Machine Learning Model to Predict Osteoporotic Spine with Hounsfield Units on Lumbar Computed Tomography

Kyoung Hyup Nam, Il Seo, Dong Hwan Kim, Jae Il Lee, Byung Kwan Choi, In Ho Han

Department of Neurosurgery, Medical Research Institute, Pusan National University Hospital, Busan, Korea

Objective: Bone mineral density (BMD) is an important consideration during fusion surgery. Although dual X-ray absorptiometry is considered as the gold standard for assessing BMD, quantitative computed tomography (QCT) provides more accurate data in spine osteoporosis. However, QCT has the disadvantage of additional radiation hazard and cost. The present study was to demonstrate the utility of artificial intelligence and machine learning algorithm for assessing osteoporosis using Hounsfield units (HU) of preoperative lumbar CT coupling with data of QCT.

Methods: We reviewed 70 patients undergoing both QCT and conventional lumbar CT for spine surgery. The T-scores of 198 lumbar vertebra was assessed in QCT and the HU of vertebral body at the same level were measured in conventional CT by the picture archiving and communication system (PACS) system. A multiple regression algorithm was applied to predict the T-score using three independent variables (age, sex, and HU of vertebral body on conventional CT) coupling with T-score of QCT. Next, a logistic regression algorithm was applied to predict osteoporotic or non-osteoporotic vertebra. The Tensor flow and Python were used as the machine learning tools. The Tensor flow user interface developed in our institute was used for easy code generation.

Results: The predictive model with multiple regression algorithm estimated similar T-scores with data of QCT. HU demonstrates the similar results as QCT without the discordance in only one non-osteoporotic vertebra that indicated osteoporosis. From the training set, the predictive model classified the lumbar vertebra into two groups (osteoporotic vs. non-osteoporotic spine) with 88.0% accuracy. In a test set of 40 vertebrae, classification accuracy was 92.5% when the learning rate was 0.0001 (precision, 0.939; recall, 0.969; F1 score, 0.954; area under the curve, 0.900).

Conclusion: This study is a simple machine learning model applicable in the spine research field. The machine learning model can predict the T-score and osteoporotic vertebrae solely by measuring the HU of conventional CT, and this would help spine surgeons not to under-estimate the osteoporotic spine preoperatively. If applied to a bigger data set, we believe the predictive accuracy of our model will further increase. We propose that machine learning is an important modality of the medical research field.

Key Words: Artificial intelligence · Machine learning · Tensor flow · Osteoporosis · Spine · Hounsfield unit.
INTRODUCTION

Osteoporosis is an important prognostic factor for spinal instrumentation. Severe osteoporosis often results in hardware failure such as screw loosening and cage subsidence. The precise diagnosis of osteoporosis before surgery is especially important when performing spinal fixation in older patients. In most cases, the diagnosis of osteoporosis is dependent on dual X-ray absorptiometry (DEXA), which often fails to accurately reflect the bone mineral density (BMD) in the degenerative spine of the elderly. Quantitative computed tomography (QCT) measures the volumetric trabecular bone density without superimposition of the cortical bone and other tissues and is reported to be more sensitive than DEXA for detecting osteoporosis in older patients. Lumbar CT is routinely performed for preoperative diagnosis, especially in fusion surgery. In lumbar CT, the Hounsfield unit (HU) of lumbar vertebral body can easily be measured by the picture archiving and communication system (PACS) system. Recent several studies have demonstrated the utility to estimate bone mineral density using HU from lumbar CT. We hypothesize that HU of the cancellous bone in lumbar CT is associated with the QCT value in the basic concept, and hence HU could be a representative value of the regional BMD.

Artificial intelligence (AI) has recently been a hot issue in the 4th industrial revolution. Machine learning (ML) is a key element of AI and is used as a tool for medical research. In particular, the core function of machine learning is prediction and classification. ML has been applied in several pharmaceutical area for diagnosis, imaging analysis and treatment optimization such as IBM Watson, most popular machine learning. If this function of prediction and classification is applied to the medical field, it would help in the diagnosis and treatment of patients. Based on the prediction function of MLs, we developed to estimate the T-score of lumbar spine, and classify osteoporotic vertebra and non-osteoporotic vertebrae the using HU of lumbar CT.

The purpose of the current study is to predict osteoporosis with preoperative lumbar CT through the machine learning model.
rior to the superior endplate, mid-vertebral body, and just superior to the inferior endplate. The ROI was drawn encapsulating only cancellous bone excluding cortical edge, osseous abnormalities, and voids such as vascular channels. Mean HU values were calculated in the ROI for each vertebral level (Fig. 1).

The measurements of HU were obtained by two observers independently. During the measurements, the observers were blinded to the QCT scores of the patient. Interobserver coefficient was 0.991 (95% confidence interval, 0.998–0.993) and $p$ value was below 0.001.

---

**Fig. 1.** The HU measurement with drawing of elliptical ROI on conventional lumbar CT scan. For mean HU assessment, the largest ROI was drawn excluding the cortical bone and vascular markings on 3 areas from each vertebra (superior and inferior to end plate, middle of the body). A: Sagittal image. B: Axial image: upper area. C: Axial image: middle area. D: Axial image: lower area. U: upper area, M: middle area, L: lower area, HU: Hounsfield units, ROI: regions of interest, CT: computed tomography.

**Fig. 2.** The Tensor flow and Python were used for machine learning tools and the Tensor flow user interface was developed at our institute. HU: Hounsfield units.
PyCham was used as an Integrated Development Environment to develop the predictive model. The Python language was used for coding the algorithm. Google TensorFlow (version 0.12; Google, Mountain View, CA, USA) was used as the machine learning library. The Tensor graphic user interface (Tensor GUI; Winform_parent version) was employed for easy code generation for machine learning and creating the server in order to predict a T-score and osteoporosis with the new data (Fig. 2).

Our models consisted of single layer with three input node (age, sex, and HU) and one output node (BMD of QCT). First, we applied a multi-variable linear regression algorithm to predict the T-score of each vertebra with HU of conventional CT. Age, sex, and the average HU of the vertebral body were used as independent variables. T-scores of QCT at each matched vertebra was used as a dependent variable. The numbers of vertebra for training set and test set were 158 and 40, respectively. We didn’t use validation set considering small dataset. The learning rate was 0.0001 and the number of training was 200000. We used minibatch for training with 100 of size. Mean squared error was used as cost function. Adam optimizer was used as an optimizer algorithm. After running the model, we assessed the final cost. The predictive T-scores of the test set was saved as a csv file and excel file and compared with the T-score of QCT.

Next, we applied a logistic regression algorithm to classify osteoporotic or non-osteoporotic vertebra with HU of lumbar CT. Age, sex, and the average HU of vertebral body were also used as independent variables. Osteoporosis and non-osteoporosis of each vertebra based on the criteria of QCT were decided as labels for binomial classification. The numbers of vertebra for training set and test set were 158 and 40. The learning rate was 0.0001. The number of training number was 10000. Batch size was 100 and cross entropy was used as cost function. Gradient descent optimizer was used. After running the model, we evaluated for the accuracy of training and test sets. F1 score, precision, and recall were evaluated considering imbalanced dataset. In addition, receiver operating curve (ROC) and area under the curve (AUC) was evaluated.

| No | T-score of QCT | Predict value from ML | No | T-score of QCT | Predict value from ML |
|----|----------------|-----------------------|----|----------------|-----------------------|
| 1  | -3.8           | -3.4                  | 21 | -3.7           | -3.5                  |
| 2  | -4.3           | -3.7                  | 22 | -4.5           | -4.2                  |
| 3  | -3.8           | -3.2                  | 23 | -4.8           | -4.5                  |
| 4  | -4.5           | -4.9                  | 24 | -3.6           | -2.4                  |
| 5  | -4.3           | -4.0                  | 25 | -1.8           | -3.0                  |
| 6  | -5.4           | -5.6                  | 26 | -3.9           | -4.6                  |
| 7  | -3.9           | -3.8                  | 27 | -3.9           | -4.4                  |
| 8  | -0.1           | 0.6                   | 28 | -4.1           | -4.6                  |
| 9  | -2.5           | -2.7                  | 29 | -2.7           | -2.5                  |
| 10 | -4.0           | -3.7                  | 30 | -2.7           | -3.3                  |
| 11 | -4.2           | -3.9                  | 31 | -2.8           | -2.8                  |
| 12 | -4.2           | -4.1                  | 32 | -4.6           | -4.7                  |
| 13 | -4.5           | -3.7                  | 33 | -5.5           | -5.1                  |
| 14 | -5.3           | -5.3                  | 34 | -4.8           | -4.8                  |
| 15 | -4.5           | -4.3                  | 35 | -3.9           | -3.9                  |
| 16 | -5.5           | -5.7                  | 36 | -3.5           | -4.0                  |
| 17 | -4.1           | -4.3                  | 37 | -3.6           | -3.8                  |
| 18 | 1.5            | 0.6                   | 38 | -3.9           | -3.9                  |
| 19 | -3.5           | -2.8                  | 39 | -3.7           | -3.5                  |
| 20 | -3.7           | -3.3                  | 40 | -4.0           | -4.2                  |

QCT: quantitative computed tomography, ML: machine learning
RESULTS

In the predictive model with multi-variable regression algorithm, the training cost and test cost was 31.227 and 4.45875, respectively. The predictive T-scores of the 40-test set was described in Table 1. Only one non-osteoporotic vertebra was incorrectly assessed as osteoporotic spine. Our data indicated

![Fig. 3. The created AI server to predict the T-score of the new vertebra by entering new dataset including age, sex, and average HU of lumbar CT. The “guess” indicates predictive T-score value. AI : artificial intelligence, HU : Hounsfield units, CT : computed tomography.](image)

![Fig. 4. ROC curve for machine learning prediction for osteoporosis based on HU measurement compared with QCT. ROC : receiver operating curve, AUC : area under the curve, HU : Hounsfield units, QCT : quantitative computed tomography.](image)

### Table 2. Logistic regression algorithm classifies the result to osteoporosis and non-osteoporosis with machine learning

| No | Result of QCT | Predict result from ML | No | Result of QCT | Predict result from ML |
|----|---------------|------------------------|----|---------------|------------------------|
| 1  | 0             | 0                      | 21 | 0             | 0                      |
| 2  | 0             | 0                      | 22 | 0             | 0                      |
| 3  | 0             | 1                      | 23 | 0             | 0                      |
| 4  | 0             | 0                      | 24 | 0             | 1                      |
| 5  | 0             | 0                      | 25 | 1             | 0                      |
| 6  | 0             | 0                      | 26 | 0             | 0                      |
| 7  | 0             | 0                      | 27 | 0             | 0                      |
| 8  | 1             | 1                      | 28 | 0             | 0                      |
| 9  | 1             | 1                      | 29 | 1             | 1                      |
| 10 | 0             | 0                      | 30 | 1             | 1                      |
| 11 | 0             | 0                      | 31 | 1             | 1                      |
| 12 | 0             | 0                      | 32 | 0             | 0                      |
| 13 | 0             | 0                      | 33 | 0             | 0                      |
| 14 | 0             | 0                      | 34 | 0             | 0                      |
| 15 | 0             | 0                      | 35 | 0             | 0                      |
| 16 | 0             | 0                      | 36 | 0             | 0                      |
| 17 | 0             | 0                      | 37 | 0             | 0                      |
| 18 | 1             | 1                      | 38 | 0             | 0                      |
| 19 | 0             | 0                      | 39 | 0             | 0                      |
| 20 | 0             | 0                      | 40 | 0             | 0                      |

QCT : quantitative computed tomography, ML : machine learning, 0 : osteoporosis, 1 : non-osteoporosis
that the created server can predict the T-score of new vertebrae by entering a new dataset including age, sex, and average HU of lumbar CT (Fig. 3).

In the predictive model with logistic regression algorithm, the accuracy of prediction to classify osteoporotic and non-osteoporotic spine was 88.0%. In a test set of 40 vertebrae, the classification accuracy was 92.5% (precision, 0.939; recall, 0.969; F1 score, 0.954); predictions for only three vertebrae were wrong. ROC curve showed that our model is a reliable classifier of spinal osteoporosis (AUC, 0.900) (Fig. 4). The final cost was 0.327243716. The classification result of test set is presented in Table 2. The prediction server also can classify the new vertebra as osteoporosis and non-osteoporosis by entering the new dataset (Fig. 5).

DISCUSSION

Osteoporosis is the major cause of hardware failures such as screw loosening and cage subsidence; hence, a clinical diagnosis of osteoporosis is very important in spinal instrumentation. DEXA and QCT are the most common tools employed to measure BMD. However, DEXA is unable to exactly predict spinal BMD since degeneration, aortic calcification, and soft tissue calcification over-estimate BMD, especially the elderly. Even though clinical finding may indicate osteoporosis, DEXA results may still indicate a normal BMD. Therefore, DEXA of hip is recommended in identification of osteoporosis in older people. However, the ability of DEXA to detect osteoporosis decrease since hip also contain cortical bone and degenerative changes and. Considering all the above problems, DEXA of the hip and spine often result in over-estimating the spinal and hip BMD. If spine surgeon performs spinal instrumentation overlooking under-estimated osteoporotic spine, the possibility of hardware failure can increase and appropriate time of osteoporosis medication can be missed. Thus, the main purpose of our ML model is to avoid the under-estimation of osteoporosis due to disadvantage of DEXA in spinal instrumentation.

QCT measures the BMD of the trabecular bone in spine and is more sensitive for detecting osteoporosis. Despite the high sensitivity of QCT, many hospitals are unequipped. Thus, our machine learning model can help to measure spinal BMD in a hospital without the QCT equipment. Even though the T-score is not the exact BMD of QCT, the predictive T-score is a value approximating the T-score of QCT. Furthermore, the T-score of our model does not overestimate the real BMD in patients with spinal degeneration or aortic calcification. In our study, age, sex, and the average HU of lumbar CT were used as independent variables. The T-score is the BMD at the site when compared to young normal reference mean. Hence, age and sex are related to the T-score and these data are used as independent variables. The average HU of lumbar CT is a strong indicator to relate with BMD and T-score of QCT. QCT measures the volumetric density of the vertebra and our HU of lumbar CT is measured as a cross-sectional area in two dimensions. However, we measured three different levels of the vertebra and reflected indirectly the density of the whole vertebral body. Lumbar CT is routinely performed in all patients undergoing lumbar spine fusion. In addition, spinal surgeons can easily measure the HU of lumbar CT by the PACS system. The merit of our model can predict the spinal T-score and osteoporosis of patients with these simple data.

The limitation of our predictive model was followed as below. First, an accuracy of 92% is still low for medical application. In order to improve the accuracy, a bigger data sets are required. A total of 188 data set is too less to acutely predict osteoporosis. In addition, standardized data is also required. Furthermore, ROI is different according to surgeons to draw in spite of measurement protocol. Thus, additional programs need to be developed to draw ROI of the vertebral body for more standardized data. Nevertheless, the predictive model is
valuable because it can predict very simply spinal T-score and BMD similar to that of QCT, and not under-estimate the BMD preoperatively. The second limitation is that our predictive model is not useful in cervical and thoracic instrumentation. In these surgeries, lumbar CT is routinely not performed preoperatively. QCT is performed only in lumbar spine. Hence, a label of cervical and thoracic spine, values of QCT are difficult to obtain. Therefore, our predictive model is useful only in lumbar instrumentation where lumbar CT is routinely performed.

Recent studies concerning the application of machine learning to medical fields has increased. Especially, ML has been applied to medical diagnosis including diagnostic imaging, genetic tests and electrodiagnosis. In clinical practice, IBM Watson is being used in cancer diagnosis and treatment decisions. In future, since a wide range of medical applications of ML will be inevitable, it is imperative for a physician to understand the basic concepts of ML. ML algorithms can be divided into two major categories: unsupervised learning and supervised learning. Supervised learning is the ML task of learning a function that maps an input to an output based on example input-out pairs. Among the many supervised learning algorithms, linear regression and logistic regression algorithms are used in our predictive model. Linear regression is a linear approach for modeling the relationship between independent and dependent variables, and this model can predict the dependent variable such as T-score with new independent variables such as age, sex, and HU of lumbar CT. In logistic regression, the dependent variable is categorized. By binary logistic regression, the output can be classified into two values, “0” and “1”, which represent the outcome as pass/fail, alive/dead, good prognosis/bad prognosis, benign/malignant, etc. In our logistic regression model, the output is classified as osteoporotic and non-osteoporotic spine by input of age, sex, and HU. Our predictive model is the simplest ML model. A physician can also implement a medical ML model through simple coding using Google TensorFlow and Python. In our opinion, to develop a more valuable medical predictive model, it is necessary to understand the basic concepts of AI and ML. In addition, physicians should be encouraged to collecting bigger data sets to improve the accuracy and value of a medical predictive model.

CONCLUSION

This study is a simple machine learning model applied to the spine research field. The machine learning model predicts the T-score and osteoporotic vertebrae by measuring the HU of conventional CT and helps spinal surgeons not to underestimate the osteoporotic spine preoperatively. If bigger data is collected, we believe that the predictive accuracy of our model will further increase. We therefore propose that machine learning will be an important modality of the medical research field and is not solely an engineering area.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

INFORMED CONSENT

Informed consent was obtained from all individual participants included in this study.

AUTHOR CONTRIBUTIONS

Conceptualization : KHN, IHH
Data curation : IS, DHK
Formal analysis : DHK, JIL
Methodology : KHN, IHH
Project administration : BKC, IHH
Visualization : KHN
Writing - original draft : KHN, IHH
Writing - review & editing : BKC, IHH

References

1. Bastanlar Y, Ozucal M : Introduction to machine learning. Methods Mol Biol 1107 : 105-128, 2014
2. Cheng Q, Zhu YX, Zhang MX, Li LH, Du PY, Zhu MH : Age and sex effects on the association between body composition and bone mineral density in healthy Chinese men and women. Menopause 19 : 448-455, 2012
3. Choi MK, Kim SM, Lim JK: Diagnostic efficacy of Hounsfield units in spine CT for the assessment of real bone mineral density of degenerative spine: correlation study between T-scores determined by DEXA scan and Hounsfield units from CT. *Acta Neurochir (Wien)* 158: 1421-1427, 2016

4. Coe JD, Warden KE, Herzig MA, McAfee PC: Influence of bone mineral density on the fixation of thoracolumbar implants. A comparative study of transpedicular screws, laminar hooks, and spinous process wires. *Spine (Phila Pa 1976)* 15: 902-907, 1990

5. Deo RC: Machine learning in medicine. *Circulation* 132: 1920-1930, 2015

6. Ebbesen EN, Thomsen JS, Beck-Nielsen H, Nepper-Rasmussen HJ, Mosekilde L: Lumbar vertebral body compressive strength evaluated by dual-energy X-ray absorptiometry, quantitative computed tomography, and ashing. *Bone* 25: 713-724, 1999

7. Erickson BJ, Korfiatis P, Akkus Z, Kline TL: Machine learning for medical imaging. *Radiographics* 37: 505-515, 2017

8. Forsting M: Machine learning will change medicine. *J Nucl Med* 58: 357-358, 2017

9. Han SS: Internal fixation in the osteoporotic spine. *Spine (Phila Pa 1976)* 19(24 Suppl): 2415-2420, 1994

10. Hu SS: Correlation between bone mineral density measured by dual-energy X-ray absorptiometry and Hounsfield units measured by diagnostic CT in lumbar spine. *J Korean Neurosurg Soc* 54: 384-389, 2013

11. Jergas M, Breitenseher M, Glüer CC, Black D, Lang P, Grampp S, et al.: Which vertebrae should be assessed using lateral dual-energy X-ray absorptiometry of the lumbar spine. *Osteoporos Int* 5: 196-204, 1995

12. Kotoku J: An introduction to machine learning. *Igaku Butsuri* 36: 18-22, 2016

13. Lee S, Chung CK, Oh SH, Park SB: Correlation between bone mineral density measured by dual-energy X-ray absorptiometry and Hounsfield units measured by diagnostic CT in lumbar spine. *J Korean Neurosurg Soc* 54: 384-389, 2013

14. Lochmüller EM, Bürklein D, Kuhn V, Glaser C, Müller R, Glüer CC, et al.: Mechanical strength of the thoracolumbar spine in the elderly: prediction from in situ dual-energy X-ray absorptiometry, quantitative computed tomography (QCT), upper and lower limb peripheral QCT, and quantitative ultrasound. *Bone* 31: 77-84, 2002

15. Masud T, Langley S, Wiltshire P, Doyle DV, Spector TD: Effect of spinal osteophytosis on bone mineral density measurements in vertebral osteoporosis. *BMJ* 307: 172-173, 1993

16. Matsukawa K, Abe Y, Yanai Y, Yato Y: Regional Hounsfield unit measurement of screw trajectory for predicting pedicle screw fixation using cortical bone trajectory: a retrospective cohort study. *Acta Neurochir (Wien)* 160: 405-411, 2018

17. Mounach A, Abayi DA, Ghazi M, Ghaziani I, Noujaidi A, Achemal L, et al.: Discordance between hip and spine bone mineral density measurement using DXA: prevalence and risk factors. *Semin Arthritis Rheum* 38: 467-471, 2009

18. Nguyen ND, Eisman JA, Center JR, Nguyen TV: Risk factors for fracture in nonosteoporotic men and women. *J Clin Endocrinol Metab* 92: 955-962, 2007

19. Nonaka K, Uchiyama S: Assessment of volumetric bone mineral density and geometry for hip with clinical CT device. *Clin Calcium* 21: 1003-1009, 2011

20. Reynolds RJ, Day SM: The growing role of machine learning and artificial intelligence in developmental medicine. *Dev Med Child Neurol* 60: 858-859, 2018

21. Schreiber JJ, Anderson PA, Rosas HG, Buchholz AL, Au AG: Hounsfield units for assessing bone mineral density and strength: a tool for osteoporosis management. *J Bone Joint Surg Am* 93: 1057-1063, 2011

22. Somashekhar SP, Sepulveda MJ, Puglielli S, Norden AD, Shortliffe EH, Rohit Kumar C, et al.: Watson for oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. *Ann Oncol* 29: 418-423, 2018

23. Suzuki K: Pixel-based machine learning in medical imaging. *Int J Biomed Imaging* 2012: 792079, 2012

24. Suzuki K, Yan P, Wang F, Shen D: Machine learning in medical imaging. *Int J Biomed Imaging* 2012: 123727, 2012

25. Yamagata M, Kitahara H, Minami S, Takahashi K, Isobe K, Moriya H, et al.: Mechanical stability of the pedicle screw fixation systems for the lumbar spine. *Spine (Phila Pa 1976)* 17(3 Suppl): S51-S54, 1992

26. Younes M, Ben Hammouda S, Jguirim M, Younes K, Zmour S, Béjia I, et al.: Discordance between spine and hip bone mineral density measurement using DXA in osteoporosis diagnosis: prevalence and risk factors. *Tunis Med* 92: 1-5, 2014