Annals of Nuclear Medicine
https://doi.org/10.1007/s12149-020-01524-0

ORIGINAL ARTICLE

The utility of a deep learning-based algorithm for bone scintigraphy in patient with prostate cancer

Yuki Aoki[1] · Michihiro Nakayama[1] · Kenta Nomura[1] · Yui Tomita[1] · Kaori Nakajima[1] · Masaaki Yamashina[1] · Atsutaka Okizaki[1]

Received: 13 July 2020 / Accepted: 8 September 2020

Abstract

Objective Bone scintigraphy has often been used to evaluate bone metastases. Its functionality is evident in detecting bone metastasis in patients with malignant tumor including prostate cancer, as appropriate treatment and prognosis are dependent on the presence and degree of bone metastasis. The development of a deep learning-based algorithm in the field of information processing has been remarkable in recent years. We hypothesized that a deep learning-based algorithm is useful in diagnosing osseous metastases in patients with prostate cancer using bone scintigraphy. Thus, this study aims to examine the utility of deep learning-based algorithm in detecting bone metastases in patients with prostate cancer, as compared with nuclear medicine specialists.

Methods In total, 139 serial patients with prostate cancer, who underwent whole-body bone scintigraphy, were enrolled in this study. Each scintigraphy examination was evaluated visually and independently by nuclear medicine specialists; this was also analyzed using a deep learning-based algorithm. The number of abnormal uptakes was assessed by the nuclear medicine specialists and with a software which used the deep learning-based algorithm, and the per-patient detection rate and the per-region detection rate were then calculated. The software automatically analyzed bone scintigraphy for the presence or absence of osseous metastasis in individual patients, for the 12 body regions. The detection rates analyzed separately by the nuclear medicine specialists and using the software were then compared. The sensitivity, specificity, and accuracy by the specialist and with the software were calculated.

Results The sensitivity, specificity, and accuracy by the nuclear medicine specialists were 100%, 94.9% and 97.1%. On the other hand, they with the software were 91.7%, 87.3% and 89.2%. No statistically significant difference was determined between the per-patient detection rates assessed by the specialists versus the software. In regional assessment, there was also no statistically significant difference between most of the per-region detection rates (10 of 12 regions) by the specialists versus the results obtained by the software.

Conclusions The software with the deep learning-based algorithm might be used as diagnostic aid in the evaluation of bone metastases for prostate cancer patients.

Keywords Bone scintigraphy · Bone metastases · Prostate cancer · Deep learning

Introduction

Bone scintigraphy has been identified as a useful imaging modality in detecting bone metastasis [1, 2]. The importance of bone scintigraphy in the detection of osseous metastasis in patients with malignant tumor, including prostate cancer, cannot be overstated, as appropriate treatment and prognosis are dependent on the presence and extent of bone metastasis [3–5]. Quantitative evaluation of bone metastases for patients with prostate cancer is important; therefore, Extent of Disease (EOD) was proposed as a method of evaluation. EOD is a semiquantitative index based on a 5-point scale, which is evaluated with the number of bone metastases [6]. EOD is simple but not sufficient for a precise estimation of the prognosis; thus, Erdi et al. [7] have proposed the use of the bone scan index (BSI). BSI is able to quantify bone

✉ Atsutaka Okizaki
okizaki@asahikawa-med.ac.jp

[1] Department of Radiology, Asahikawa Medical University, Midorigaoka-higashi 2-1-1-1, Asahikawa, Hokkaido 078-8510, Japan

Published online: 21 September 2020

metastases by calculating a percentage of abnormal uptake area, derived from bone metastases, in a whole-body bone scan. Wakabayashi et al. [8] and Dennis et al. [9] reported that BSI reflects the activity of bone metastasis; however, it requires lengthy procedures and has low reproducibility, as its calculation was not performed automatically when Erdi et al. first proposed the BSI [7]. Recently, BSI can be computed using artificial intelligence, which is implemented by an artificial neural network (ANN) [10, 11]. ANN is a component of artificial intelligence that is used to simulate the functioning of the human brain. ANN requires a training phase, wherein it learns how to recognize patterns in the data visually, audibly, or textually. In the training phase, we use a yes/no question type using binary numbers to instruct the ANN what to look for and what to output. This means that the learning information needs to be specified by a human, which can introduce bias or errors. Therefore, although this approach facilitates the rapid calculation of BSI as compared to manual procedures, it may also cause insufficient accuracy in the interpretation of bone scintigraphy. The ANN may not be as accurate as the nuclear medicine specialists.[12, 13].

The development of deep learning in the field of information processing, which is applicable in speech recognition, image recognition, and translation, has been remarkable over the past 8 years [14–16]. Deep learning uses a multi-layered algorithm to determine the settings and combinations of features by artificial intelligence itself. A large amount of supervised data can lead to favorable results without operator decision. Some authors reported that some attempts have already been made to apply deep learning for medical fields as well [17–19].

We hypothesized that the deep learning-based algorithm may be useful in the diagnosis of osseous metastases in patients with prostate cancer using bone scintigraphy. In Japan, a software for the evaluation of bone scintigraphy using a deep learning-based algorithm was recently made available [20]. The purpose of this study is to assess the software's utility in the detection of bone metastases in patients with prostate cancer, as compared with the findings of nuclear medicine specialists.

Materials and methods

Patients

Between 2009 and 2018, 139 patients with prostate cancer, who underwent bone scintigraphy in this institution, were included in the study. Their mean age was 74 years (age range 49–91 years). This study was approved by our institutional research ethics committee in accordance with the principles of the Declaration of Helsinki (approval number: 17212). Informed consent was not deemed necessary by the ethics committee because of the retrospective and noninvasive nature of this study.

Bone scintigraphy

Bone scintigraphy was performed 4 h after intravenous injection of 740 MBq (20 mCi) of technetium-99m-hydroxymethylene diphosphonate (^{99m}Tc-HMDP). ^{99m}Tc-HMDP was obtained from Nihon Medi-Physics Co. Ltd (Nishinomiya, Japan). Whole-body bone scintigraphy (anterior and posterior views, scan speed 15 cm/min, matrix size, 256 × 1024) was performed using a gamma camera system [Discovery NM 630 (GE Healthcare JAPAN, Tokyo)], equipped with low-energy high-resolution parallel hole collimators. Following these acquisitions, data processing was performed on a dedicated data processing unit [Xeleris 4.0 (GE Healthcare JAPAN, Tokyo)].

Gold standard

The gold standard was established by reviewing each patient's medical record and performing a clinical follow-up. The follow-up period was set at least 1 year. Patients with bone metastases were judged with MRI or clinical information. The sensitivity, specificity and accuracy by the specialists and with the software were evaluated.

Diagnostic supporting software

The software using a deep learning-based algorithm was provided by Nihon Medi-Physics Co., Ltd, which can be used with registration in Japan. A deep learning-based algorithm was used for skeleton segmentation and detection of hotspot extraction. The size of inputs was 512 × 1024 pixels. The proposed system employs butterfly-type networks (BtrflyNets). BtrflyNets fuses two U-Nets into a single network that can process both anterior and posterior images simultaneously. BtrflyNets for skeleton segmentation and hotspot extraction are different networks; but some parts are similar. The difference is the size of the input and output images and the number of output layers. The input for skeletal segmentation is an anterior–posterior pair of images of the whole body, while the input for hotspot extraction is a pair of anterior–posterior patch images. In skeleton segmentation, the output of the anterior skeleton consists of 13 layers and the output of the posterior skeleton consists of 12 layers. On the other hand, the output of the hotspot extraction is 3 layers of images [20]. The software has automatically analyzed bone scintigraphy for the presence or absence of bone metastasis in individual patients in 12 regions (skull, cervical vertebrae, thoracic vertebrae, lumbar vertebrae, sacrum, pelvis, rib, scapula, humerus, femur, sternum, and clavicle).

Evaluation of bone scintigraphy

Each bone scintigraphy study was evaluated visually and independently by two nuclear medicine specialists, who had at least 15 years of experience in this field. If two specialists had different opinions, the results were determined with discussion. The nuclear medicine specialists were blind to clinical information including location of any painful site. The whole-body bone scintigraphy studies were also analyzed using the software. Finally, the number of abnormal uptakes was calculated for both the nuclear medicine specialists and the software. The per-patient detection rate and the per-region detection rate were defined as follows:

(The per-patient detection rate) = (the number of patients with abnormal uptake)/(the number of all patients).

(The per-region detection rate) = (the number of regions with abnormal uptake)/(the number of all regions).

Statistical analysis

The χ^2 test or Fisher's exact test was used to compare detection rates of bone metastasis. A kappa score was also used in evaluating per-patient agreement between with software and by specialists. A value of $p < 0.05$ was considered to be statistically significant. All statistical analyses were performed using EZR (Saitama Medical Center, Jichi Medical University, Saitama, Japan), which is a graphical user interface for R (the R Foundation for Statistical Computing, Vienna, Austria). More precisely, it is a modified version of R commander designed to add statistical functions frequently used in biostatistics [21].

Results

The analyses using the software were automatically performed on all patients. Figures 1 and 2 demonstrate with and without bone metastasis, respectively. In the majority of cases, no difference was determined between the

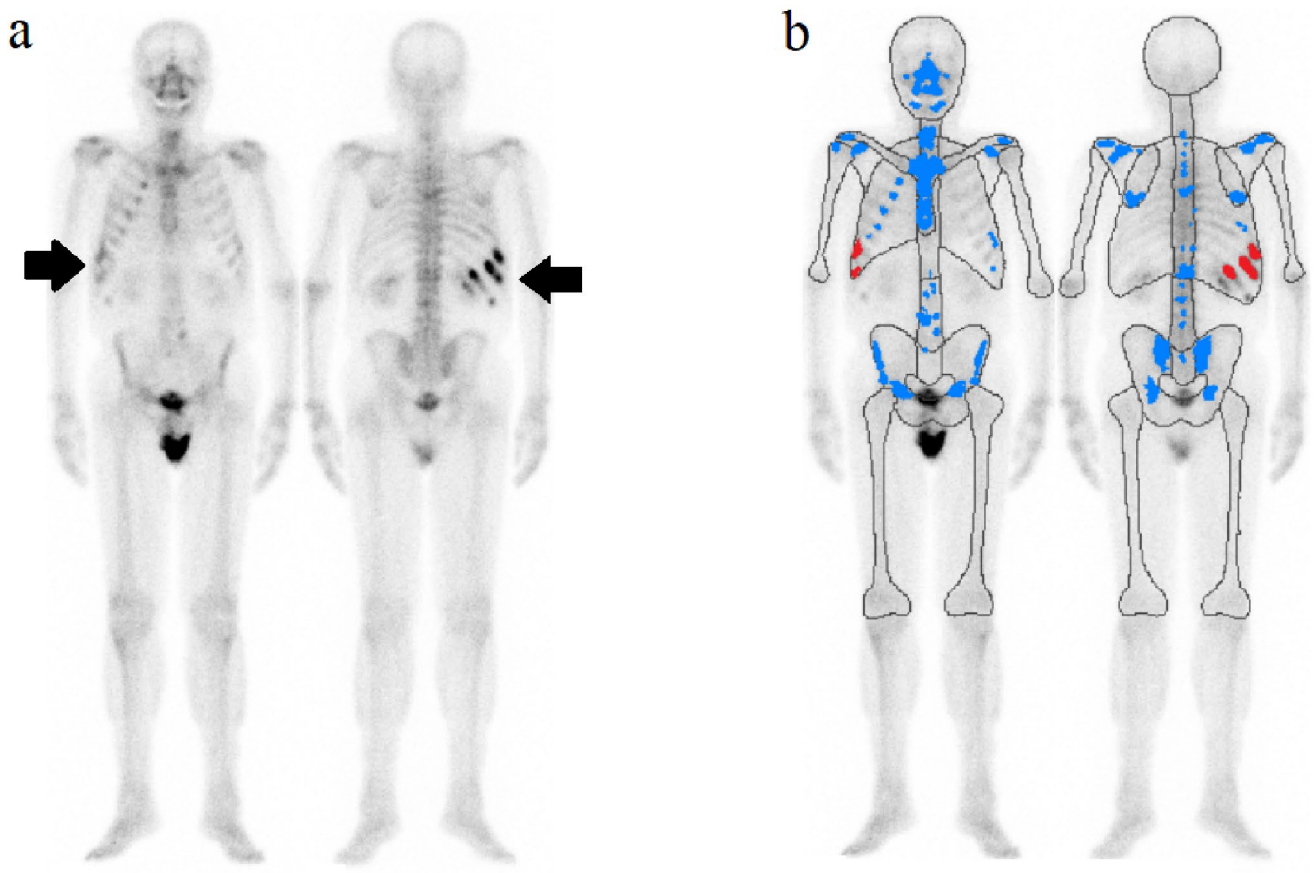

Fig. 1 **a** Shows an example of bone scintigraphy with multiple sites of bone metastases. Abnormal uptakes in the right ribs (from 9 to 12) were observed (black arrows). **b** Shows the result with the software on the same patient. Physiological uptakes and degenerative changes are indicated in blue; bone metastases are in red

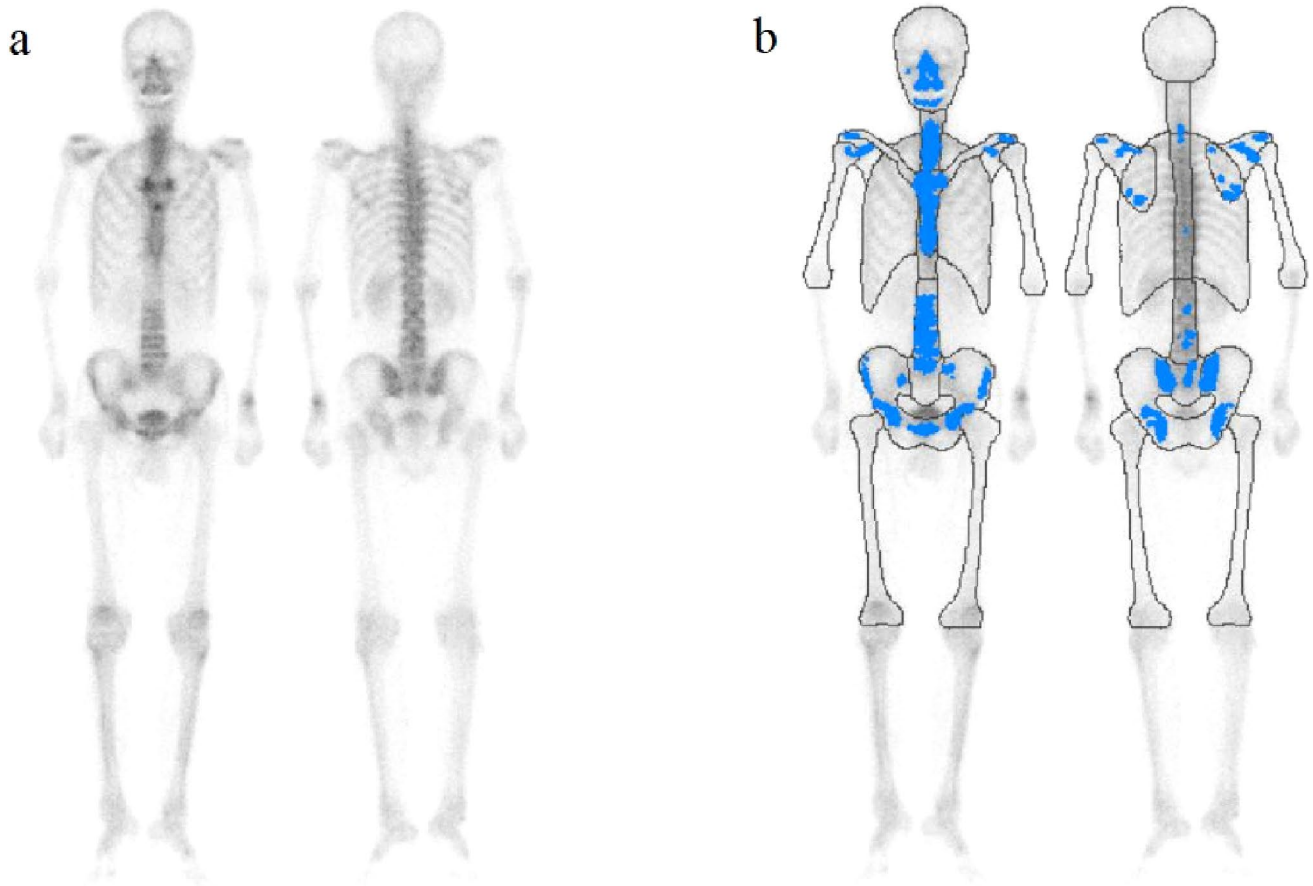

Fig. 2 **a** Shows an example of bone scintigraphy without bone metastases. **b** Also shows the result with the software on the same patient. Only physiological uptakes are indicated in blue

Table 1 Per patient detected bone metastasis

	Meta (+)	Meta (−)	Detection rate (%)	*p* value
Specialist	64	75	46.0	1
Software	65	74	46.8	

Meta (+) positive for bone metastasis, meta (−) negative for bone metastasis

interpretations made by the specialists versus the software. Table 1 shows the per-patient detection rates by the specialists and by the software. There was no statistically significant difference between the per-patient detection rates by the specialists and by the software. The kappa score on a per-patient basis was 0.812, indicating a high level of inter-rater reliability. The regional assessments made by the specialists and by the software are shown in Table 2. The per-region detection rates using the software were significantly lower than those reported by the specialists in the regions of the cervical spine and sacrum; however, there were no statistically significant differences among the other regions (10 of 12 regions). The sensitivity, specificity, and accuracy by the nuclear medicine specialists were 100% (60 of 60), 94.9% (75 of 79) and 97.1% (135 of 139). On the other hand, they with the software were 91.7% (55 of 60), 87.3% (69 of 79) and 89.2% (124 of 139).

Discussion

This study has demonstrated that the sensitivity, specificity, and accuracy with the software were slightly less than that by the specialists; however, there was no statistically significant difference in per-patient detection rate of bone metastasis between the nuclear medicine specialists and the software. The kappa score was 0.812; therefore, the software showed non-inferiority to the specialists for assessing the presence or absence of bone metastases on a per-patient basis. The per-region detection rates in the cervical spine and in the sacrum using the software were statistically significantly lower than the rates determined by the specialists.

Table 2 Per region detected bone metastasis

Segment	Evaluation	Meta (+)	Meta (−)	Detection rate (%)	*p* value
Skull	Specialist	12	127	8.6	0.825
	Software	10	129	7.2	
Cervical spine	Specialist	29	110	20.9	0.00018*
	Software	7	132	5.0	
Clavicle	Specialist	6	133	4.3	0.1042
	Software	14	125	10.1	
Scapula	Specialist	25	114	18.0	0.0832
	Software	14	125	10.1	
Humerus	Specialist	12	127	8.6	0.4862
	Software	8	131	5.8	
Rib	Specialist	46	93	33.1	0.605
	Software	41	98	29.5	
Sternum	Specialist	19	120	13.7	0.153
	Software	29	110	20.9	
Thoracic spine	Specialist	44	95	31.7	0.6
	Software	39	100	28.1	
Lumbar spine	Specialist	40	99	28.8	0.589
	Software	35	104	25.2	
Sacral spine	Specialist	33	106	23.7	0.0113*
	Software	16	123	11.5	
Pelvic bone	Specialist	55	84	39.6	0.621
	Software	50	89	36.0	
Femur	Specialist	25	114	18.0	1
	Software	25	114	18.0	

Meta (+) positive for bone metastasis, meta (−) negative for bone metastasis

*$p < 0.05$

One reason for the relatively lower per-region detection rates in the cervical spine and the sacrum might be attributed to the difficulty in assessing for the three-dimensional superimposition of bones with whole-body bone scintigraphy. For example, portions of the cervical spine might be superimposed with thyroid cartilage and sternum, and the sacrum might be superimposed with the physiological uptake of urine in the bladder. Although there was no statistically significant difference, the per-region detection rate in the scapula using the software was also lower than the rate determined by the specialists. This might be explained by the influence of the overlap of the scapula with the clavicle and/or ribs. In addition, the software's per-region detection rates in the sternum and clavicle tended to be higher as compared with those of the specialists. Alternatively, the software's per-region detection rate for the humerus and femur had almost the same results as those of the specialists. The sternum and clavicle are superimposed with vertebral bodies and the scapula, while the humerus and femur are not superimposed over any other bones. This suggests that the superimposition of bones might affect the detection rates. More precise evaluation might be facilitated using single photon emission computed tomography (SPECT), due to minimization of overlapping bones; however, SPECT analysis was not performed in this study. As SPECT requires additional acquisition time, some patients cannot tolerate a prolonged period of lying in the supine position (e.g., due to pain and non-cooperation).

Several limitations should be noted in this study. First, this study was retrospective and limited to a single institution. In principle, the software can be utilized in most circumstances, and does not depend on patient profile or conditions, as the deep learning-based algorithm was prepared using Japanese patients with prostate cancer. However, multicenter randomized trials are needed to confirm the robustness of the software. Second, the sample size was not large. Further study with a larger subject cohort should be considered to assess the practicality and benefit of the software. Third, other algorithms may produce more accurate interpretation, because deep learning-based algorithms are evolving rapidly, and more useful algorithms may be discovered in the future. Fourth, the number of bone metastases has not been evaluated. Because isolation of a bone metastasis from adjacent others was sometimes difficult with only bone scintigraphy. A more detailed study can be performed with MRI in future. Finally, no biopsy was performed in this study; however, the interpretation by the specialists might be reliable, because they had more than 15 years of experience in this field.

In conclusion, the software with a deep learning-based algorithm may be a useful diagnostic aid in the evaluation of bone metastases in patients with prostate cancer.

Funding Atsutaka Okizaki is currently receiving funding (from Bayer Yakuhin, Ltd, FUJIFILM Toyama Chemical Co., Ltd., Nihon Medi-Physics Co., Ltd).

Compliance with ethical standards

Conflict of interest For the remaining authors none were declared.

References

1. Love C, Din AS, Tomas MB, Kalapparambath TP, Palestro CJ. Radionuclide bone imaging: an illustrative review. Radiographics. 2003;23(2):341–58.
2. Van den Wyngaert T, Strobel K, Kampen WU, Kuwert T, van der Bruggen W, Mohan HK, et al. The EANM practice guidelines for bone scintigraphy. Eur J Nucl Med Mol Imaging. 2016;43(9):1723–38.
3. Norgaard M, Jensen AO, Jacobsen JB, Cetin K, Fryzek JP, Sorensen HT. Skeletal related events, bone metastasis and survival of prostate cancer: a population based cohort study in Denmark (1999–2007). J Urol. 2010;184(1):162–7.

4. Coleman RE. Clinical features of metastatic bone disease and risk of skeletal morbidity. Clin Cancer Res. 6243s;12(20 Pt 2):6243s–s62496249.
5. National Comprehensive Cancer Network. Prostate cancer version 1.2020. https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf.
6. Soloway MS, Hardeman SW, Hickey D, Raymond J, Todd B, Soloway S, et al. Stratification of patients with metastatic prostate cancer based on extent of disease on initial bone scan. Cancer. 1988;61(1):195–202.
7. Erdi YE, Humm JL, Imbriaco M, Yeung H, Larson SM. Quantitative bone metastases analysis based on image segmentation. J Nucl Med. 1997;38(9):1401–6.
8. Wakabayashi H, Nakajima K, Mizokami A, Namiki M, Inaki A, Taki J, et al. Bone scintigraphy as a new imaging biomarker: the relationship between bone scan index and bone metabolic markers in prostate cancer patients with bone metastases. Ann Nucl Med. 2013;27(9):802–7.
9. Dennis ER, Jia X, Mezheritskiy IS, Stephenson RD, Schoder H, Fox JJ, et al. Bone scan index: a quantitative treatment response biomarker for castration-resistant metastatic prostate cancer. J Clin Oncol. 2012;30(5):519–24.
10. Langsteger W, Rezaee A, Pirich C, Beheshti M. (18)F-NaF-PET/CT and (99m)Tc-MDP bone scintigraphy in the detection of bone metastases in prostate cancer. Semin Nucl Med. 2016;46(6):491–501.
11. Nakajima K, Edenbrandt L, Mizokami A. Bone scan index: a new biomarker of bone metastasis in patients with prostate cancer. Int J Urol. 2017;24(9):668–73.
12. Ulmert D, Kaboteh R, Fox JJ, Savage C, Evans MJ, Lilja H, et al. A novel automated platform for quantifying the extent of skeletal tumour involvement in prostate cancer patients using the Bone Scan Index. Eur Urol. 2012;62(1):78–84.
13. Takahashi Y, Yoshimura M, Suzuki K, Hashimoto T, Hirose H, Uchida K, et al. Assessment of bone scans in advanced prostate carcinoma using fully automated and semi-automated bone scan index methods. Ann Nucl Med. 2012;26(7):586–93.
14. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Lake Tahoe: Curran Associates Inc.; 2012.
15. Hinton G, Deng L, Yu D, Dahl G, Mohamed A-R, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. Signal Proc Mag IEEE. 2012;29:82–97.
16. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. Proc Adv Neural Inf Process Syst. 2014;27:3104–12.
17. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP, et al. Deep learning in radiology. Acad Radiol. 2018;25(11):1472–80.
18. Kather JN, Pearson AT, Halama N, Jager D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med. 2019;25(7):1054–6.
19. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med. 2018;24(10):1559–677.
20. Shimizu A, Wakabayashi H, Kanamori T, Saito A, Nishikawa K, Daisaki H, et al. Automated measurement of bone scan index from a whole-body bone scintigram. Int J Comput Assist Radiol Surg. 2020;15(3):389–400.
21. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplant. 2013;48(3):452–8.