Using Machine Learning to Predict Bone Mineral Density from Dual-energy X-ray Absorptiometry Images of the Lumbar Spine

Nikola Kirilov[1], Elena Kirilova[2], Evgeniy Krastev[3]

[1]Medical Faculty, Medical University of Pleven, 5800, Bulgaria,
[2]Medical Faculty, University Prof Dr Assen Zlatarov, 1 Prof Yakimov Blvd., 8010 Burgas, Bulgaria
[3]Faculty of Mathematics and Informatics, University of Sofia St. Kliment Ohridsky, 5 James Bourchier Blvd., 1164 Sofia, Bulgaria

kirilov_9@abv.bg

Abstract. Machine learning is widely used nowadays in many fields of science. Of particular interest is its application in the image processing for image classification and prediction. Image data is extensively used and generated in medicine and healthcare, especially by radiological and medical imaging examinations. Bone mineral density (BMD) is a value, which is acquired through dual-energy x-ray absorptiometry scans (DEXA) of the lumbar spine using low energy x-ray beams. The objective of this paper is to create a convolutional neuronal network model using popular open-source machine learning frameworks like TensorFlow in Python to predict BMD values from DEXA images of the lumbar spine. The machine learning neuronal network is trained with a large set of image data and tested with a testing split, assessing its accuracy through mean absolute error and the standard deviation of absolute error. Furthermore, the predicted values are correlated to the actual ones in order to examine the predictive accuracy of the model.

Keywords: Machine Learning, Convolutional Neuronal Network, Image Processing, Dual-Energy X-Ray Absorptiometry, Bone Mineral Density.

1   Introduction

Nowadays machine learning has been used in large variety of scientific fields. With the advancement of computer power it becomes even more incorporated into everyday life. One of the most interesting areas, which have a potential to take advantage of this new technology, are medicine and healthcare. The brightest examples of such an application are the automated interpretation of electrocardiograms, disease identification and diagnosis, personalized treatment, drug discovery etc. [1].

Probably, one of the most astonishing use of machine learning is in the Radiology and Medical imaging. This problem could be compared with the plain application of machine learning in the image processing, given the fact how much image data is being generated everyday worldwide [2]. The main tasks of the computer are to classify images into groups or to be able to predict continuous values after analyzing an image. This is accomplished with artificial neuronal networks (ANN) and most frequently convolutional neural network (CNN). CNN is a class of deep neural networks used to analyze image data, which are multilayer and fully connected. Each neuron is connected to all neurons in the next layer, a fact that makes them susceptible to over-fitting data. CNN's are a crucial part of the sub-field of machine learning called deep learning [3]. The neuronal networks are included in a model with several layers. The last one are then trained with a training dataset and tested using a testing dataset to inspect its predictive accuracy. There are many open-source frameworks available for the implantation of such models including Keras and TensorFlow [4].

In Radiology, deep learning supplies the research and diagnostic process with state-of-the-art detection, segmentation, classification, and prediction facilitating the work of physicians and scientists [5, 6]. Radiology and Medical imaging includes broad spectrum of methods: x-ray, computed tomography, magnetic resonance tomography, ultrasound etc. One of the subjects of this clinical specialty is the measurement of the bone density, quality and fracture risk [7]. There are several approaches for the accomplishment of this assessment, namely, dual-energy x-ray absorptiometry (DEXA), quantitative computed tomography and ultrasound [8, 9, 10]. The "gold-standard" for bone density measurement is the DEXA, which uses a low dose beam to acquire images of the lumbar spine and to compute the bone mineral density (BMD) values [11, 12, 13]. Previous authors have studied if BMD could be predicted from CT images using deep learning and CNN, acquiring promising results [14].

In this paper, we are going to demonstrate the use of machine learning to predict the BMD values from DEXA images of the lumbar spine. The purpose of Section 2 is to provide the materials and methods used to create CNNs, which we train and consequently test using a significantly large dataset. In Section 3, we study the results and performance of the CNN in predicting the BMD values, which shows a strong correlation to the actual values.

2 Materials and methods

In our study, we used 4,894 images of the lumbar spine acquired from a DEXA densitometer. These images were used to train a CNN to predict the BMD value of a scan image. CNN is a type of neural networks commonly used for analyzing image data, which has many advantages over other network types. TensorFlow

is a free and open-source Python library, which could easily implement CNNs and in combination with another Python's libraries like Open Source Computer Vision Library (OpenCV) could be used for the purpose of our study.

The image data was converted to the PNG format with resolution 800x800 pixels, which were subsequently resized to 256 x 256 pixels before being fed to the model (see Fig.1).
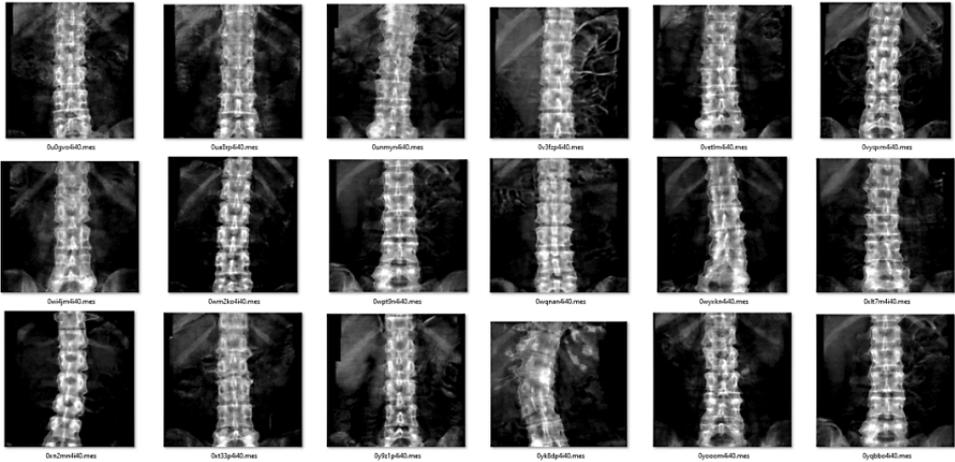

Fig. 1. Image data used to train and test the model.

Each image's actual measured BMD was recorded in a CSV file. The image paths and their corresponding BMD values were put in two lists in Python. The paths list has been looped. Next, images have been loaded and resized with OpenCV. The chosen resolution of 256 x 256 pixels for the training data was found as optimal due to reduced predictive accuracy using lower resolution and extreme process time using resolutions higher than 256 pixels. The data used for training was 75% and the remaining 25% of the data was used for testing the model. Partitioning the data into training and testing splits was done using the Scikit-learn library.
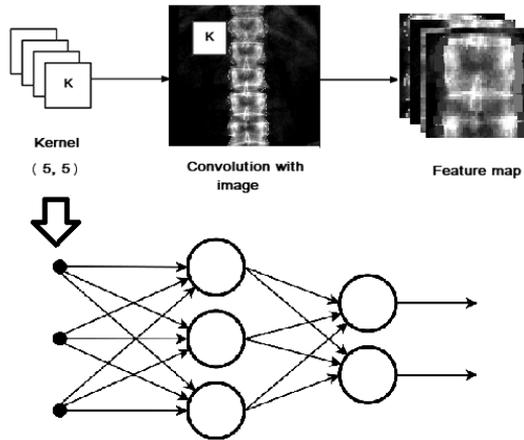
Fig. 2. Convolution of the original images with the Kernel creating a feature map, which is fed to the CNN layers for training.

The input data was a multi-dimensional array with size 4894 x 256 x 256 x 3 (image count, height, width, input channels), which is a cornerstone in machine learning called a tensor. The CNN consisted of three layers with filter numbers respectively (32, 64, 128) suited to the image resolution. The first layer learned 32 filters, the second 64 and the last 128, increasing while approaching the output. The use of more filters did not show any benefit, but increased processing time. The tensor was passed through each convolutional layer of our CNN in order to abstract the image to a feature map, convoluting it with a (5, 5) kernels. The kernel size was chosen according to the input resolution of the image. Rectified Linear Unit (ReLu) was used as an activation function (see Fig.2).

A regression was performed with mean absolute error (MAE) as a loss function. To optimize the training process of the CNN we used Adam's optimization algorithm instead of the classical stochastic gradient descent procedure, which improved the training speed greatly. The model's prediction accuracy was then assessed by the MAE and the standard deviation of absolute error (STD of AE) of the testing set. Additionally the predicted values were correlated with the actual values and Pearson's correlation coefficient was calculated.

3   Results

From all the 4,894 images 3,670 of them were used to train the model in 15 epochs. The training loss and validation loss percentages decreased significantly after the first 3 epochs and kept doing so until the 10th epoch where some overfitting started to appear (see Fig.3). After then the training loss kept declining

while validation loss showed abrupt increase. As a result, no further iterations were carried out. In the final epoch, the model calculated loss of 9.5 % and validation loss of 6.75 %.
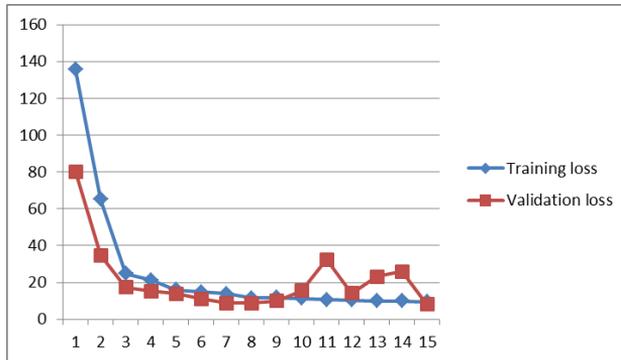

Fig. 3. Training loss and validation loss values in percent (%) during each epoch of training.

The testing was done using the remaining 1,224 of total 4,894 images. It yielded MAE of 8.19 % and STD of AE 6.75%, meaning that the network had a maximum deviation of 6.75 % off the actual BMD values. All 1,224 predicted values were correlated to the actual BMD values. A Pearson's correlation coefficient was calculated to assess their relationship (see Table 1). There was a positive correlation between the two variables, $r = 0.818$, $n = 1,224$, $p = 0.000$, which could be considered a strong positive correlation as seen on Fig. 4.

Table 1. Correlation between predicted BMD and actual BMD.

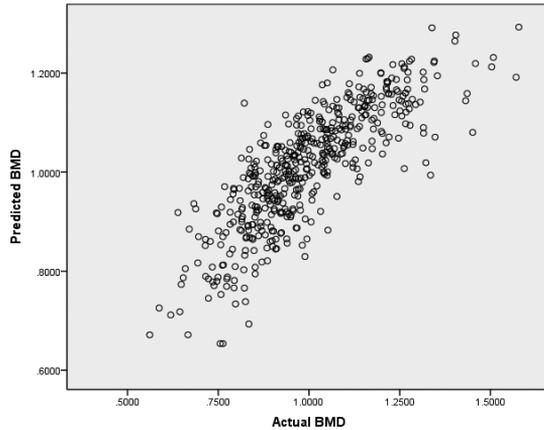|		Predicted BMD	Actual BMD
Predicted BMD	Pearson Correlation	1	0.818[**]
	Sig. (2-tailed)		0.000
	N	1,224	1,224
Actual BMD	Pearson Correlation	0.818[**]	1
	Sig. (2-tailed)	0.000	
	N	1,224	1,224

Fig. 4. Scatterplot of the predicted BMD and actual BMD.

4 Conclusion

Machine learning is a powerful technique, which finds application in many fields of science, including medicine and medical image analysis. Open-source frameworks like TensorFlow make algorithms available and easy to use. Our study showed that CNNs could be trained to predict BMD values from DEXA images of the lumbar spine showing good accuracy and great potential. The results were strong correlated to the actual data, which supported the statement furthermore.

5 Acknowledgment

References

1. Deo R. C. Machine Learning in Medicine. Circulation. 2015; 132(20):1920-1930. doi:10.1161/CIRCULATIONAHA.115.001593Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016).
2. Lezoray O., C. Charrier, C. Hubert, L. Sébastien. (2008). Machine Learning in Image Processing – Special Issue Editorial. EURASIP J. Adv. Sig. Proc. 2008. 10.1155/2008/927950.Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010).
3. Krishna M, M. Neelima, M. Harshali, M. Venu. (2018). Image classification using Deep learning. International Journal of Engineering & Technology. 7. 614. 10.14419/ijet.v7i2.7.10892.
4. Bisong E. (2019). TensorFlow 2.0 and Keras. 10.1007/978-1-4842-4470-8_30.
5. Montagnon, E., Cerny M., Cadrin-Chênevert, A. et al. Deep learning workflow in radiology: a primer. Insights Imaging 11, 22 (2020). https://doi.org/10.1186/s13244-019-0832-5.

6. Yoo T.K., Kim S.K., Kim D.W., Choi J.Y., Lee W.H., Oh E., Park E.C. Osteoporosis risk prediction for bone mineral density assessment of postmenopausal women using machine learning. Yonsei Med J. 2013 Nov; 54(6):1321-1330.
7. Smets, J., Shevroja E., Hügle T., Leslie W.D. and Hans D. (2021), Machine Learning Solutions for Osteoporosis—A Review. J Bone Miner Res. https://doi.org/10.1002/jbmr.4292.
8. Yasaka, K., Akai H., Kunimatsu A. et al. Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network. Eur Radiol 30, 3549–3557 (2020). https://doi.org/10.1007/s00330-020-06677-0.
9. Atanasov A., St. Kuzmanova, A. Batalov, S. Tsvetkova. Axial osteoporosis in premenopausal women with rheumatoid arthritis according to the mobility index, Rheumatology, (2000) 30 – 33.
10. R. Karalilova, A. Batalov, Z. Batalov. Management of Aromatase inhibitors – induced bone loss in postmenopausal women with hormone receptor positive breast cancer. Osteoporosis International (2014), Vol 25, Supplement 2: P583.
11. Leslie W.D., Shevroja E., Johansson H., et al. Risk-equivalent T-score adjustment using lumbar spine trabecular bone score (TBS): The Manitoba BMD Registry. Osteoporos Int. 2018b; 29:751– 758.
12. McCloskey E.V., Odén A., Harvey N.C., Leslie W.D., Hans D., Johansson H., Barkmann R., Boutroy S., Brown J., Chapurlat R., Elders P.J.M., Fujita Y., Glüer C.C., Goltzman D., Iki M., Karlsson M., Kindmark A., Kotowicz M., Kurumatani N., Kwok T., Lamy O., Leung J., Lippuner K., Ljunggren Ö., Lorentzon M., Mellström D., Merlijn T., Oei L., Ohlsson C., Pasco J.A., Rivadeneira F., Rosengren B., Sornay-Rendu E., Szulc P., Tamaki J., Kanis J.A. A meta-analysis of trabecular bone score in fracture risk prediction and its dependence on FRAX. J Bone Miner Res. 2016; 31:940–948.
13. McCloskey E.V., Odén A., Harvey N.C., Leslie W.D., Hans D., Johansson H., Kanis J.A. Adjusting fracture probability by trabecular bone score. Calcif Tissue Int. 2015b; 96:500–509.
14. Yoo T.K., Kim S.K., Kim D.W., Choi J.Y., Lee W.H., Oh E., Park E.C. Osteoporosis risk prediction for bone mineral density assessment of postmenopausal women using machine learning. Yonsei Med J. 2013 Nov; 54(6):1321-1330.