

Age prediction from low resolution, dual-energy X-ray images using convolutional neural networks

Kamil Janczyk ^{1,2,*}, Tomasz Neumann ¹, Jacek Rumiński, ^{1,‡} Natalia Głowacka ¹ and Piotr Wiśniewski ³

¹ Dept. of Biomedical Engineering, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology;

² Medical Engineering Division, Maritime Advanced Research Centre;

³ Medical University of Gdansk;

* Correspondence: kamil.janczyk@pg.edu.pl

‡ These authors contributed equally to this work.

Abstract: Age prediction from X-rays is an interesting research topic important for clinical applications such as biological maturity assessment. It is also useful in many other practical applications, including sports or forensic investigations for age verification purposes. Research on these issues is usually carried out using high-resolution X-ray scans of parts of the body, such as images of the hands or images of the chest. In this study, we used low-resolution, dual-energy, full-body X-ray absorptiometry images to train deep learning models to predict age. In particular, we proposed a pre-processing framework and adapted many partially pre-trained convolutional neural network (CNN) models to predict the age of children and young adults. We used a new data set of 910 multispectral images weakly annotated by specialists. The experimental results show that the proposed pre-processing techniques and the adapted approach to the CNN model achieved a discrepancy between chronological age and the predicted age of around 15.56 months for low-resolution whole-body X-rays. Furthermore, we found that the main factor that influenced age prediction scores was spatial features, not multispectral features.

Keywords: Dual-energy absorption; convolutional neural network; bone age; age prediction

1. Introduction

Dual-energy X-ray absorptiometry (DXA) is the most widely used clinical tool to diagnose osteoporosis. It permits a precise in vivo quantification of mineralization in bones surrounded by soft tissue. DXA utilizes two different x-ray energies in order to separate bone from soft tissue. The lower-energy beam is preferentially attenuated by bone; however, the contributions of soft tissue to beam attenuation can be mathematically derived. Radiation exposure to the patient is very small, usually of a magnitude similar to the daily background radiation. Older generation DXA scanners provided low-resolution images used solely for patient positioning verification or regions of interest (ROI) placement. The obtaining of images of much better quality from modern DXA scanners leads to the development of many new diagnostic applications, such as the trabecular bone score or vertebral morphometry [1].

Special DXA protocols permit quantification of bone mineralization, fat mass, and fat distribution. Improving the image quality and shortening the scanning time allowed performing a full-body scan to quantify its components. Contemporary DXA images of the full-body, containing information about the quantity and spatial distribution of body components, open the field for research on the new use of this old method.

This work is a part of the larger project that addresses the following research question: How deep learning methods could improve the diagnostic knowledge extraction from whole-body, low resolution, dual-energy X-ray absorptiometry data, potentially related with a lower radiation dose for patients. Related diagnostic knowledge could be represented in many forms. For example, it could be organised as a set of parameters including: bone density (e.g. bone mineral density (BMD), bone mineral content (BMC)), fat-density, muscle-density, osteoporosis scores, bone age, etc. The scanners used for testing allow to



Citation: Janczyk, K.; Neumann, T.; Rumiński, J.; Głowacka N. and Wiśniewski P. Age prediction from low resolution, dual-energy X-ray images using convolutional neural network. *Preprints* 2022, 1, 0. <https://doi.org/>

Received:

Accepted:

Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

obtain a digital version of this image, but it is only used to check correct positioning of the patient setting and correct placement of the regions of interest (ROI). The bone density results are computed by specialized software based on the X-ray attenuation data that the detector will receive. For BMD measurement, it is important to eliminate soft tissue input, a measurement based on bone alone. This is possible by scanning at two different x-ray photon energies, followed by mathematical manipulation of the recorded signal - different attenuations for bones and soft tissues at different photon energies are used. Furthermore, the radiation dose for DXA is considerably lower than the radiation dose for conventional radiography. Many other information and dependencies can be extracted from such images that are not used in determining BMD. Many researchers focused on extracting information from DXA and X-Ray images.

In [3] the authors present the study that describes the comparison of the use of DXA images together with X-ray images to assess bone age in children using Bland and Altman statistical methods. A high agreement was obtained between the bone age assessments performed by DXA and the X-ray. Other studies also considered the problem of using statistical methods, this time involving only BMD measurements. Castillo et al. in [4] investigated whether there are correlations between age and gender and BMD, and whether you can use this solution in forensic anthropology. The research was carried out on 70 subjects and shows bone mineral density to be a useful technique for gender and age data in forensic anthropology. Neural networks can also be used to analyze X-ray images. Navega et al. in [5] used neural networks to create a method to estimate age at death. The data sample consisted of 100 femora of female individuals. The authors used a modified general regression neural network as a model for age prediction. Using a nonprobabilistic artificial neural network, the mean absolute difference achieved between the real and estimated age ranged from 9.19 to 12.03 years, depending on the variables used in the modelling. The use of classic X-ray images is also applicable. Lee et al. in [6] proposed using hand x-ray images to predict bone age. The images were pre-processed (region of interest (ROI) segmentation, standardization and preprocessing of input radiographs) and analysed by the neural network - GoogLeNet. The solution was validated using 8325 data samples. The accuracy achieved by the classification model was: within 2 years (<98%) and within 1 year (<90%) for a group of women and men. A similar solution was presented by Igloukov et al. in [7]. Bone age prediction was performed using a database provided by the Radiological Society of North America (RSNA). Initially, the images were processed: hand masks were extracted and all other objects were removed (positive mining utilizing U-Net architecture), as well as key points were detected, and the hands were placed on the images in one orientation and size (network inspired by the VGG model). Bone age assessment was performed using two types of convolutional neural network (CNN), regression and classification, as well as image division into three areas (whole hand, carpal bones, metacarpals and proximal phalanges). The smallest mean absolute error (MAE) was obtained for the ensemble of classification and regression models and the ensemble of regional models. This solution had the best accuracy with MAE equal to 6.10 months. Human age can also be estimated on the basis of chest images using CNN. In solution [8], Karargyris et al. created such a solution based on the DenseNet-169 model, pre-trained on ImageNet. They used the Chest X-ray dataset from the National Institutes of Health, which includes 10,000 chest images from 30,000 different people. They obtained the results: for a margin of ± 4 years, a sensitivity equal to 0.6745 and for ± 9 years - 0.9441. Similarly, Xue et al. [9] used X-rays to classify gender. The dataset contained 2,066 chest X-ray images (1097 women and 969 men). The images were pre-processed and then the features were extracted from them using the neural networks: AlexNet, VggNet, GoogLeNet and ResNet (for comparison). The classification was carried out using the Support Vector Machine (SVM) and Random Forest. The best result was obtained for the VGGNet-16 + SVM classifier, accuracy was 86.6% with 5 times cross-validation. Spine X-ray images can also be used to classify gender. Xue et al. in [10] developed a sequential CNN model, trained from scratch using spine images. The accuracy of 88.9% was obtained for the original spine images. Data analysis using



the DenseNet-121 model pre-overtrained in the ImageNet data range gave an accuracy of 99% for cervical images and 98% for lumbar images. In [11] Marouf et al. proposed a hybrid methodology for gender classification and bone age estimation, using the trained VGG-16 model and the RSNA dataset. They have achieved an accuracy of 99% for the gender classification and for age classification MAD 0.50 years and RMS 0.67 years. In [12] Mumtaz et al. proposed to examine method based on CNN and left-hand radiographs of the children to determine the gender. Using class activation mapping (CAM) they showed that the lower part of the palm around the wrists (wrist) is more important to determine the gender of the child. The model accuracy was 98%, given the unfinished skeleton of the children. Age can also be determined from whole-body CT scans. Nguyen et al. in [13] they presented a solution based on deep hierarchical features. They used data from an anonymous hospital that contained 813 whole-body CT bone images. They tested the fine-tuned VGGNet, ResNet and GoogLeNet models and modified model which is based on VGGNet for automatic BAA. The best result was achieved for modified VGGNet with hierarchical features - MAE: 4.856 months. Castillo et al. proposed model for the bone age prediction base of the VGG16 with attention mapping-focused architecture. Using the RSNA dataset [14] they achieved MAE equal to 11.45 months for both genders. Liu et al. evaluated a proposed two-stage bone age assessment network with ranking learning using the same dataset. They achieved an average MAE of 6.05 months [15]. Salim and Hamza introduced a two-stage approach for bone age assessment using segmentation and ridge regression [16]. The MAE of the proposed solution was equal to 6.38 months.

Here, we focus on age prediction for children and young adults from dual-energy X-ray images. It can be important in many applications, including proper assessment of the child's development. The agreement between chronological age and developmental advancement is an important indicator of the health condition of children. However, under pathological conditions, age predictions based on other methods (e.g. bone age) may better reflect biological maturity [2]. In this work, our contributions are as follows:

- We build a dataset of low-resolution, dual-energy X-ray absorptiometry images with annotations done by specialists. This is probably the one of the first dataset with whole-body, dual-energy X-ray absorptiometry images (to the best of our knowledge).
- We propose a CNN-based age prediction framework for low-resolution, whole-body, dual-energy X-ray absorptiometry images.
- As a result of experiments, we demonstrate that age can be successfully predicted from low-resolution images, so potentially even less radiation energy could be used at the data acquisition phase.

The rest of the paper is structured as follows. In Section II, the methodology used in the study is described. The results are presented in Section III and discussed in Section IV. The study is concluded in Section V.

2. Materials and Methods

2.1. Data acquisition

All data used in this study was obtained from the Hologic QDR 4500 X-ray Bone Densitometer. This model has a single switched pulse, dual-energy X-ray tube operating at 100 and 140kVp. Average tube current is equal to 5mA at 50% duty cycle. It can be used to scan the lumbar spine (in anterior-posterior and lateral projections), the proximal femur, and the whole body. Maximum scan area is 195 cm x 65 cm. Each original sample generated by this device contains high- and low-resolution data. In this study we focus on raw, low-resolution, whole-body data. Each sample contained six not preprocessed images corresponding to high- and low-energy beam passing through a rotating drum filter with three different segments with radio-opacities equivalent to tissue, bone, and air and then a patient (Figure 1). In the remaining part of the paper, we will refer to these by combining keywords of energy level and segment of drum filter (e.g. low bone image corresponds to low energy passing through segment of filter with radio-opacities equivalent to bone).



The height of low-resolution images was variable from 84 to 128 pixels. The width was constant and equaled 106 pixels.

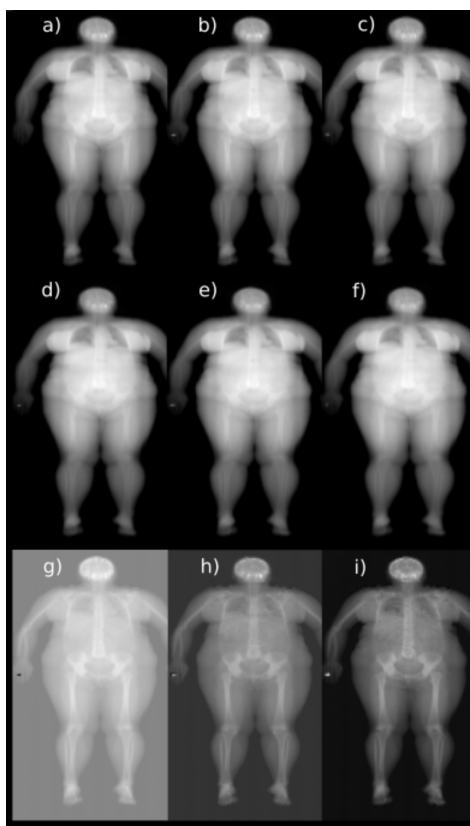


Figure 1. The first six images correspond to low- and high-energy beam passing through different segments of filter and patient: a) low air, b) low tissue, c) low bone, d) high air, high tissue, f) high bone. Images g), h) and i) are result of the subtraction of high-energy images from low-energy images.

2.2. Dataset

The dataset consists of 1,455 samples of the full body DXA scan, but according to the objective of this study we only used examples from young subjects (patients up to 230 months old at the time of the DXA scan). After reduction, the data set includes 910 samples, 570 for males and 370 for females (mean age was equal to 140.4 ± 42.1 months). This dataset will be referenced further in the text as the DXA dataset. Considering a small number of samples, the influence of data augmentation on the analysed deep learning models was tested. The data augmentation operations (selected experimentally) were: rotation of the image by an angle of 0 to 5 degrees and random horizontal flip of the image. The augmented data set will be further referenced in the text as the augDXA dataset. The age of subjects at the time of the DXA scan, which was used as labels, was determined by subtracting the birth date from the date of the DXA scan. All retrospectively used clinical and image data were collected from 2015 to 2020 at Gdansk Medical University. The use of DXA scans in the study was approved by the Independent Bioethics Committee for Scientific Research at Medical University of Gdańsk.

To obtain an appropriate representation of the data in the training and validation sets, the samples were categorised by age of the subject as follows: group 1 - age form 0 to 96 months (170 samples), group 2 - age form 97 to 132 months (215 samples), group 3 - age form 133 to 156 months (177 samples), group 4 - age form 157 to 180 months (168 samples), and group 5 - age form 181 to 230 months (184 samples) Figure 2. The samples in all groups were proportionally divided into training set (80%) and validation set (20%) in a 5-fold cross-validation experiment. Validation samples were not used during the training. The

validation set was used as a test set (during the training the validation set was used only to monitor the training process, it was not used in any decision taking e.g. early stopping, etc.).

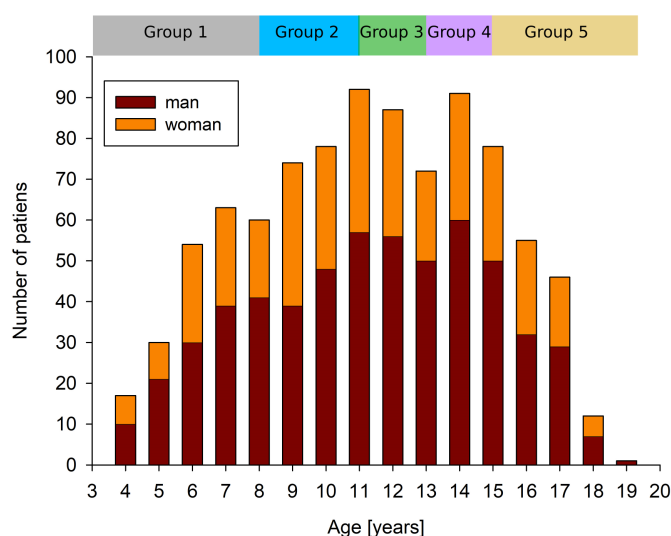


Figure 2. The data in groups with a similar number of samples were divided proportionally into training and validation datasets based on age.

2.2.1. Data preparation

Three different variants of the three-channel images were investigated. In the first case the arrays corresponding to high energy were subcontracted from arrays corresponding to low energy for each filter. In the second case, the three-channel images were assembled from the low- and high-energy arrays separately. In third variant, the images were assembled from three copies of the same arrays (e.g. three copies of low tissue array).

Next, in all cases, the obtained images were normalized globally to a value from -1 to 1 channel-wise. The normalisation range was selected on the basis of the model used. Lastly, the images were cropped symmetrically and padded to size 128x128 pixels. Each channel of a single image was padded with constant value equal to the mean of 24 pixels selected from the background image.

2.3. Tested model and training procedure

In our study we adopted different base models (MobileNetV2 [18], VGG16 [19], InceptionV3 [20], ResNet50V2 [21] and InceptionResNetV2 [22]) without top layers, loaded with 'imagenet' weights and added layers (Figure3). Added layers included: global average pooling layer (GA) or depthwise convolution layer (DW) with 4x4 or 2x2 filter (depending on base model), two dropout layers with 0.35 dropout parameter, dense layer with 128 filters and regression layer with 1 filter. The ReLU activation function was used in both convolution layers and the linear activation function was used in the depthwise convolution layer. The number of convolutional and dropout layers as well as the number of filters and dropout parameters were selected after multiple experiments. The loss function was defined as the mean square error, and the metric as the mean absolute error.

The first experiment (Experiment I) consisted of a 5-fold cross-validation using different base models with added layers and the DXA dataset. In the second experiment (Experiment II), the influence of data augmentation on selected model (MobileNetV2 DW) was evaluated. In both experiments, the number of epochs was set to 200, and the base learning rate was set to 0.0001. All layers in the base models were frozen. The number of parameters in each model is presented in Table 1. In the third experiment (Experiment III) we evaluated the one, selected model on RSNA 2017[17] bone age dataset. The depthwise convolution layer was adjusted to larger images (12x12 filter), and the number of epochs

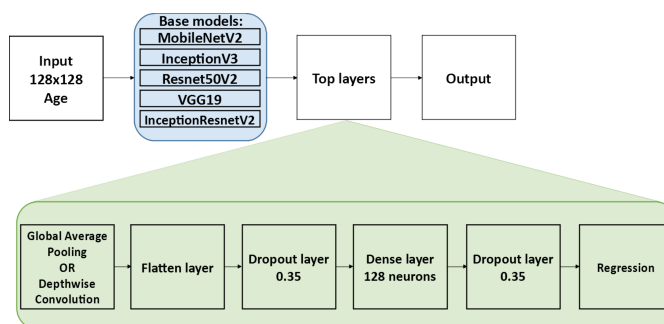


Figure 3. Scheme of proposed models used in the study. The blue block represents the base models, and the green block represents the top layers.

was reduced to 50. Additionally, we evaluated the model presented in [14] (Attention-based Model) using the DXA and augDXA datasets. The images were padded to 224x224 pixels and the model was adapted to the same size. In the fourth experiment (Experiment IV) we investigated whether the usage of the DXA images improves prediction of age in comparison to single-energy X-ray images. For this purpose, the 5-fold cross-validation (the MobileNetV2 DW model) was conducted using the images obtained by concatenating the low- and high-energy arrays separately. Similarly, the images consisting only of the copies of arrays corresponding to low/high energy passing one of the filters were tested. All calculations were performed at the NVIDIA DGX-1 station available at the AI Living Lab, Gdansk University of Technology.

Table 1: Number of parameters for each of the adapted models.

Adapted Model	Total parameters	Trainable parameters
MobileNetV2 DW	2,443,841	185,857
MobileNetV2 GA	2,422,081	164,097
VGG16	14,789,185	74,497
InceptionV3	22,075,425	272,641
ResNet50V2	23,862,017	297,217
InceptionResNetV2	54,541,281	204,545

3. Results

The performance of each model was evaluated with the mean absolute error (MAE) [23] expressed in months. For each fold, the experiment was conducted five times and the mean MAE value and related standard deviation were calculated. Additionally, the mean MAE value and standard deviation was calculated for all folds. The vanilla gradient method [24] was used to inspect the resulting models, which body parts are processed to obtain the final regression results. To use the vanilla gradient, the regression model (MobileNetV2 DW) was converted to a classification model.

3.1. Experiment I - DXA dataset

The results of Experiment I are presented in Table 2. The best mean result is achieved for the MobileNetV2 DW model. The calculated average MAE ranges from 13.39 months to 16.67 months (the mean MEA is equal to 15.56 months). The lowest standard deviation calculated for all folds is achieved for the ResNet50V2 DW model - 0.58 months. It is important to underline that the mean MAE obtained for the MobileNetV2 DW is lower by 1.52 months than MobileNetV2 GA. Figure 4 shows the learning curves for MobileNetV2 DW.



Table 2: Average values and standard deviations of MAE obtained in Experiment I - different models, DXA dataset. The mean MAE and standard deviation in last row were calculated from all folds.

Fold	MobV2 DW	MobV2 GA	VGG16 DW	Inception V3 DW	ResNet 50v2 DW	Inception ResNetV2 DW
1	13.39 ± 0.11	15.70 ± 0.09	15.64 ± 0.10	15.63 ± 0.07	14.67 ± 0.16	15.27 ± 0.10
2	16.29 ± 0.09	17.73 ± 0.06	16.25 ± 0.05	16.72 ± 0.13	15.72 ± 0.15	16.99 ± 0.14
3	16.58 ± 0.07	17.58 ± 0.42	16.49 ± 0.13	17.14 ± 0.37	16.03 ± 0.10	17.56 ± 0.08
4	14.85 ± 0.01	16.31 ± 0.07	16.74 ± 0.05	17.09 ± 0.03	15.95 ± 0.10	17.69 ± 0.12
5	16.67 ± 0.07	18.11 ± 0.07	18.30 ± 0.03	18.85 ± 0.06	16.39 ± 0.11	18.17 ± 0.07
\overline{MAE}	15.56 ± 1.26	17.08 ± 0.92	16.69 ± 0.89	17.09 ± 1.04	15.75 ± 0.58	17.14 ± 1.00
MAE[months]: mean and standard deviation						

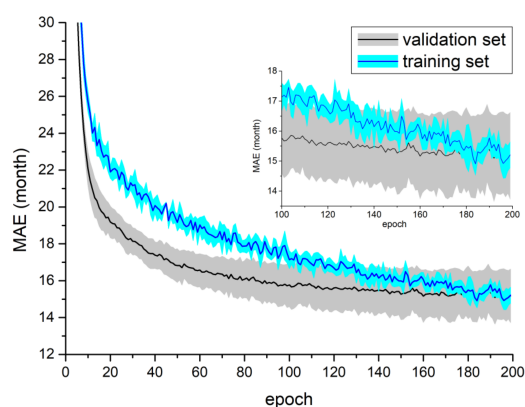


Figure 4. Training process for the MobileNetV2 DW model using DXA dataset - mean and standard deviation calculated from all folds in single test.

3.2. Experiment II - augDXA dataset

The results of Experiment II are presented in Table 3. After data augmentation, improvement can be observed (decrease of average MAE up to 1.2 months) for each fold with the exception of first (the model MobileNetV2 DW). The calculated average MAE ranges from 13.77 months to 15.82 months (the mean MEA is equal to 15.02 months). Figure 5 shows the learning curves for the model MobileNetV2 DW using augDXA dataset.

3.3. Experiment III - RSNA 2017 challenge dataset and Attention Model

The results of Experiment III are presented in Table 4. The MobileNetV2 DW model obtained a mean MAE equal to 16.29 months using the RSNA 2017 dataset. Figure 6 shows the learning curves for the MobileNetV2 DW model for the RSNA 2017 dataset. The Attention-based model achieved mean MAE is equal to 19.48 months for the DXA dataset and 17.05 months for the augDXA dataset. It should be noted that the mean MAE obtained for the Attention-based model (the DXA dataset) is greater by 3.92 months than the MobileNetV2 DW model. The learning curves are shown in Figure 7 - the temporal increase of MAE for the validation dataset can be observed after the fourth epoch.

Table 3: Average values and standard deviations of MAE obtained in Experiment II - MobileNetV2 DW model, augDXA dataset. The mean MAE and standard deviation in the last row were calculated from all folds.

Fold	MAE[months]: mean and SD
1	13.77 ± 0.12
2	15.82 ± 0.09
3	15.75 ± 0.11
4	14.30 ± 0.04
5	15.47 ± 0.09
<i>MAE</i>	15.02 ± 0.83

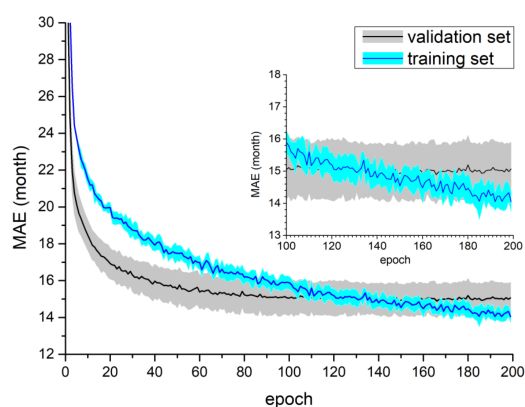


Figure 5. Training process for the MobileNetV2 DW model using augDXA dataset - mean and standard deviation calculated from all folds in a single test.

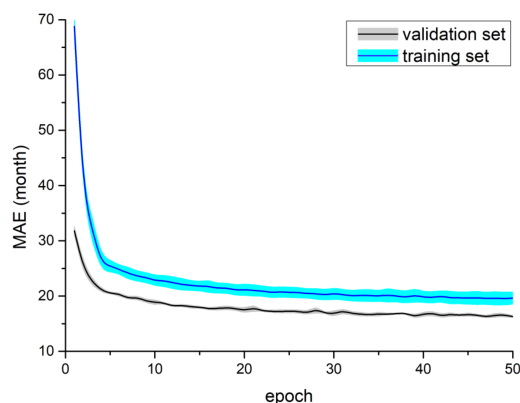


Figure 6. Training process for the MobileNetV2 DW model using the RSNA 2017 challenge dataset - mean and standard deviation calculated from five tests.

3.4. Experiment IV - the single energy x-ray images

The results of Experiment IV are presented in Table 5. Regardless of the type of image used in the prediction, the obtained values of mean MAE and standard deviation differ only slightly (the maximum difference in mean MAE is equal to 0.5 months).

3.5. Vanilla gradient

The results of the Vanilla gradient algorithm are shown in Figure 8. Analysing the results of the algorithm, it can not be clearly stated which body parts are processed to

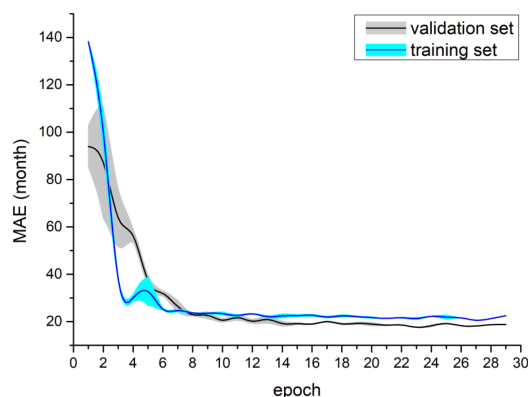


Figure 7. Training process for Attention model using DXA dataset - mean and standard deviation calculated from five tests.

Table 4: Mean values and standard deviations of MAE obtained in Experiments I, II and III.

Summary		
Model	Dataset	MAE[months]: mean and SD
MobileNetV2 DW	DXA	15.56±1.26
	augDXA	15.02±0.83
	RSNA 2017	16.29±0.19
Attention Model	DXA	19.48±0.77
	augDXA	17.05±0.55
	RSNA 2017	11.45[14]

Table 5: Mean values and standard deviations of MAE obtained in the Experiment IV - MobileNetV2 DW, different images.

Images	MAE[months]: mean and SD
DXA	15.56±1.26
high energy	15.78±1.13
low energy	15.43±1.04
high air	15.93±1.15
low air	15.61±1.27
high tissue	15.74±1.13
low tissue	15.52±1.29
high bone	15.89±1.04
low bone	15.47±1.04

obtain the final classification result, but it can be noticed that the features determining the classification of the DXA image were mostly within the outline of the body. This confirms that the model predicts, based on the image of the human body, not the background.

4. Discussion

The proposed solution achieved satisfying results for the age estimation using the full-body, low-resolution DXA scan - mean MAE equal to 15.56 months. First, the results obtained are the best among all evaluated models. Second, it is not expected to obtain MAE = 0, due to differences in body development of young subjects (including pathological changes). For example, hypothyroidism delays skeletal development and thyrotoxicosis accelerates bone maturation [25]. The analyzes of samples with greater differences between

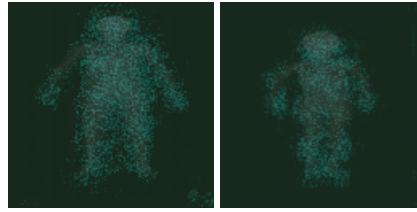


Figure 8. The results of the vanilla gradient algorithm were placed on top of the original image, and its colour was changed to teal.

actual and predicted age could, for example, indicate a connection between MAE and different disorders such as the bias in bone mineralization during maturation. Therefore, the difference between chronological age and DXA related biological age could be potentially important for diagnostics purposes [25], [26].

It is necessary to point out that the prediction of age using single x-ray and DXA images yields similar results. This may suggest that spatial (geometrical) features are the dominant factor in age prediction using the investigated CNNs. The results of the Vanilla gradient algorithm seem to confirm this statement, as the distribution of the points of interest, though not perfectly even, lacks any clusters. The issues related to connection of DXA images and different parameters (e.g. BMD) will be the subject of future work.

Different configurations of the models were investigated. Firstly, we investigated different base models, which proved to be efficient for other computer vision tasks in healthcare (MobileNetV2 [18], VGG16 [19], InceptionV3 [20], ResNet50V2 [21] and InceptionResNetV2 [22]). All of these base models showed similar results. However, introducing the depthwise convolution layer instead of the global pooling layer improved the model quality, reducing the MAE value by 1.52 months. Investigated data augmentation methods also reduced the MAE value (by 0.54 months).

The proposed framework was compared with known experiments for the RSNA 2017 dataset. The result of the cross-check analysis showed the importance of adjusting model architecture to the analyzed dataset. For example, the validation results of the retrained Attention-based model [14] for our dataset showed much worse results than for the proposed model. The difference in the related mean MAE values was approximately 4 months. In the cross-check we validated our framework for the RSNA 2017 dataset. The results obtained were worse by less than 5 months (mean MAE value). Other different models were proposed for the RSNA 2017 dataset. For example, in [15] the authors achieved mean MAE equal to 6.05 months. In another paper [16] authors obtained mean MAE equal to 6.38. However, it is important to underline that the images in RSNA 2017 dataset have minimal resolution 512x512 pixels and represent a slightly different domain (only hand). For comparison purposes only, the size hand in images used for this study is, e.g. about 15x15 pixels.

For the MobileNetV2 DW model, the list of five samples from validation set with the greatest MEA was determined for each fold in two out of five repetitions of experiment I (total 25 samples - 14 females and 11 males). In both repetitions, exactly the same samples were determined, although their position on the list was different (with the exception of sample with the greatest MEA for each fold - they remained on same position). The difference between the mean age of all samples and the mean age of the 25 worst samples was within the standard deviation. Similarly, also for augDXA DW the list of five samples from the validation set with the greatest MEA was determined for each fold. Of the total of 25 samples on the list in experiment I, 17, repeated on the list in experiment II (including the sample with the greatest MEA for each fold). Ten of the samples with the greatest MAE were submitted for consultation with the radiologist. As a result of the consultations, it was indicated that the majority of those samples were obtained from subjects too tall or too short for their age. This also supports the statement that the convolutional neural network predicts age on the basis of geometric features.

5. Conclusions

It was shown that it is possible to estimate the age of the subject based on the low-resolution image obtained from DXA scan, although the convolutional neural network focuses on spatial features in the area of the whole body. As it was experimentally verified, for low-resolution images the multipsectral features did not improve the final results. It is worth noting that both the usage of the depthwise convolution layer and data augmentation improved results. Future work should focus on extraction information correlated dual-energy X-ray absorptiometry, by taking into account the different parameters (e.g. bone mineral density) and additional metadata (e.g. gender), coupled with image segmentation. We also plan to expand the conducted study, by investigating different regression models. It is necessary to point out that all data used in this study were collected with the same densitometer. To properly validate the proposed solution, the DXA data from the other sources should be obtained and tested. Lastly, the additional experiments should be conducted using higher resolution images (e.g. DXA scans of body parts).

Author Contributions:

Conceptualization, Jacek Rumiński, Kamil Janczyk and Tomasz Neumann; methodology, Jacek Rumiński, Kamil Janczyk; software, Kamil Janczyk; validation, Jacek Rumiński, Tomasz Neumann; formal analysis, Jacek Rumiński, Piotr Wiśniewski; investigation, Kamil, Janczyk, Jacek Rumiński; resources, Tomasz Neumann, Piotr Wiśniewski; writing—original draft preparation, Kamil Janczyk, Natalia Głowacka; writing—review and editing, Kamil Janczyk, Piotr Wiśniewski, Jacek Rumiński, Tomasz Neumann, Natalia Głowacka; visualization, Kamil Janczyk, Tomasz Neumann; supervision, Jacek Rumiński; project administration, Jacek Rumiński; data acquisition, Piotr Wiśniewski.

Funding: This research received no external funding

Institutional Review Board Statement: This study was approved by the Independent Bioethics Committee for Scientific Research at Medical University of Gdańsk (Approval No. NKBBN/124/2022).

Acknowledgments: This work has been partially supported by Gdansk University of Technology, the Faculty of Electronics, Telecommunications and Informatics.

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

Abbreviations

The following abbreviations are used in this manuscript:

DXA	dual-energy x-ray absorptiometry
MAE	mean absolute error
CNN	convolutional neural network
ROI	regions of interest
BMD	bone mineral density
BMC	bone mineral content
RSNA	Radiological Society of North America

References

1. Briot, K. DXA Parameters: Beyond Bone Mineral Density. *Joint Bone Spine*, **2013**, vol. 80, no. 3, pp. 265–269,
2. Satoh, M. Bone Age: Assessment Methods and Clinical Applications. *Clinical Pediatric Endocrinology*, **2015**, vol. 24, no. 4, pp. 143–152,
3. Hepe, D.H.M., et al. Bone Age Assessment by Dual-Energy X-Ray Absorptiometry in Children: An Alternative for X-Ray? *The British Journal of Radiology*, **2012** vol. 85, no. 1010, pp. 114–120,
4. Castillo, R. F.; Maria del C.L.R. Assessment of Age and Sex by Means of DXA Bone Densitometry: Application in Forensic Anthropology. *Forensic Science International*, **2011**, vol. 209, no. 1-3, pp. 53–58,
5. Navega, D., et al. DXAGE: A New Method for Age at Death Estimation Based on Femoral Bone Mineral Density and Artificial Neural Networks. *Journal of Forensic Sciences*, **2017**, vol. 63, no. 2, pp. 497–503,
6. Lee, H., et al. Fully Automated Deep Learning System for Bone Age Assessment. *Journal of Digital Imaging*, **2017**, vol. 30, no. 4, pp. 427–441,
7. Igloukov, V. I.; Rakhlin, A.; Kalinin, A. A. & Shvets, A. A. Paediatric bone age assessment using deep convolutional neural networks. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, Cham, Switzerland, 2018, pp. 300–308.

8. Karargyris, A.; Kashyap, S.; Wu, J. T.; Sharma, A.; Moradi, M. & Syeda-Mahmood, T. Age prediction using a large chest x-ray dataset. *In Medical Imaging 2019: Computer-Aided Diagnosis*, 2019, Vol. 10950.
9. Xue, Z.; Antani, S.; Long, L. R. & Thoma, G. R. Using deep learning for detecting gender in adult chest radiographs. *In Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications* 2018, Vol. 10579
10. Xue Z.; Rajaraman S.; Long R.; Antani S.; and Thoma G. Gender Detection from Spine X-Ray Images Using Deep Learning, 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS), pp. 54-58, Karlstad University, Sweden, 2018
11. Marouf M.; Siddiqi R.; Bashir F. & Vohra B.; Automated Hand X-Ray Based Gender Classification and Bone Age Assessment Using Convolutional Neural Network, 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), pp. 1-5, Sukkur IBA University Pakistan, 2020.
12. Kaloi, M.A.; He, K. Child Gender Determination with Convolutional Neural Networks on Hand Radio-Graphs. 2018, arXiv:1811.05180
13. Nguyen H.; Soohyung K. Automatic Whole-body Bone Age Assessment Using Deep Hierarchical Features. 2019, arXiv:1901.10237
14. Castillo, J.; Tong, Y.; Zhao, J. & Zhu, F. (2018). RSNA Bone-age Detection using Transfer Learning and Attention Mapping.
15. Liu, B., et al. Bone Age Assessment Based on Rank-Monotonicity Enhanced Ranking CNN. *IEEE Access*, vol. 7, 2019, pp. 120976–120983.
16. Salim, I. & Ben Hamza A. Ridge Regression Neural Network for Pediatric Bone Age Assessment. *Multimedia Tools and Applications*, 2021, vol. 80, no. 20, pp. 30461–30478.
17. Halabi, S.S., et al. The RSNA Pediatric Bone Age Machine Learning Challenge. *Radiology*, 2019, vol. 290, no. 2, pp. 498–503.
18. Sandler, M; et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA June 2018.
19. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014, arXiv 1409.1556.
20. Szegedy, Ch. et al. Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA, 2016
21. He K.; Zhang X.; Ren S. & Sun, J. Identity Mappings in Deep Residual Networks. 2016, arXiv:1603.05027
22. Szegedy, C.; Ioffe, S.; Vanhoucke, V. & Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017, San Francisco, USA
23. Geron, A. *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems*, 2nd ed.; O'Reilly, Sebastopol, California, USA, 2019
24. Sicara/Tf-Explain, GitHub, Available online: github.com/sicara/tf-explain (accessed on 7 Apr. 2021)
25. Williams G. R. Actions of thyroid hormones in bone. *Endokrynologia Polska*, 2009, vol.60, pp. 380–388.
26. Hill, R.J. et al. Bone Health in Children with Inflammatory Bowel Disease: Adjusting for Bone Age. *Journal of Pediatric Gastroenterology & Nutrition*, 2009, vol. 48, no. 5, pp. 538–543.