, achieving an F1-Score result of approximately 0.75 to classify the bone structure as degraded or normal according to its TBS range.**

***Keywords*—Osteoporosis; Dual Energy X-ray Absorptiometry (DXA); Trabecular Bone Score (TBS); Classification; Convolutional Neural Network (CNN)**

I. Introduction

Osteoporosis is a skeletal disease characterised by low bone mineral density (BMD) and deterioration of bone structure and strength, which increases the risk of fracture and mortality. Osteoporosis is most prevalent in postmenopausal women, although the prevalence of this disease is growing in both genders due to longer life expectancy [1].

BMD measurement calculated using Dual Energy X-ray Absorptiometry (DXA), is the gold standard test for diagnosing osteoporosis. This modality uses low-intensity beams to measure BMD, so the radiation dose is much lower than conventional X-ray, but produces low-quality images that are not useful for diagnosis based on a physician's visual examination. An assessment of bone structure is recommended as more than half of all fragility fractures occur despite normal BMD values [2], [3].

Regarding bone structure analysis, 3D medical imaging modalities, such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) provide a three-dimensional, high-resolution view. Although CT is the best option, the cost and radiation dose are very high, motivating the exploration of the use of DXA images [4].

Another method of estimating bone structure is the Trabecular Bone Score (TBS), which is an assessment of bone microarchitecture obtained by texture analysis of DXA images [5]. A lower TBS is associated with an increased likelihood of fragility fractures, independently of the BMD value [6]. TBS ranges have been established, with a TBS $\geq$ 1.350 considered healthy, a TBS between 1.200 and 1.350 considered partially degraded, and a TBS $\leq$ 1.200 defining a degraded microarchitecture [7].

In recent years, there has been a notable expansion in the utilisation of neural networks and artificial intelligence in the field of medicine, particularly in the context of medical imaging. These technologies have increased the capacity of medical professionals to be supported, offering enhanced accuracy and information about various medical conditions, including osteoporosis [8].

In particular, convolutional neural networks (CNNs) have demonstrated remarkable capabilities in the identification of bone lesions, the estimation of BMD and the prediction of fracture risk in various medical imaging modalities. Such systems not only provide clinicians with valuable information for the early detection and management of disease but also improve treatments and the quality of life for patients.

The primary challenge in utilising CNNs is the necessity for a substantial quantity of labelled data for the training of models. However, in the field of medicine, particularly in the context of osteoporosis, the availability of a large dataset is often limited. Transfer learning is a technique whereby knowledge acquired during pre-training is transferred to a new task. By fine-tuning pre-trained CNNs on osteoporosis-specific datasets, the learned features can be leveraged to improve model performance with limited data, thereby enhancing the generalisation and robustness of the solution [9]. In this field, pre-trained CNN architectures such as VGG, AlexNet, SqueezeNet, ResNet and DenseNet have gained prominence due to their versatility and effectiveness. These models have been trained on large-scale image datasets through extensive training, demonstrating their suitability for this domain [10].

The main contribution of the proposed work is the development of ad hoc architectures and the fine-tuning of pre-trained CNN models for the detection of Degraded bone architecture in DXA images. To the best of our knowledge, this is the first work to classify DXA samples according to bone structure quality based on TBS values using CNNs.

This article includes a review of related works which is presented in Section II, followed by a description of the used image dataset, the pre-processing and augmentation techniques, and the proposed CNN models for the classification in Section III. Next, Sections IV and V describe and analyse the results obtained after training and testing each model. Finally, Section VI presents the conclusions and future works.

II. Related Works

The field of research into the detection and diagnosis of osteoporosis using artificial intelligence techniques has been a significant area of study in recent years, with numerous works addressing it from different perspectives. Below, some CNN-based articles that classify images from different image modalities into categories related to osteoporosis assessment, are presented[1].

Among the published works on the classification of osteoporosis using dental panoramic images, [11] employs ResNet and EfficientNet CNN models, which are trained exclusively on images, and are assembled with clinical variables (accuracy 0.845). In contrast, [12] applies transfer learning to AlexNet, VGG16 and GoogLeNet models, resulting in accuracy values of 0.74-0.79.

Some studies utilise knee X-ray images, for example [13] present a classifier based on a CNN with multiple blocks and skip connections, obtaining an accuracy of 0.826 for classification into normal or osteoporotic categories. Furthermore, [14] reports an accuracy of 0.911 using transfer learning of pre-trained CNNs, including AlexNet, VGG16, VGG19 and ResNet also for classification into diagnostic categories.

Concerning hip and lumbar spine X-ray image analysis, [15] focuses on the prediction of osteoporosis by implementing a segmentation using a U-Net architecture and classification using DenseNet121. This approach achieves an accuracy of 0.74. Besides, [16] proposes a six-layer CNN architecture, which achieved sensitivity values of 0.853. Furthermore, [17] addresses the identification of fractures, the prediction of BMD and the assessment of fracture risk in X-ray images of the spine and hip. The study utilises pre-trained CNNs and achieves accuracy results of 0.862, 0.95 and 0.90, respectively. Although the aforementioned works do not employ DXA images, the CNN architectural solutions and fine-tuning have in some way inspired the solution proposed in this work.

Regarding studies carried out with DXA images, [18] proposes a CNN architecture to classify images according to their BMD value, achieving an accuracy of 0.98. [19] attempts to predict fracture risk, and detect scoliosis, and abnormalities, with an accuracy of 0.52, 0.94, and 0.82, respectively. Conversely, the objective of [20] is to distinguish images of healthy bones from those with osteoporosis, achieving a training accuracy of 0.90. Although [18], [19] and [20] use CNN-based solutions and DXA image modalities, they do not perform a classification using TBS as a label.

Conversely, the work [21] predicts BMD and TBS values through the architecture of ResNet50 CNNs using CT images. The results demonstrate that the obtained BMD values exhibit a strong correlation, whereas the obtained TBS values exhibit a moderate correlation.

Considering that to our knowledge there are no works that perform the classification of DXA images using solutions based on CNN and TBS as a label, the presented works were not used in order to compare the results with ours, but have served as inspiration for the developed strategy.

III. Materials and Methods

The following subsections present the dataset, the pre-processing and the resampling techniques used in this work. Finally, three neural network approaches are presented. The first two are based on pre-trained CNN models, while the last one is based on CNN models but with simple ad hoc architectures.

A. Dataset

We performed a retrospective study of 1469 patients[2], with a mean age of 64.61 $\pm$ 10.7 (standard deviation) years, using spine DXA images produced using a General Electric Lunar Prodigy Advance® equipment.

The study set comprises one raw image per patient (from the total of 1469 images), exported from enCORE v17 software platform, of which 1098 were labelled as Normal and 371 as Degraded. This unbalance is expected according to the disease incidence [22]. The classification was based primarily on the range of TBS values calculated using TBS iNsight version 3.0.2.0. The ranges were presented in Section I. No images corresponding to severe scoliosis cases, prosthetics or other conditions that affected the ROI segmentation performed by the software were included. The main reason behind this is that the calculated TBS under such circumstances is unreliable.

The spatial resolution of the images is approximately 300 x 280 pixels and is represented in 8-bit greyscale. The region of interest is defined as the lumbar vertebrae L1 to L4, as defined by the same software used to export the images.

Because not all images have the same dimensions, the images were resized to 224x224 pixels. The lumbar spine area, including the L1-L4 vertebrae, is positioned in the centre of the crop to eliminate part of the background and other vertebrae, thus ensuring that the relevant information (L1-L4) is visible (Fig. 1).

The dataset was randomly divided into training, validation and test sets. A total of 181 samples were reserved for testing the model, of which less than 25% belong to the Degraded class. The remaining samples were used to train and validate the model.

[1]The accuracy is the value shown due is the metric used in most papers.

[2]The data was obtained from "Instituto de Diagnóstico e Investigaciones Metabólicas" (IDIM), Buenos Aires, Argentina

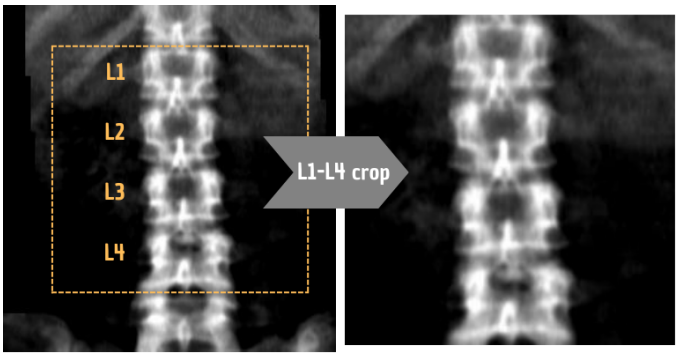

Fig. 1. Example of an image resized to 224x224 pixels by cropping the L1-L4 zones.

B. Data Augmentation

Data augmentation techniques were applied to the available training and validation image datasets to increase the number of samples, thereby reducing the difference between classes. Training the models with the original dataset can lead to underfitting and poor generalisation due to the complexity of the model and the limited number of images.

The augmentation process was made by applying by horizontal flipping, random rotations from -10° to 10°, and random darkness and brightness (10% to 40%). While there are many possible transformations, we select and apply those that do not change the relative relationship between pixel properties or those that do not alter the texture of the images. This is because the texture is one of the most used indicators for analysing bone structure on 2D images, and any change alters the sample and therefore the final result [23], [7].

As can be seen, no pronounced rotations or vertical flips were performed, since according to the patient positioning guides [24], there are no cases where this is possible. Furthermore, no zoom operations have been applied, as the areas of interest (L1-L4) may be lost. Fig. 2 shows an example of the result of applying some filters on one of the samples.

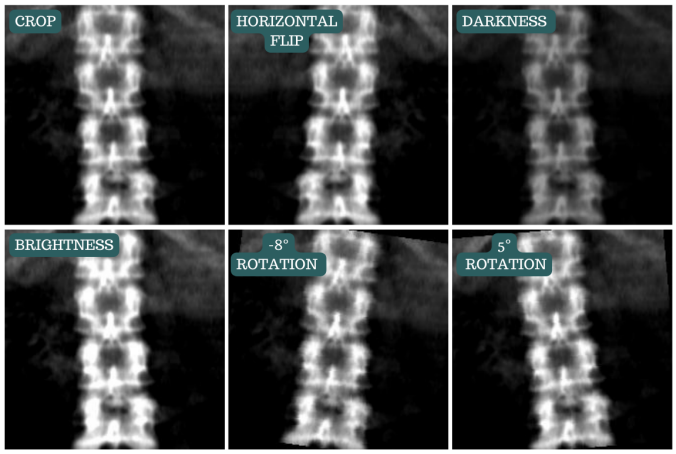

Fig. 2. Example of applied filters to data augmentation.

As previously mentioned, the dataset was divided into three distinct subsets: training, validation, and testing. The data augmentation techniques presented above were applied only to the training and validation sets.

The number of samples belonging to the Degraded class was increased to a greater extent than that of the Normal class. This was done by applying more filters to improve the difference in the number of samples between the two classes. After augmentation, 37% of the samples in the training and validation datasets belong to the Degraded class. Table I shows the number of original and augmented samples belonging to each class, the partitioning of the data into training (Train), validation (Val) and test, the partial sums of the Normal and Degraded samples (PS1), and the partial sums of the partitions (PS2). In addition, it can also be seen that no augmentation was applied to the test set.

TABLE I. NUMBER OF ORIGINAL AND AUGMENTED SAMPLES FOR EACH CLASS

	Normal original samples	**Degraded** original samples	**Normal** augmented samples	**Degraded** augmented samples	**PS1**
Train	979	148	225	563	1915
Val	140	21	32	80	273
Test	139	42	0	0	181
PS2	1258	211	257	643	2369

C. Approaches 1 and 2: Pre-trained CNN Models

Several pre-trained models have shown good performance in radiological image classification tasks such as [14], [11]. For this work, five pre-trained models were selected: AlexNet [25], ResNet-18 [26], VGG-16 [27], DenseNet-121 [28] and SqueezeNet [29].

Since these models were originally designated to classify the ImageNet dataset, which contains 1000 classes, it was necessary to modify the final fully connected (FC) layer to output two classes and incorporate a softmax function to generate probabilities for each class (Normal and Degraded). Additionally, a dropout layer of 0.4 rate value was added for regularisation purposes. These models were pre-trained on the ImageNet dataset, meaning their weights were initialised rather than randomly assigned.

These pre-trained models were retrained using two approaches. The first one involved retraining all layers to update the entire set of network weights. The second approach, known as fine-tuning, entailed retraining only the reshaped layers.

Each model has its own unique architecture, with different combinations of layers, resulting in varying depths and number of parameters. The number of trainable parameters for the models with approach 1 varies from 7.364×10^5 in the case of SqueezeNet to 1.343×10^8 in the case of VGG-16. In contrast, when training the models with fine-tuning of the second approach, the trainable parameters vary from 1.026×10^3 in the case of ResNet-18 to 8.194×10^3 in the case of VGG-16 and AlexNet.

All models were retrained using a batch size of 24, the stochastic gradient descent (SGD) optimiser with a learning rate (LR) of 0.0001, and a momentum of 0.9. The retraining of all layers of the models (approach 1) was carried out for

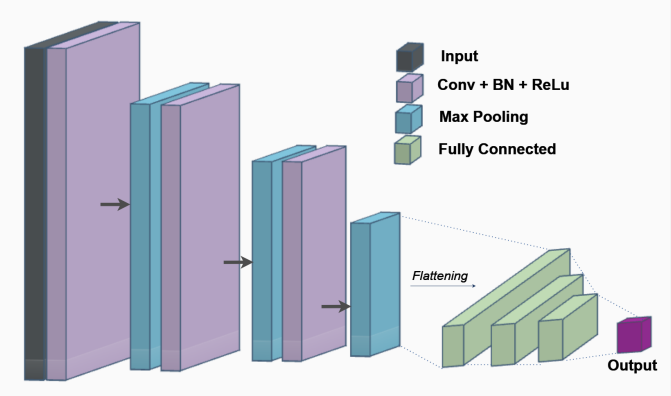

Fig. 3. Representation of the basis of the proposed models. The model comprises convolutional blocks which are followed by fully connected layers.

50 epochs, while the retraining with fine-tuning (approach 2) was carried out for 20 epochs.

D. Approach 3: Proposed CNN Model

In addition to the pre-trained models detailed above, four distinct architectures of CNNs were proposed. The four models presented below were based on the architecture shown in Fig. 3, with varying numbers of blocks. The blocks are composed of a convolutional layer (Conv), batch normalisation (BN) and a rectified linear unit (ReLu) activation function, followed by a max pooling layer to reduce the dimensions. After the flattening operation, fully connected layers are added with varying numbers of neurons. Finally, a single output neuron is added where a sigmoid function is applied to obtain a binary classification. The hyperparameter tuning for all models was performed using Ray Tune library algorithms. The selected hyperparameters are shown in the respective model description.

The architectures and hyperparameters of each model are described in Table II. This table presents the number of convolutional blocks along with the number of filters in each block, as well as the number of FC layers and the number of neurons in each. Additionally, it details the batch size used to train each model. All models were trained using the SGD optimizer, and the table specifies the learning rate (LR) and momentum values used for each model. The kernel size for all convolutional layers in all models is 3x3, with a stride of 1. A flattening layer was also added before the FC layers. All models were trained for a maximum of 200 epochs, with early stopping triggered if the loss did not improve after 15 consecutive epochs.

IV. Results

This section presents the results of testing all the models after training. Considering the positive class as Degraded and the negative as Normal, the following metrics were calculated for each model: True Positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN), Accuracy, Sensitivity, Specificity, F1-Score and Area Under the Receiving Operator-Curve (AUC) [30]. Furthermore, the time required in seconds to train for 10 epochs is given for each model.

Taking into account the imbalance of the dataset, mentioned above, we choose F1-Score as the metric that better shows how well the models correctly predict the Degraded class considering at the same time the prediction of the Normal class.

Table III shows the results of testing the pre-trained CNN models after retraining the whole layers (approach 1). It can be seen that all models show high values for F1-Score, being SqueezeNet the model that reached the maximum value of **0.759**.

The results obtained testing the pre-trained CNN models after retraining only the reshaped layers (approach 2) are shown in Table IV. It can be observed that the F1-Score values are lower than those achieved from models trained with approach 1. Furthermore, the training time drops considerably due to the smaller number of trainable parameters. In this instance, the DenseNet-121 model exhibited the most optimal performance in terms of the F1-Score, achieving a value of **0.658**.

Finally, Table V shows the results of testing the proposed architectures (approach 3). It can be seen that in general good metric values were obtained, with Model 4 exhibiting the highest performance, achieving an F1-Score of **0.747**.

The results of the best models (in terms of F1-Score greater than 0.7) are resumed in Fig. 4, which shows the F1-Score along with Sensitivity and training time. The best models are the ones that are located in the upper-right corner. It can be seen that the best is SqueezeNet followed by Model 4.

All the results were obtained after training and testing all the models using Python (version 3.7.16) language and PyTorch (version 1.13.1) framework running on an Ubuntu 20.04.6 LTS computer with GTX 1070 GPU and Intel(R) Core i7-7700K × 4.20 GHz CPU, under comparable Operating System load conditions.

V. Discussion

Observing the results, it can be seen at first sight that pre-trained CNN models that were re-trained with approach 1 have a very good performance in terms of F1-Score while they have high sensitivity values. Between these models, SqueezeNet has the best F1-Score and the time needed for train 10 epochs was relatively low in comparison with the times achieved by VGG16 or DenseNet-121.

When the retraining is done only in the reshaped CNNs (approach 2), DenseNet-121 had the best F1-Score, but it can be seen that the general performance, in terms of F1-Score, is poorer compared with all the models of approach 1, with a similar training time to that obtained retraining SqueezeNet entirely.

Among the simpler architectures proposed, model 4 achieved the best performance. This model could stand out due to its architecture and the selected hyperparameters. A slightly higher number of filters in the last layers could have helped to capture important texture features. In addition, having a larger number of neurons in the FC layer with a simpler structure could reduce complexity and the risk of overfitting. Consequently, the model also has the largest number of trainable parameters, which means it has the ability to learn more

TABLE II. SUMMARY OF AD-HOC ARCHITECTURE DETAILS AND HYPERPARAMETERS FOR EACH PROPOSED MODEL (APPROACH 3)

	Model 1	Model 2	Model 3	Model 4
Convolutional blocks	5	3	3	3
Filters per block	8, 16, 32, 64, 128	8, 16, 32	8, 16, 32	8, 28, 36
FC layers	3	3	2	2
Neurons per layer	1000, 100, 1	500, 100, 1	100, 1	1000,1
Batch size	24	24	24	48
LR	0.001	0.0001	0.001	0.0001
Momentum	0.6	0.7	0.9	0.9
Trainable parameters	3.400×10^6	1.087×10^7	2.169×10^6	2.434×10^7

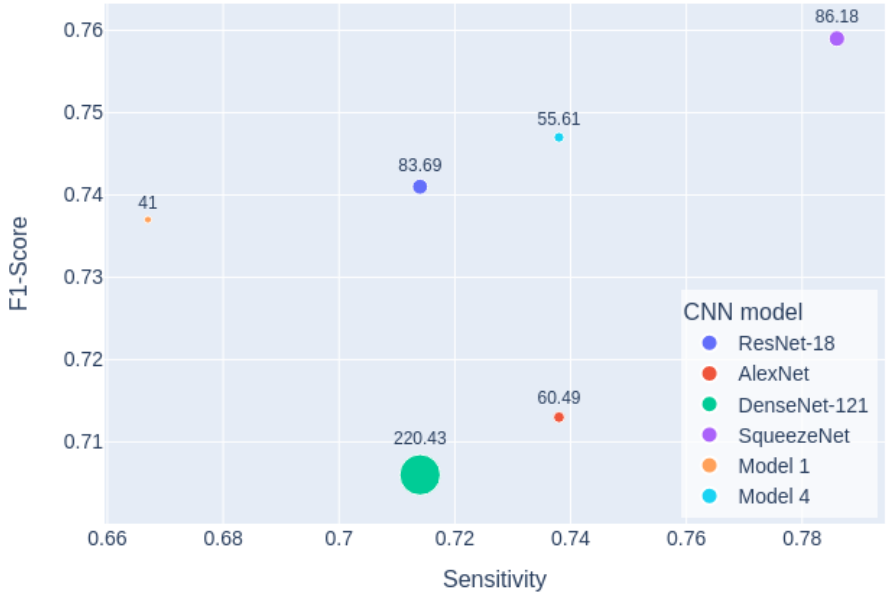

Fig. 4. Classification results of models that achieved an F1-score value greater than 0.7. The X and Y axes represent sensitivity and F1-Score values, respectively, while the diameter of the circles represents the training time of 10 epochs.

TABLE III. RESULTS OF TESTING PRE-TRAINED MODELS AFTER RETRAINING THE WHOLE LAYERS (APPROACH 1)

	TP	FP	TN	FN	Accuracy	Sensitivity	Specificity	F1-Score	AUC	Time(s) 10 epochs
ResNet-18	30	9	130	12	0.884	0.714	0.935	0.741	0.816	83.69
AlexNet	31	14	125	11	0.862	0.738	0.899	0.713	0.815	60.49
VGG-16	29	17	122	13	0.834	0.690	0.878	0.659	0.784	400.03
DenseNet-121	30	13	126	12	0.862	0.714	0.906	0.706	0.810	220.43
SqueezeNet	33	12	127	9	0.884	0.786	0.914	***0.759***	0.828	86.18

representations and features from the data. Finally, a larger batch size, together with a lower learning rate and higher momentum, could cause more stable and gradual learning, facilitating convergence to optimal values for the weights.

Comparing the different strategies, it can be seen that ad hoc CNN models performed around 14% better than the retrained models with approach 2, and the time is about 40% lower comparing the best results of both approaches.

Finally, the proposed ad hoc CNNs performed almost the same as pre-trained models trained with approach 1 and reduced the time needed for training by around 35%. Furthermore, it should be noted that the proposed ad hoc models were trained entirely with the images from the dataset proposed in this work, while the architectures of approaches 1 and 2 have a complex pre-training with a large dataset such as ImageNet.

TABLE IV. RESULTS OF TESTING PRE-TRAINED MODELS AFTER RETRAINING ONLY THE RESHAPED LAYERS (APPROACH 2)

	TP	FP	TN	FN	Accuracy	Sensitivity	Specificity	F1-Score	AUC	Time(s) 10 epochs
ResNet-18	22	10	129	20	0.834	0.524	0.928	0.594	0.726	39.25
AlexNet	25	14	125	17	0.829	0.595	0.899	0.617	0.747	28.14
VGG-16	19	6	133	23	0.840	0.452	0.957	0.567	0.705	148.2
DenseNet-121	25	9	130	17	0.856	0.595	0.935	***0.658***	0.765	93.59
SqueezeNet	21	2	137	21	0.873	0.500	0.986	0.646	0.743	44.73

TABLE V. RESULTS OF TESTING AD HOC ARCHITECTURE MODELS (APPROACH 3)

	TP	FP	TN	FN	Accuracy	Sensitivity	Specificity	F1-Score	AUC	Time(s) 10 epochs
Model 1	28	6	133	14	0.889	0.667	0.957	0.737	0.934	41.00
Model 2	29	14	125	13	0.850	0.690	0.899	0.682	0.888	44.62
Model 3	23	7	132	19	0.856	0.548	0.950	0.639	0.828	36.80
Model 4	31	10	129	11	0.884	0.738	0.928	***0.747***	0.936	55.61

VI. CONCLUSIONS AND FUTURE WORKS

Several CNN-based strategies to distinguish between Normal and Degraded trabecular bone structure from DXA images have been presented.

Among the different approaches proposed, simpler architecture CNNs were more adequate in comparison with pre-trained CNN models, since they reached the same performance as the best pre-trained CNN, requiring considerably less training effort and hence less computational resources. This is consistent with the fact that in previous works, algorithms for basic texture patterns search worked better than those related to complex textures [31], [32].

Considering future works, at first, we plan to develop an automatic method to segment vertebrae and compare the effectiveness of this segmentation with the one made by the software, which in some cases requires manual adjustments. Also, individual (L1-L4) vertebrae training is planned in order to assess the trabecular bone quality for each vertebra. As part of the ongoing project in which this work was done, more classes will be included, besides Degraded and Normal, to have more detailed information about the intermediate conditions of bone quality degradation. At last, regarding data augmentation, we plan to study the rotational invariance of the texture in order to increase the rotation range.

ACKNOWLEDGMENT

This work was supported by a Fellowship of Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and Escuela Doctoral de la Universidad de Jaén (EDUJA).

REFERENCES

[1] I. Foessl, H. P. Dimai, and B. Obermayer-Pietsch, "Long-term and sequential treatment for osteoporosis," *Nature Reviews Endocrinology*, vol. 19, no. 9, pp. 520–533, 2023.

[2] R. Karunanithi, S. Ganesan, T. Panicker, M. P. Korath, and K. Jagadeesan, "Assessment of bone mineral density by dxa and the trabecular microarchitecture of the calcaneum by texture analysis in pre-and postmenopausal women in the evaluation of osteoporosis," *Journal of Medical Physics*, vol. 32, no. 4, pp. 161–168, 2007.

[3] J. Compston, A. Cooper, C. Cooper, N. Gittoes, C. Gregson, N. Harvey, S. Hope, J. A. Kanis, E. V. McCloskey, K. E. Poole *et al.*, "Uk clinical guideline for the prevention and treatment of osteoporosis," *Archives of osteoporosis*, vol. 12, pp. 1–24, 2017.

[4] J. E. Adams, "Advances in bone imaging for osteoporosis," *Nature Reviews Endocrinology*, vol. 9, no. 1, pp. 28–42, 2013.

[5] E. Shevroja, J.-Y. Reginster, O. Lamy, N. Al-Daghri, M. Chandran, A.-L. Demoux-Baiada, L. Kohlmeier, M.-P. Lecart, D. Messina, B. M. Camargos *et al.*, "Update on the clinical use of trabecular bone score (tbs) in the management of osteoporosis: results of an expert group meeting organized by the european society for clinical and economic aspects of osteoporosis, osteoarthritis and musculoskeletal diseases (esceo), and the international osteoporosis foundation (iof) under the auspices of who collaborating center for epidemiology of musculoskeletal health and aging," *Osteoporosis International*, vol. 34, no. 9, pp. 1501–1529, 2023.

[6] L. Pothuaud, P. Carceller, and D. Hans, "Correlations between grey-level variations in 2d projection images (tbs) and 3d microarchitecture: applications in the study of human trabecular bone microarchitecture," *Bone*, vol. 42, no. 4, pp. 775–787, 2008.

[7] B. C. Silva, W. D. Leslie, H. Resch, O. Lamy, O. Lesnyak, N. Binkley, E. V. McCloskey, J. A. Kanis, and J. P. Bilezikian, "Trabecular bone score: a noninvasive analytical method based upon the dxa image," *Journal of Bone and Mineral Research*, vol. 29, no. 3, pp. 518–530, 2014.

[8] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," *Frontiers in Public Health*, vol. 11, p. 1273253, 2023.

[9] A. W. Salehi, S. Khan, G. Gupta, B. I. Alabduallah, A. Almjally, H. Alsolai, T. Siddiqui, and A. Mellit, "A study of cnn and transfer learning in medical imaging: Advantages, challenges, future scope," *Sustainability*, vol. 15, no. 7, p. 5930, 2023.

[10] P. Kora, C. P. Ooi, O. Faust, U. Raghavendra, A. Gudigar, W. Y. Chan, K. Meenakshi, K. Swaraja, P. Plawiak, and U. R. Acharya, "Transfer learning techniques for medical image analysis: A review," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 79–107, 2022.

[11] S. Sukegawa, A. Fujimura, A. Taguchi, N. Yamamoto, A. Kitamura, R. Goto, K. Nakano, K. Takabatake, H. Kawai, H. Nagatsuka *et al.*, "Identification of osteoporosis using ensemble deep learning model with panoramic radiographs and clinical covariates," *Scientific reports*, vol. 12, no. 1, p. 6088, 2022.

[12] T. Nakamoto, A. Taguchi, and N. Kakimoto, "Osteoporosis screening support system from panoramic radiographs using deep learning by convolutional neural network," *Dentomaxillofacial Radiology*, vol. 51, no. 6, p. 20220135, 2022.

[13] A. Kumar, R. C. Joshi, M. K. Dutta, R. Burget, and V. Myska, "Osteonet: A robust deep learning-based diagnosis of osteoporosis using x-ray images," in *2022 45th International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, 2022, pp. 91–95.

[14] I. M. Wani and S. Arora, "Osteoporosis diagnosis in knee x-rays by transfer learning based on convolution neural network," *Multimedia Tools and Applications*, vol. 82, no. 9, pp. 14 193–14 217, 2023.

[15] S.-W. Feng, S.-Y. Lin, Y.-H. Chiang, M.-H. Lu, and Y.-H. Chao, "Deep learning-based hip x-ray image analysis for predicting osteoporosis," *Applied Sciences*, vol. 14, no. 1, p. 133, 2023.

[16] B. Zhang, K. Yu, Z. Ning, K. Wang, Y. Dong, X. Liu, S. Liu, J. Wang, C. Zhu, Q. Yu *et al.*, "Deep learning of lumbar spine x-ray for osteopenia and osteoporosis screening: A multicenter retrospective cohort study," *Bone*, vol. 140, p. 115561, 2020.

[17] C.-I. Hsieh, K. Zheng, C. Lin, L. Mei, L. Lu, W. Li, F.-P. Chen, Y. Wang, X. Zhou, F. Wang *et al.*, "Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning," *Nature communications*, vol. 12, no. 1, p. 5472, 2021.

[18] A. Z. Mohammed and L. E. George, "Osteoporosis detection using convolutional neural network based on dual-energy x-ray absorptiometry images," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 1, pp. 315–321, 2023.

[19] T. Nissinen, S. Suoranta, T. Saavalainen, R. Sund, O. Hurskainen, T. Rikkonen, H. Kröger, T. Lähivaara, and S. P. Väänänen, "Detecting pathological features and predicting fracture risk from dual-energy x-ray absorptiometry images using deep learning," *Bone reports*, vol. 14, p. 101070, 2021.

[20] N. K. Kirilov and E. K. Kirilova, "Classifying dual-energy x-ray absorptiometry images using machine learning," in *2021 56th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)*. IEEE, 2021, pp. 53–54.

[21] K. Yoshida, Y. Tanabe, H. Nishiyama, T. Matsuda, H. Toritani, T. Kitamura, S. Sakai, K. Watamori, M. Takao, E. Kimura *et al.*, "Feasibility of bone mineral density and bone microarchitecture assessment using deep learning with a convolutional neural network," *Journal of Computer Assisted Tomography*, vol. 47, no. 3, pp. 467–474, 2023.

[22] N. Salari, H. Ghasemi, L. Mohammadi, M. H. Behzadi, E. Rabieenia, S. Shohaimi, and M. Mohammadi, "The global prevalence of osteoporosis in the world: a comprehensive systematic review and meta-analysis," *Journal of orthopaedic surgery and research*, vol. 16, pp. 1–20, 2021.

[23] K. Zheng and S. Makrogiannis, "Bone texture characterization for osteoporosis diagnosis using digital radiography," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 1034–1037.

[24] A. Bazzocchi, F. Ponti, U. Albisinni, G. Battista, and G. Guglielmi, "Dxa: Technical aspects and application," *European journal of radiology*, vol. 85, no. 8, pp. 1481–1492, 2016.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[28] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[29] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[30] Ž. Vujović *et al.*, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021.

[31] R. Abdala, M. Gonzalez, M. B. Zanchetta, V. Longobardi, M. Sesta, and J. M. Massa, "Classification of dxa samples into tbs ranges using machine learning and pyradiomics," in *Proceedings of the n-th edition of some conference*. ASBMR 2022, Annual Meeting (2022), 2022. [Online]. Available: https://www.endoweb.net/images/pdfs/ ClassificationofDXASamplesIntoTBSRangesUsingMachineLearningandPyradiomics.pdf

[32] M. Gonzalez, J. M. Massa, and N. de Martino, "Bone quality classification in dxa images using pyradiomics and machine learning," in *17th International Symposium on Medical Information Processing and Analysis*, vol. 12088. SPIE, 2021, pp. 396–402.