



OPEN

Deep neural network based artificial intelligence assisted diagnosis of bone scintigraphy for cancer bone metastasis

Zhen Zhao^{1,3}, Yong Pi^{2,3}, Lisha Jiang¹, Yongzhao Xiang¹, Jianan Wei², Pei Yang¹, Wenjie Zhang¹, Xiao Zhong¹, Ke Zhou¹, Yuhao Li¹, Lin Li¹, Zhang Yi²✉ & Huawei Cai¹✉

Bone scintigraphy (BS) is one of the most frequently utilized diagnostic techniques in detecting cancer bone metastasis, and it occupies an enormous workload for nuclear medicine physicians. So, we aimed to architecture an automatic image interpreting system to assist physicians for diagnosis. We developed an artificial intelligence (AI) model based on a deep neural network with 12,222 cases of ^{99m}Tc-MDP bone scintigraphy and evaluated its diagnostic performance of bone metastasis. This AI model demonstrated considerable diagnostic performance, the areas under the curve (AUC) of receiver operating characteristic (ROC) was 0.988 for breast cancer, 0.955 for prostate cancer, 0.957 for lung cancer, and 0.971 for other cancers. Applying this AI model to a new dataset of 400 BS cases, it represented comparable performance to that of human physicians individually classifying bone metastasis. Further AI-consulted interpretation also improved human diagnostic sensitivity and accuracy. In total, this AI model performed a valuable benefit for nuclear medicine physicians in timely and accurate evaluation of cancer bone metastasis.

Advanced malignant carcinomas, such as breast cancer, prostate cancer, and lung cancer, frequently develop into bone metastasis. Thus the early detection of bone metastasis holds a valuable benefit for choosing the treatment strategy to obtain a better overall survival period and improved life quality of patients^{1,2}. Bone scintigraphy (BS) with ^{99m}Tc-MDP is one of the most commonly utilized diagnostic techniques to identify bone metastasis in cancer patients since it has merit for whole-body detection and high sensitivity³. The latest national survey has reported that more than 1.15 million bone scans were annually performed in China, which occupies a great workload for nuclear physicians. However, the limited resolution of BS images makes the interpretation is time-consuming and experience-dependent work and has the disadvantages of subjectivity, error distinctive, and unsatisfied efficiency.

Recently, the development of artificial intelligence (AI) is creeping into every facet in modern life by its advances in big-data retrieval and explicit feature evaluation, which is ideal for medical image analysis⁴⁻⁶. With the help of deep neural networks (DNNs), the computational methods allow an algorithm to program itself by learning from a large set of examples that demonstrate the desired behavior, removing the need to specify rules explicitly^{7,8}. Compared to traditional image processing methods, deep learning is more reliable and efficient, since it could automatically extract image features instead of hand-crafted features^{9,10}. By now, the AI with deep neural networks have achieved great success in the applications of medical image analysis¹¹, such as contouring of nasopharyngeal carcinoma volumes¹², retinopathy of prematurity screening¹³ and diagnosing of breast ultrasonography images¹⁴. Considering these advances, the application of DNNs-based AI system in analysis of nuclear BS image is worth pursuing.

In this study, we constructed an AI model based on a DNN with 12,222 cases of ^{99m}Tc-MDP bone scintigraphy images from patients with definite clinical conclusions. Then, the diagnostic performance and the consulting potential of AI model for improving human diagnostic accuracy and efficiency were evaluated.

¹Laboratory of Clinical Nuclear Medicine, Department of Nuclear Medicine, West China Hospital of Sichuan University, No. 37 Guo Xue Alley, Chengdu 610041, People's Republic of China. ²Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, People's Republic of China. ³These authors contributed equally: Zhen Zhao and Yong Pi. ✉email: zhangyi@scu.edu.cn; hw.cai@yahoo.com

Methods

Collection, inclusion, and exclusion of patients. This study with retrospective information collection was approved by the Institutional Ethics Committee of West China Hospital in Sichuan University. We collected 13,477 cases of BS images from patients suspected to have bone metastasis and underwent whole-body BS between January 1st, 2016, and June 30th, 2018. Then, cases with improper injection, improper imaging process, the patients who had definite primary bone tumor, and the ones did not undergo follow-up examinations were excluded.

Scanning process and diagnostic criteria of cases. Whole-body anterior and posterior views were performed using two gamma cameras (GE Discovery NM/CT 670 and Philips Precedence 16 SPECT/CT). The patient received 555 to 740 MBq of technetium-99 m methylene diphosphonate (^{99m}Tc -MDP; purchased from Syncor Pharmaceutical Co., Ltd, Chengdu, China) by intravenous injection, and the images were obtained approximately 3 h post injection. The gamma cameras were equipped with low-energy, high-resolution, parallel-hole collimators. The scan speed was 16–20 cm/min, and the matrix size was 256×1024 . The energy peak was centered at 140 keV with 15% to 20% windows.

Each BS examination contained two images of anterior and posterior views with resolutions of 256×1024 . All images collected for data set were DICOM format and interpreted for the presence or absence of bone metastasis via consensus by two nuclear medicine physicians with more than 10 years of experience. The follow-up scans were used to observe whether hot spots had disappeared, remained unchanged, or decreased or increased in size and intensity. The final clinical assessment of bone metastasis was patient-based analysis, it would be determined by clinical assessments which based on the follow-up bone scans and the other radiographic images (Such as SPECT/CT, MRI, and CT et al.)¹⁵. When no clinical or radiographic signs was found in image, data would be considered Grade 0 with no probability of bone metastasis. If the visible hot spots have confirmed fractures or degenerative changes in gathered clinical records, or the spots disappeared, remained unchanged, or decreased in size and intensity on the follow-up scan and the other radiographic images indicates these lesions leaned away from malignancy and toward the low probability of bone metastasis as Grade 1. However, cases in Grade 2 represents visible hot spots, with localization, distribution, and intensity not typical of degenerative changes or fractures, not substantially changed in scintigraphic follow-up and the radiographic modalities are equivocal, but the overall clinical judgement indicates probable bone metastasis. Grade 3 images had typical scintigraphic or radiographic patterns for bone metastasis and the patient's medical record states bone metastasis. Thus, Grade 2 and Grade 3 were identified as bone metastasis.

Cohorts and network architecture for AI model. To obtain an accurate testing results, 12,222 images were randomly assigned to three cohorts: (a) a training cohort of 9776 patients for DNNs construction, (b) a validation cohort of 1223 patients for optimization of the DNNs hyperparameters, and (c) a testing cohort of 1223 patients to test the performance of the model. As shown in Table 1, the images used in our study contained 6021 cases with lung cancer, 1844 cases with prostate cancer, 2100 cases with breast cancer, and 2257 cases with other cancers (37 kinds of cancers were listed in Supplementary Table S1).

Then, we proposed a multi-input convolutional neural network (CNN) which can accept multiple images as input. The original images (DICOM format) were resized to 256×768 and Hu matrix were normalized to [0, 1] before going to the model. Previous studies indicate that fine-tuning with pre-trained networks is an effective method for training CNNs^{16,17}. In this study, several ImageNet pretrained networks are explored and ResNet-50 has been chosen to extract high-level features from input images. Fully connection layer was removed from the final layer of ImageNet pretrained network ResNet-50 for feature extraction. The proposed network contains three parts. In the first part, ResNet-50 network was employed to extract high-level features. In the second part, max aggregation operator was used to aggregate high-level features extracted from two images. Since hotspots in the images usually present variant scales. Inspired by spatial pyramid pooling, three pooling layers with different kernel size were used to capture different scale information. In the final part, two fully connected layers were applied to classify the features into metastasis or non-metastasis. The detailed network architecture is shown in Fig. 1.

Evaluation of AI performance. Performance of the automated AI model was evaluated by the ROC analysis and AUC measurement using the testing cohort containing another 1223 cases. Total cases were divided as 4 subgroups by cancer types: prostate cancer (15.13%), breast cancer (17.17%), lung cancer (49.22%), and other cancers (18.48%), while the sensitivity, specificity, accuracy, PPV, and NPV in each cancer were calculated respectively. Gender and age related diagnostic performance was conducted to investigate whether these factors would affect the results by comparing the AUC values of male versus female, and patient's age < 60 years versus ≥ 60 years in these patients.

Then, an individual interpreting competition between AI and three nuclear physicians who had more than 5 years' experience was carried out. A new dataset containing 200 cases with cancer bone metastasis and 200 without metastasis were randomly chosen from 2786 examinations with confirmed conclusion between July and October 2018 in West China Hospital. In this competition, AI and physicians were blinded to the ground truth and distribution of patients, and interpreted images without extra radiologic and medical information, but only based on BS images. To further estimate the potential value of AI model, one hundred days later, these three physicians were required to re-interpreting the same test cohort of 400 cases, and they would give the final judgement after consulting AI's result. The time-cost, diagnostic sensitivity, specificity, accuracy, PPV, and NPV of AI system and physicians were evaluated, respectively.

Characteristic	Lung cancer	Prostate cancer	Breast cancer	Other cancers	Total
Training cohort					
Number	4817	1474	1680	1805	9776
Sex					
Male	2699	1474	10	1238	5421
Female	2118	0	1670	567	4355
Age					
Mean + SD	58.5 + 10.8	71.0 + 8.7	51.4 + 10.4	56.1 + 12.9	58.7 + 12.3
< 60 years	2400	137	1296	1054	4887
≥ 60 years	2417	1337	384	751	4889
Skeletal lesions					
No metastasis	2697	756	1066	1138	5657
Metastasis	2120	718	614	667	4119
Metastasis rates	44.01%	48.71%	36.55%	36.95%	42.13%
Validating cohort					
Number	602	185	210	226	1223
Sex					
Male	336	185	0	147	667
Female	266	0	210	79	556
Age					
Mean + SD	58.6 + 10.5	71.5 + 8.4	50.3 + 10.3	57.0 + 12.4	58.8 + 12.2
< 60 years	304	15	175	122	616
≥ 60 years	298	170	35	104	607
Skeletal lesions					
No metastasis	337	95	133	142	707
Metastasis	265	90	77	84	516
Metastasis rates	44.02%	48.65%	36.67%	37.17%	42.19%
Testing cohort					
Number	602	185	210	226	1223
Sex					
Male	357	185	0	150	692
Female	245	0	210	76	531
Age					
Mean + SD	58.5 + 10.9	71.0 + 8.1	51.5 + 10.0	56.3 + 13.4	58.8 + 12.3
< 60 years	303	14	163	139	619
≥ 60 years	299	171	47	87	604
Skeletal lesions					
No metastasis	337	95	133	142	707
Metastasis	265	90	77	84	516
Metastasis rates	44.02%	48.65%	36.67%	37.17%	42.19%

Table 1. The distribution of training, validation, and testing cohorts for AI modeling.

Statistical analysis. In this study, the comparisons of accuracy, sensitivity, specificity, PPV, and NPV between each cancer type were evaluated using the Chi-square test. All analyses were performed by using statistical software SPSS 21.0 (SPSS Inc, Chicago, IL, USA). Statistical significance was considered at the value of $P < 0.05$.

Ethics approval and consent to participate. This retrospective study was performed in accordance with the Declaration of Helsinki declaration and its later amendments or comparable ethical standards. The ethical permission for the retrospective study was obtained at the Biomedical Research Ethics Committee of West China Hospital of Sichuan University (Approval No. 2019–317), and the requirement to obtain informed consent was waived.

Results

Patient characteristics. The flow diagram of case inclusion in this study is shown in Supplementary Table S2. At the beginning, we collected 13,477 images from individual patients who were suspected to have bone metastasis and underwent whole-body BS. Then, 796 patients who had primary bone tumor were excluded, 147 cases were excluded because of poor image quality, and 312 patients were excluded for they did not undergo

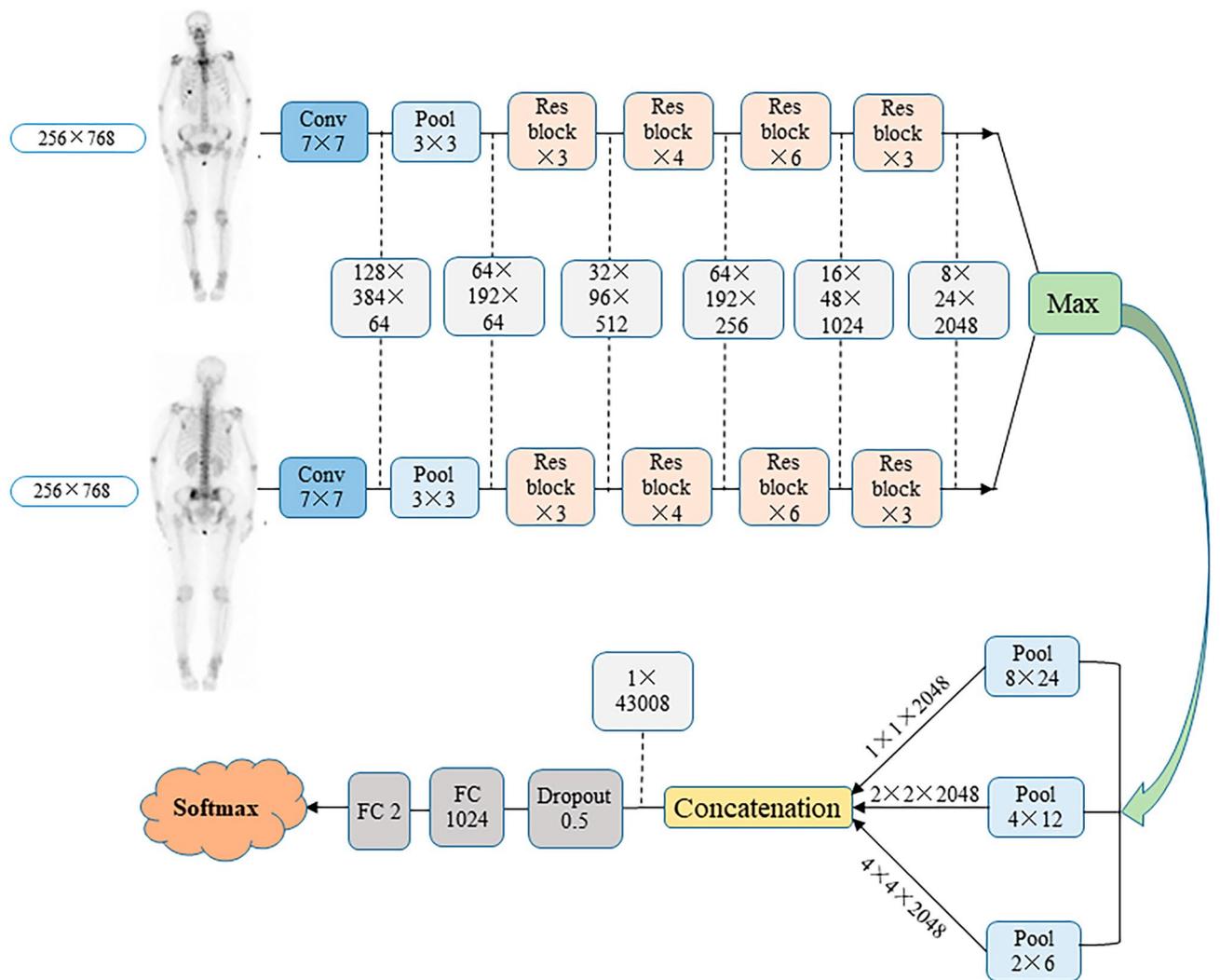


Figure 1. The architecture of convolutional neural network for AI model.

follow-up examinations. Finally, 12,222 patients (median age, 58.7 ± 12.3 years; male, 6781; female, 5441) were collected and stratified sampling for dataset, including 9776 cases for training, 1223 cases for validating, and another 1223 cases for testing.

Performance of the AI model. After training and validating process, our AI model indicated considerable diagnostic accuracy of 93.38% in cancer bone metastasis in total of 1223 testing cases, which is better than other models in previous reports (Table 2).

As shown in Fig. 2, in subgroups divided by cancer types, our AI model displayed considerable high accuracy measured by AUC value, which was 0.955 for prostate cancer, 0.988 for breast cancer, 0.957 for lung cancer, and 0.971 for the other cancers. The age-based analysis indicated no significant diagnostic differences of bone metastasis in patients with breast cancer, lung cancer, and other cancers. However, statistically different diagnostic accuracy was investigated in patients between ≥ 60 years old (AUC = 0.938) and < 60 years old (AUC = 0.992) in prostate cancer group ($P < 0.05$). A probable reason might be the older ages of patients (71.0 ± 8.1 years) than other groups ($P < 0.01$), thus the increased risk of benign diseases in aging patients, such as osteophyte, arthrosis, osteoporotic fracture, and postoperative change, also displayed hot spots in BS and thus decreased the diagnostic accuracy of bone metastases. In addition, except for sexuality-related breast cancer and prostate cancer, there were no significant differences in the diagnosis of bone metastasis between male and female patients in lung cancer and other cancer groups.

There are still 81 misdiagnosed cases were found in the testing cohort of 1223 cases (6.62%), including 38 false-negative (3.11%) and 43 false-positive (3.51%) cases (Supplementary Table S3). Lesion number, size, and adjacent diffused signal were the major influence factors in false-negative cases. On the other hand, fracture, inflammation, degenerative, and postoperative change were the main reasons for the false-positive cases in our test.

	EB	BN1	BN2	Ours
Country	Swedish	Japan	Japan	China
Training cohort	795	904	1532	9776
Bone metastases	33%	16%	42%	42%
Age	66 ± 12	64 ± 12	64 ± 12	58 ± 12
Gender				
Male	514	457	790	5421
Female	281	447	742	4355
Cancer types				
Prostate	431	267	451	1474
Breast	217	383	624	1680
Lung	/	/	/	4817
Others	147	254	457	1805
Validating cohort	NM	NM	NM	1223
Testing cohort	384	257	503	1223
Performance measured by AUC				
AUC of cancer types				
Prostate	0.939	0.949	0.957	0.955
Breast	0.847	0.91	0.924	0.988
Lung	–	–	–	0.957
Other	0.77	0.861	0.914	0.971
Total	0.858	0.91	0.932	0.964
AUC of gender				
Male	0.877	0.912	0.934	0.963
Female	0.831	0.91	0.932	0.966
AUC of age				
< 60 years	–	–	–	0.979
≥ 60 years	–	–	–	0.949

Table 2. Comparison of diagnostic performance for cancer bone metastasis by our and previous AI models. EB EXINbone¹⁸, BN1 BONENAVI version 1¹⁹, BN2 BONENAVI version 2²⁰.

Human vs. AI. The comparison of diagnostic performance of AI and human physicians were shown in Fig. 3. In the interpreting competition between AI model and three qualified nuclear medicine physicians, AI model cost only 11.3 s to complete the interpretation of 400 cases, while three physicians spent 116, 140, and 153 min, respectively, to accomplish the same work, which is corresponding to a time savings of 99.88%. Then, compared with the highest performance of three physicians, AI model manifested improved accuracy (93.5% vs. 89.00%) and sensitivity (93.5% vs. 85.00%) in calculating metastases in total cases ($P < 0.001$), but the specificity between AI model (93.50%) and human (94.50%) were not significantly different. However, after consulting the AI result, physician-1 and physician-3 indicated improved diagnostic performance, especially in finding the missed lesions and reducing the false-negative rate.

In detailed error analysis, we collected 13 cases with correct interpretation by AI but misdiagnosed by all three physicians. Among these cases, 11 patients were found to have small lesions (diameter for a few millimeters) or insufficient resolution of radioactive uptake, were ignored or judged as benign by humans (Supplementary Fig. S1). The other 2 patients who had osteoporotic vertebral compression fracture were misdiagnosed as metastases by humans (Supplementary Fig. S2). Interestingly, there were 6 cases misdiagnosed by AI but correctly interpreted by all three physicians. One patient with diffuse skeletal metastasis and two patients with humerus metastases were misdiagnosed as benign by AI (Supplementary Fig. S3). Then, one patient with multiple fractures and one patient with postoperative bone change, were misdiagnosed as malignant lesions by AI model (Supplementary Fig. S4); while the last misdiagnosed case was caused by the catheter on the patient.

Discussion

Despite the advent of various imaging modalities, such as PET/CT and multiparameter MRI, have been developed to detect skeletal metastasis, bone scintigraphies with ^{99m}Tc-MDP remains one of the most effective diagnostic techniques for its considerable sensitivity and cost performance^{21,22}. Skeletal imaging occupies 61.3% of 2.09 million of SPECT scans annually in China, and most of them were not fused with CT by the limited device utilization²³. Thus, the diagnosis of BS planar image is still a challenge for the nuclear medicine physicians in China. Fortunately, an automated system might be an effective tool to overcome this dilemma. In this study, we constructed an AI model with deep neural network based on 12,222 cases to extract image features, and evaluated its efficiency for diagnosing cancer bone metastasis with BS images. This model simultaneously improved diagnostic performance and time–cost for interpreting images, and the AI consulting system could potentially

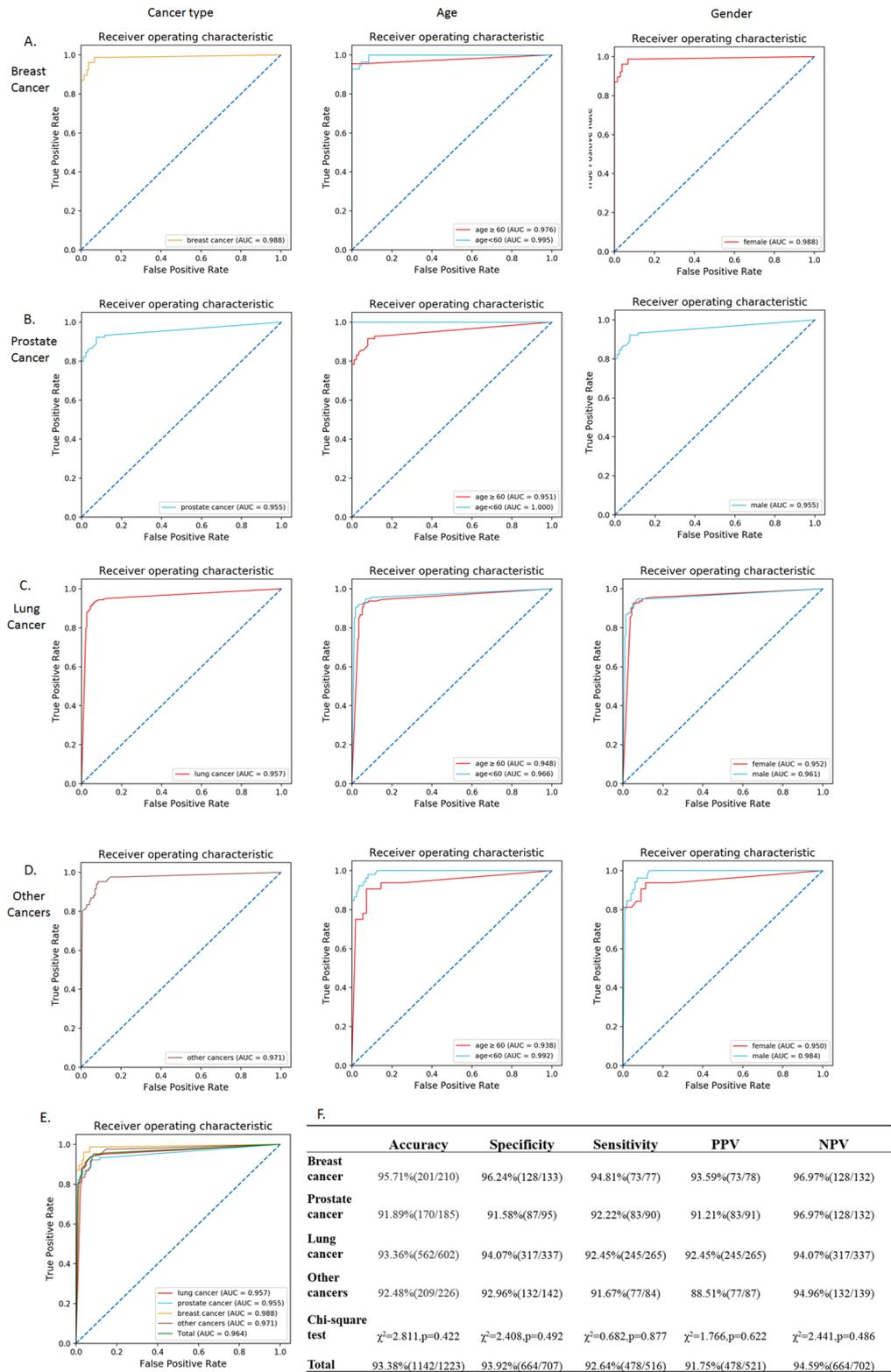


Figure 2. Diagnostic performance of AI model in BS interpretation assessed by ROC analysis for cancer types, age (B) and gender (C) factors. (A) Breast cancer; (B) prostate cancer; (C) lung cancer; (D) other cancers; E&F. Summary of total cases. AI artificial intelligence, ROC receiver operating characteristic, AUC area under the curve, PPV positive predictive value, NPV negative predictive value.

	Accuracy		Specificity		Sensitivity	
	Human diagnosis	AI consulted diagnosis	Human diagnosis	AI consulted diagnosis	Human diagnosis	AI consulted diagnosis
Physician 1	84.25%(337/400)	88.50%(354/400)	94.50%(189/200)	96.00%(192/200)	74.00%(148/200)	81.00%(162/200)
Physician 2	89.00%(356/400)	90.50%(362/400)	93.00%(186/200)	95.00%(190/200)	85.00%(170/200)	86.00%(172/200)
Physician 3	84.00%(336/400)	85.75%(343/400)	89.50%(179/200)	86.50%(173/200)	78.50%(157/200)	85.00%(170/200)
AI Model	93.50%(374/400)		93.50%(187/200)		93.50%(187/200)	

	PPV		NPV		Total time (min)	
	Human diagnosis	AI consulted diagnosis	Human diagnosis	AI consulted diagnosis	Human diagnosis	AI consulted diagnosis
Physician 1	93.08%(148/159)	95.29%(162/170)	78.42%(189/241)	83.48%(192/230)	140 min	159 min
Physician 2	92.39%(170/184)	94.51%(172/182)	86.11%(186/216)	87.16%(190/218)	116 min	119 min
Physician 3	88.20%(157/178)	86.29%(170/197)	80.63%(179/222)	85.22%(173/203)	153 min	188 min
AI Model	93.50%(187/200)		93.50%(187/200)		11.3 sec	

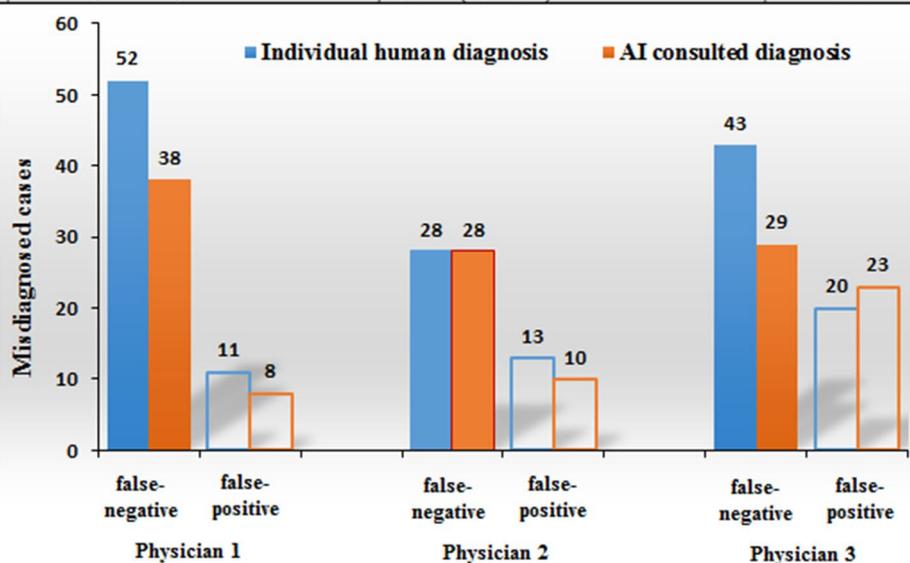


Figure 3. Diagnostic performance of individual diagnosis by three human physicians and human-AI consulted interpretation of 400 cases.

improve physicians' diagnostic skills specially for younger physicians who lacked experience. Besides, by the first time, lung cancer was separated as an individual subgroup for AI analysis and indicated diagnostic accuracy of 93.36%, which seems promising for clinical use in the future study.

Generally, deep neural networks with sufficient valid dataset is usually conducive for improving the final outcomes for AI analysis²⁴. In this study, a dataset contained 12,222 BS examinations from 40 cancer types, which is the largest dataset for single-center BS image interpreting by now, was used to construct the DNN for AI modeling. Compared with traditional methods using hand-crafted features, the use of multi-input deep convolutional neural network allows AI model to follow the natural distribution, reduced subjective judgment of physicians, better generalization performance, and closer to the usual clinical environment. For example, previous studies^{15,25} usually excluded cases that could be misleading during the training process, such as patients with large bladder, sternotomy, or fracture. However, there were not any atypical cases were excluded in our dataset to help the AI model come closest to a real index. Thus, as expected, our AI model represented improved diagnostic accuracy of AUC values (0.964) compared with other BS diagnostic AI models in previous reports (0.858, 0.91, and 0.932)^{18–20}. Notably, although the AI model have made false-negative of 8 cases in navigating small lesions in testing cohort, it displayed better capability in small lesion recognition than humans in following competition.

Although the AI model was able to efficiently improve the detection of missed small metastatic lesions by human and beneficial to reduce the readers' error rates of BS interpretation, there are several limitations should be noted. First, the estimations by our AI model were based on BS images only. The false-negative and false-positive cases have still appeared, which may be due to small lesion number, lesion size, lesion adjacent to physiological uptakes like bladder, and diffused skeletal metastasis manifested by diffused homogenous uptakes. These kinds of cases were also tricky for nuclear medicine physicians to interpret based on BS images only. However, in "real" clinical works, the patients' medical records, such as injury history, surgical record, characteristics of other

imaging modalities, and the results of laboratory tests, must be considered to obtain accurate BS interpretation. According to this, the construction of a new AI model based on the fused SPECT/CT bone images is currently undergoing by our team, and we hope the addition of fused reference CT and medical records would effectively reduce the diagnostic errors. Secondly, the unsatisfied capability in recognizing add-ons on patients, such as a catheter, is still a noticeable disadvantage of this AI model but easy for physicians. Thirdly, our study just focused on the performance of the AI model on the diagnosis of absence or presence of bone metastasis to assist nuclear medicine physicians' interpretation. However, a series previous study demonstrated that the bone scan index (BSI) calculated by artificial neural networks is an effective biomarker for predicting the prognosis or survival of some malignant cancers^{26–29}. Whether our AI model could be beneficial to the assessment of the prognosis or survival of some malignant cancers like BSI, it might require more concentration on lesion-based analysis. Last but not least, the retrospectively acquired database was collected from only one hospital for the present work. The patients at our hospital might not be considered typical of other centers, and the findings might be considered to be relatively institution-specific. A prospective multi-center study will also be needed to evaluate whether the AI model would be able to show satisfactory performance on BS images acquired with different gamma cameras, protocols, interpretive styles, and incidence of metastatic disease. These processes require considerable time for collecting more clinical data and will be studied in future works.

Conclusions

Our AI model achieved considerable time-efficiency, accuracy, specificity and sensitivity in diagnosis of bone metastasis in patients with lung cancer, prostate cancer, breast cancer, and other cancers. With further assessment and validation, this model could facilitate diagnosing programs and help physicians improve the diagnostic efficiency and accuracy of bone metastasis, particularly in remote or low-resource areas, leading to a beneficial clinical impact.

Data availability

Data confirming the results of this study are presented in the manuscript and are available from the corresponding author upon reasonable request.

Received: 15 May 2020; Accepted: 24 September 2020

Published online: 12 October 2020

References

1. Yin, J. J., Pollock, C. B. & Kelly, K. Mechanisms of cancer metastasis to the bone. *Cell Res.* **15**, 57–62. <https://doi.org/10.1038/sj.cr.7290266> (2005).
2. Kimura, T. Multidisciplinary approach for bone metastasis: a review. *Cancers* <https://doi.org/10.3390/cancers10060156> (2018).
3. Van den Wyngaert, T. *et al.* The EANM practice guidelines for bone scintigraphy. *Eur. J. Nucl. Med. Mol. Imaging* **43**, 1723–1738. <https://doi.org/10.1007/s00259-016-3415-4> (2016).
4. Dong, M., Huang, X. & Xu, B. Unsupervised speech recognition through spike-timing-dependent plasticity in a convolutional spiking neural network. *PLoS ONE* **13**, e0204596. <https://doi.org/10.1371/journal.pone.0204596> (2018).
5. Frank, D. A., Chrysochou, P., Mitkidis, P. & Ariely, D. Human decision-making biases in the moral dilemmas of autonomous vehicles. *Sci. Rep.* **9**, 13080. <https://doi.org/10.1038/s41598-019-49411-7> (2019).
6. Moravcik, M. *et al.* DeepStack: expert-level artificial intelligence in heads-up no-limit poker. *Science* **356**, 508–513. <https://doi.org/10.1126/science.aam6960> (2017).
7. Nensa, F., Demircioglu, A. & Rischpler, C. Artificial intelligence in nuclear medicine. *J. Nucl. Med.* **60**, 29s–37s. <https://doi.org/10.2967/jnumed.118.220590> (2019).
8. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H. & Aerts, H. Artificial intelligence in radiology. *Nat. Rev. Cancer* **18**, 500–510. <https://doi.org/10.1038/s41568-018-0016-5> (2018).
9. Kermany, D. S. *et al.* Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **172**, 1122–1131. <https://doi.org/10.1016/j.cell.2018.02.010> (2018).
10. Visvikis, D., Cheze Le Rest, C., Jaouen, V. & Hatt, M. Artificial intelligence, machine (deep) learning and radio(geno)mics: definitions and nuclear medicine imaging applications. *Eur. J. Nucl. Med. Mol. Imaging* <https://doi.org/10.1007/s00259-019-04373-w> (2019).
11. Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88. <https://doi.org/10.1016/j.media.2017.07.005> (2017).
12. Lin, L. *et al.* Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma. *Radiology* **291**, 677–686. <https://doi.org/10.1148/radiol.2019182012> (2019).
13. Wang, J. *et al.* Automated retinopathy of prematurity screening using deep neural networks. *EBio Med.* **35**, 361–368. <https://doi.org/10.1016/j.ebiom.2018.08.033> (2018).
14. Qi, X. *et al.* Automated diagnosis of breast ultrasonography images using deep neural networks. *Med. Image Anal.* **52**, 185–198. <https://doi.org/10.1016/j.media.2018.12.006> (2019).
15. Sadik, M., Suurkula, M., Höglund, P., Järund, A. & Edenbrandt, L. Improved classifications of planar whole-body bone scans using a computer-assisted diagnosis system: a multicenter, multiple-reader, multiple-case study. *J. Nucl. Med.* **50**, 368–375. <https://doi.org/10.2967/jnumed.108.058883> (2009).
16. Wang, C. J. *et al.* Deep learning for liver tumor diagnosis part II: convolutional neural network interpretation using radiologic imaging features. *Eur. Radiol.* **29**, 3348–3357. <https://doi.org/10.1007/s00330-019-06214-8> (2019).
17. Shen, X., Zhang, J., Yan, C. & Zhou, H. An automatic diagnosis method of facial acne vulgaris based on convolutional neural network. *Sci. Rep.* **8**, 5839. <https://doi.org/10.1038/s41598-018-24204-6> (2018).
18. Ulmert, D. *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.* **62**, 78–84. <https://doi.org/10.1016/j.eururo.2012.01.037> (2012).
19. Horikoshi, H., Kikuchi, A., Onoguchi, M., Sjöstrand, K. & Edenbrandt, L. Computer-aided diagnosis system for bone scintigrams from Japanese patients: importance of training database. *Ann. Nucl. Med.* **26**, 622–626. <https://doi.org/10.1007/s12149-012-0620-5> (2012).
20. Nakajima, K. *et al.* Enhanced diagnostic accuracy for quantitative bone scan using an artificial neural network system: a Japanese multi-center database project. *EJNMMI Res.* **3**, 83. <https://doi.org/10.1186/2191-219X-3-83> (2013).

21. Davila, D., Antoniou, A. & Chaudhry, M. A. Evaluation of osseous metastasis in bone scintigraphy. *Semin. Nucl. Med.* **45**, 3–15. <https://doi.org/10.1053/j.semnuclmed.2014.07.004> (2015).
22. Urano, M. *et al.* Diagnostic utility of a computer-aided diagnosis system for whole-body bone scintigraphy to detect bone metastasis in breast cancer patients. *Ann. Nucl. Med.* **31**, 40–45. <https://doi.org/10.1007/s12149-016-1132-5> (2017).
23. Wang, J. & Li, S. A brief report on the results of the national survey of nuclear medicine in 2018. *Chin. J. Nucl. Med. Mol. Imaging* **38**, 2–4 (2018).
24. Porenta, G. Is there value for artificial intelligence applications in molecular imaging and nuclear medicine?. *J. Nucl. Med.* **60**, 1347–1349. <https://doi.org/10.2967/jnumed.119.227702> (2019).
25. Sadik, M. *et al.* Computer-assisted interpretation of planar whole-body bone scans. *J. Nucl. Med.* **49**, 1958–1965. <https://doi.org/10.2967/jnumed.108.055061> (2008).
26. Inaki, A., Nakajima, K., Wakabayashi, H., Mochizuki, T. & Kinuya, S. Fully automated analysis for bone scintigraphy with artificial neural network: usefulness of bone scan index (BSI) in breast cancer. *Ann. Nucl. Med.* **33**, 755–765. <https://doi.org/10.1007/s12149-019-01386-1> (2019).
27. Nakajima, K. *et al.* Role of bone scan index in the prognosis and effects of therapy on prostate cancer with bone metastasis: Study design and rationale for the multicenter prostatic cancer registry of standard hormonal and chemotherapy using bone scan index (PROSTAT-BSI) study. *Int. J. Urol.* **25**, 492–499. <https://doi.org/10.1111/iju.13556> (2018).
28. Anand, A. *et al.* Analytic validation of the automated bone scan index as an imaging biomarker to standardize quantitative changes in bone scans of patients with metastatic prostate cancer. *J. Nucl. Med.* **57**, 41–45. <https://doi.org/10.2967/jnumed.115.160085> (2016).
29. Kalderstam, J., Sadik, M., Edenbrandt, L. & Ohlsson, M. Analysis of regional bone scan index measurements for the survival of patients with prostate cancer. *BMC Med. Imaging* **14**, 24. <https://doi.org/10.1186/1471-2342-14-24> (2014).

Acknowledgements

This project was financially supported by the National Major Science and Technology Projects of China (2018AAA0100201), the Sichuan Provincial Science and Technology Project of the Health Planning (19PJ79), the Sichuan Science and Technology Program of China (2020JDR0042), and “1.3.5 project for disciplines of excellence in West China Hospital (ZYG18016)”.

Author contributions

All authors participated in the interpretation of the data and in the drafting of the manuscript. All authors read and approved the final version of the manuscript. Z.Y., C.H., Z.Z., and P.Y. participated in the design of the retrospective study. Z.Z., P.Y., W.J., Z.Y. and C.H. participated in the construction of AI model. C.H., Z.Z., Y.P., Z.K., and L.L. carried out the clinical data collection and statistical analysis. J.L., X.Y., W.J. and L.Y. carried out the nuclear medicine data collection for the BSI analysis. Z.Z., Z.W., and Z.X. participated in the human vs. AI competition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-74135-4>.

Correspondence and requests for materials should be addressed to Z.Y. or H.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020