Medical Image Analysis 65 (2020) 101784

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Automated diagnosis of bone metastasis based on multi-view bone scans using attention-augmented deep neural networks

Yong Pi[a,1], Zhen Zhao[b,1], Yongzhao Xiang[b], Yuhao Li[b], Huawei Cai[b,*], Zhang Yi[a,*]

[a] *Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, PR China*
[b] *Department of Nuclear Medicine, West China Hospital of Sichuan University, Chengdu 610041, PR China*

ARTICLE INFO

Article history:
Received 8 August 2019
Revised 9 January 2020
Accepted 15 July 2020
Available online 18 July 2020

Keywords:
Bone metastasis
Deep neural networks
Multi-view
Whole body bone scan

ABSTRACT

Bone scintigraphy is accepted as an effective diagnostic tool for whole-body examination of bone metastasis. However, the manual analysis of bone scintigraphy images requires extensive experience and is exhausting and time-consuming. An automated diagnosis system for such images is therefore much desired. Although automatic or semi-automatic methods for the diagnosis of bone scintigraphy images have been widely studied, they employ various steps to classify the images, including segmentation of the entire skeleton, detection of hot spots, and feature extraction, which are complex and inadequately validated on small datasets, thereby resulting in low accuracy and reliability. In this paper, we describe the development of a deep convolutional neural network to determine the absence or presence of bone metastasis. This model consisting of three sub-networks that aim to extract, aggregate, and classify high-level features in a data-driven manner. There are two main innovations behind this method; First, the diagnosis is performed by jointly analyzing both anterior and posterior views, which leads to high accuracy. Second, a spatial attention feature aggregation operator is proposed to enhance the spatial location information. A large annotated bone scintigraphy image dataset containing 15,474 examinations from 13,811 patients was constructed to train and evaluate the model. The proposed method is compared with three human experts. The high classification accuracy achieved demonstrates the effectiveness of the proposed architecture for the diagnosis of bone scintigraphy images, and that it can be applied as a clinical decision support tool.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Bone metastasis occurs when a primary tumor invades bone. Patients with the most common types of solid tumors (breast, lung, thyroid, kidney, and prostate) frequently develop bone metastasis, and the extent of the disease significantly affect survival time, morbidity, and quality of life (Mundy, 1997). When tumors have metastasized to bone, patients usually experience sequential skeletal complications, including skeletal remodeling, fractures, pain, and anemia. The choice of treatment strategy for solid tumors is partly determined by whether the presence of bone metastasis, with bisphosphonates having been shown to be efficient in inhibiting the development of bone metastasis (Coleman, 2001). Therefore, the early detection of bone metastasis is crucial for increasing survivability and the choice of appropriate treatment strategy.

Nowadays, various imaging modalities including radiography, magnetic resonance imaging (MRI), and computed tomography (CT) are available for the diagnosis of bone metastasis (Zhao et al., 2009). However, the whole-body bone scan (WBS) is still widely accepted as the standard method for surveying the existence and extent of bone metastasis since it is much cheaper than MRI while has similar performance with it for the differential diagnosis of bone metastasis (Wu et al., 2013).

The abnormalities in WBS images are called hot spots and generally appear as higher intensity signal than the surroundings; they are the key factors for diagnosing bone metastasis. Although WBS is effective for diagnosing bone metastasis, the analysis of the images is still a difficult and subjective task that requires extensive experience. This is because patients without bone metastasis can also present hot spots on WBS images. Various non-neoplastic diseases such as osteomyelitis, arthropathies, and fractures can also show abnormal characterizations on WBS images, and a great number of possible errors and different diagnoses should be considered (Bombardieri et al., 2003). Fig. 1 shows WBS images from two patients, with the images from both patients containing hot

[*] Corresponding authors.
E-mail addresses: hw.cai@yahoo.com (H. Cai), zhangyi@scu.edu.cn (Z. Yi).
[1] Co-first authors.

https://doi.org/10.1016/j.media.2020.101784
1361-8415/© 2020 Elsevier B.V. All rights reserved.

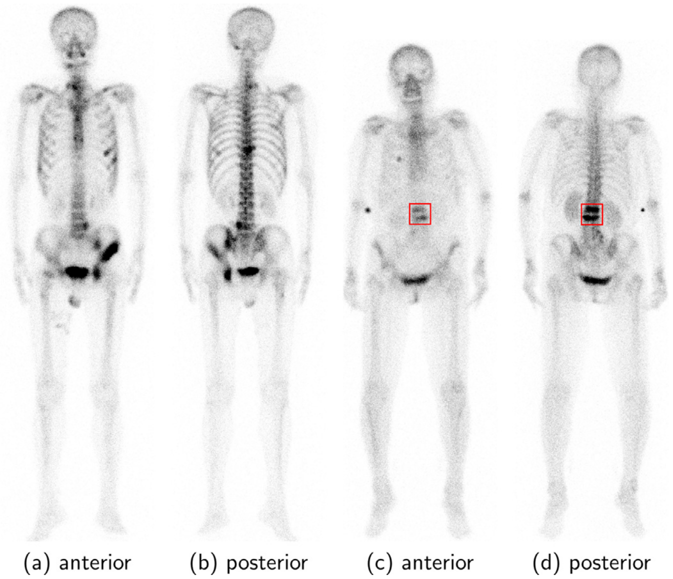


Fig. 1. WBS images of two patients. All four images contain hot spots (areas of high intensity) which are the crucial areas for diagnosing bone metastasis. The first two images are the anterior and posterior view of a gastric cancer patient with bone metastasis. The next two images belong to a lung cancer patient without bone metastasis, the two abnormal points (red rectangle) are the sign of compressive fracture instead of bone metastasis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

spots. The first two images are the anterior and posterior views of a gastric cancer patient with bone metastasis, while the next two images that belong to a patient without bone metastasis still contain two abnormal points that are the sign of compressive fracture instead of bone metastasis. Manual analysis of WBS images by physicians is subjective, time-consuming, and exhausting, and the development of an automated bone scan imaging analysis model is most desirable.

Automated bone scan analysis would be helpful for improving the reliability and efficiency of bone metastasis diagnosis. Previous computer-aided diagnosis systems for assisting physicians have been proposed (Sadik et al., 2006; 2008); however, most of these focused on the detection of hot spots, and then classified them using manually-defined features. In this study, we developed a new methodology that recognizes the two-view WBS examination in a fully automated manner, without manual preparation of features and detection of hot spots. The development of an automated diagnostic method for bone metastasis faces several obstacles, which include the following. (1) Various non-neoplastic diseases also present abnormalities on imaging findings, leading to high sensitivity and low specificity. (2) To achieve a strong generalization ability under different scenarios, a large annotated dataset is required to learn the features of bone metastasis from the data. However, none of the datasets used in previous bone scan-related research can satisfy this requirement. (3) Each WBS examination contains two images showing the anterior and posterior views. To analyze the existence of bone metastasis, the model must jointly analyze both views as a single examination.

We address the above stated challenges through the following methods. (1) We use deep convolutional neural networks (CNNs) to automatically extract high-level features from the data. Recently, CNNs have achieved great success in computer vision-related problems, and have been shown to effectively learn high-level features directly from data. Transfer learning has also been used to facilitate the training phase. (2) We construct a large-scale dataset of WBS images annotated by professional nuclear medicine physicians. This large dataset contributes to reducing overfitting and helps the model learn the key differences between the presence or absence of bone metastasis. (3) A novel architecture that receives multiple inputs is developed to jointly analyze images from anterior and posterior views.

This study makes the following contributions to the field.

1. A large-scale WBS image dataset containing 15,474 examinations labeled by professional nuclear medicine physicians was constructed for the automated analysis of bone scan images. The bone scan images in this dataset were derived from patients with a variety of metastatic cancers including breast cancer, lung cancer, and prostate cancer. The data follow a natural distribution, which helps with generalization performance. To the best of our knowledge, our dataset is an order of magnitude larger than any similar dataset used in previous bone scan-related research. The data availability is described in Section 3.3.
2. An automated bone metastasis diagnosis model based on multi-view images is proposed. This model contains three parts that aim to extract, aggregate, and classify high-level features in a data-driven manner.
3. A feature aggregation operator parameterized by a deep neural network is proposed to bind the features from the anterior and posterior views of an examination. CNNs using such an operator typically show better performance than other state-of-the-art feature aggregation operators.
4. The classification and visualization results indicate that the developed method successfully learned the characteristics of bone metastasis on WBS images. Moreover, the proposed method achieved results comparable to those of three experienced physicians in diagnosing the existence of bone metastasis, revealing that it can be applied as a useful clinical decision support tool.

2. Related work

The application of computer-aided diagnosis systems (CAD) to WBS images has been a subject of study for decades (Erdi et al., 1997). In this section, we present an overview of previous studies on the analysis of WBS images and multi-view fusion.

2.1. Traditional methods for bone scan analyzing

The existing publications on the analysis of WBS images focus mainly on three areas: automated calculation of the Bone Scan Index (BSI), automated diagnosis of bone metastasis, and segmentation of hot spots. The BSI was developed by researchers at Memorial Sloan-Kettering Cancer Center (New York, USA) during 1977 as a promising means of measuring the extent of bone metastasis (Erdi et al., 1997). However, calculation of the BSI is only semiautomatic, and requires a laborious manual process. Sadik et al. (2006) then developed an automated method for diagnosing the existence of bone metastasis that was based on image processing techniques and artificial neural networks. The proposed method contaid four steps: first, a customized algorithm was used for segmentation of the head, chest, spine, pelvis, and bladder; second, hot spots were detected using a region-specific threshold algorithm; third, fourteen hand-designed features were automatically extracted to describe the anterior and posterior images; and finally, the fourteen features were used as the input to standard multi-layer perceptrons (MLP) that were trained to classify the metastases. A total of 200 patients with a diagnosis of breast or prostate cancer were included in their dataset. The sensitivity was 90% and the specificity 74%. Sadik et al. (2008) further improved their previous method by improving the image processing techniques and increasing the size of the database. The training process was then performed on a dataset containing 810 patients, and the system showed a sensitivity of 90% and specificity of 89% on a test group of 59 patients.

The segmentation of hot spots from bone scans has also been an active area. For example, Yin and Chiu (2004) proposed a fuzzy system called the ”characteristic-point-based fuzzy inference system” (CPFIS), which chose asymmetry and brightness as the inputs to perform hot spot segmentation. The sensitivity of their system was 91.5% (227/248), and the mean number of false positives was 37.3 per image. Huang et al. (2007) used a fuzzy sets histogram thresholding method and an anatomical knowledge-based image segmentation method to find a regional threshold for extracting hot spots. The overall sensitivity of their method was 92.1% (222/241), with 7.58 false detections per image. Recently, Geng et al. (2016) proposed an interactive approach to detect and extract hot spots in the thoracic region, which was based on a new multiple instance learning (MIL) method. The F_1-score attained was 77.1% on a dataset consisting of 100 hot spots from 46 images.

Although traditional methods have made great achievements in aiding the diagnosis of bone metastasis on WBS images, the drawbacks are obvious. The classification performance of these methods relies heavily on hot spot segmentation, meaning that errors in segmentation may cause failures in subsequent classification. Furthermore, traditional methods lack robustness because of their dependency on hand-crafted features and threshold values. Moreover, the manual choosing of features is exhausting and subjective, and it is difficult to improve the performance of these approaches.

2.2. Deep neural networks for bone scan analysis

In recent years, since AlexNet (Krizhevsky et al., 2012) won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015) in 2012, deep convolutional neural networks (CNNs) have gained widespread popularity in computer vision. Since then, many important breakthroughs in computer vision tasks have been achieved following the introduction of a large number of novel CNN architectures (Szegedy et al., 2016; He et al., 2016; Huang et al., 2017; Hu et al., 2018). Deep CNNs have a number of advantages over traditional image processing methods: they automatically extract features at different levels in a data-driven manner, and there is no need for hand-crafted features, reducing the workload of physicians. Numerous studies have explored deep neural networks in a range of medical image analysis applications, including the grading of diabetic retinopathy (Zhang et al., 2019), diagnosis of breast cancer (Qi et al., 2019), and segmentation of macular edema on optical coherence tomography (Hu et al., 2019).

However, to the best of our knowledge, there are currently few studies using deep neural networks for the automated analysis of bone metastasis on WBS images. Belcher (2017) employed CNNs to classify hot spots into benign or malignant ones, reaching an accuracy of 89% on a test set. These hot spots were hand extracted from 2164 patients with prostate cancer, and 10,428 hot spots coming from the lower spine were included, as these hot spots were considered the easiest to classify. Geng et al. (2015) used a sparse autoencoder and CNNs to train an image-level classifier that classified input images into normal or suspect, achieving an accuracy of 95%, then a patch-level classifier was trained to produce a probability map of the hot spots. Finally, level set segmentation was performed on the probability map to segment hot spots. The testing dataset contained 68 suspect thoracic images containing 572 hot spots, and showed a Jaccard index of 0.8051.

Although some studies have been conducted on the automated diagnosis of bone metastasis, most of these studies performed the diagnosis on hot spots that first required segmentation from the WBS images, a process that may introduce additional errors. Moreover, these studies were based on small datasets for patients with specific cancers. In this study, a large dataset containing various cancer patients was constructed. Using this dataset, a novel architecture was developed to perform diagnosis directly on the whole WBS image, thereby rendering a robust automated diagnosis model to support clinical decision making.

2.3. Multi-view fusion

Deep learning with multi-view strategy has been gaining significant interest in recent years, especially in the medical image analysis field. Image classification tasks with natural images usually contain only one image at a time, in contrast an examination in medical imaging often comes with a set of views. For instance, screening mammogram provides two different views for each breast of a patient, whole body bone scan acquiring anterior and posterior views in an examination.

There have been lots of studies on building deep neural networks for multi-view medical image analysis, especially for automated analyzing mammography images (Jouirou et al., 2019). Carneiro et al. (2017) proposed a deep neural network to estimate the patients risk of developing breast cancer. The model uses cranio-caudal and medio-lateral oblique mammography views as inputs, getting an AUC value over 0.9 for a 3-class problem. Khan et al. (2019) proposed a multi-view feature fusion based CAD system using feature fusion technique of four views for classification of mammogram. Wang et al. (2017) developed a multi-view CNN for lung nodule segmentation from axial, coronal and sagittal views in CT images. Liu et al. (2018) presented a multi-view CNN for lung nodule type classification from CT images. Motivated by those studies, we developed an automated bone metastasis diagnosis framework using multi-view inputs. More details about the proposed method were depicted in Sections 4 and 5.

3. Dataset

To develop an automated bone metastasis diagnosis system with high performance, we constructed a large dataset of WBS images. In this section, we describe this dataset in detail, followed by the progress of the data annotation. After data annotation, the labeled examinations were partitioned into a training set, validation set, and testing set.

3.1. Materials

All of the WBS images used in this study were obtained from the Department of Nuclear Medicine of West China Hospital, Sichuan University. The data collection process proceeded as follows: first, all the digital clinical reports from Jan 2018 to Mar 2019 were collected, and sensitive patient information was erased, with the reports being identified by a check number. A total of 16,341 records were collected. Then, corresponding WBS images were exported according to the check number. Finally, a total of 16 211 examinations and their corresponding images were collected.

The collected WBS images were acquired using a gamma camera approximately three hours after an intravenous injection of ^{99m}Tc-methylene diphosphonate (25 mCi). Each WBS contains two images, the anterior and posterior views, and the original data were stored in Digital Imaging and Communications in Medicine (DICOM) format. All examinations were taken using one of two types of devices, a GE-Discovery NM/CT670 (6176 examinations) with a resolution 256 × 1024 pixels and a Philips Medical Systems scanner(9298 examinations) with a resolution 512 × 1024 pixels.

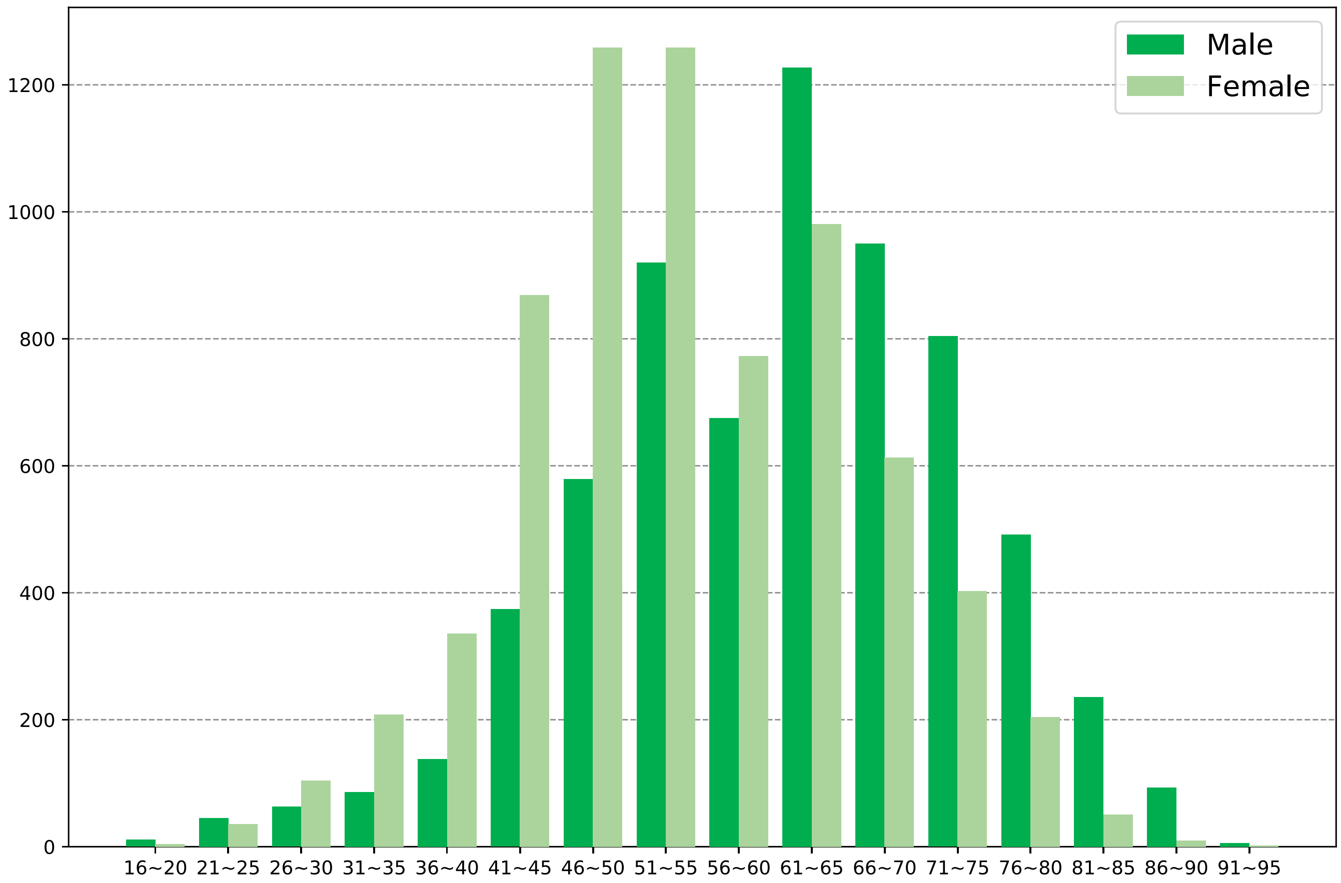


Fig. 2. Age distribution of 13,811 patients in our dataset.

Table 1
The clinical definition of labels.

Label	Clinical Descriptions
Malignant	Bone metastasis detected
Benign	No Bone metastasis detected

3.2. Data annotation

The WBS images were labelled as malignant or benign. The clinical definitions of the labels are shown in Table 1. The gold standard classification label for each examination in the dataset was based on the diagnosis in the corresponding clinical reports. These reports were written by a junior physician and were reevaluated by a senior physician based on the bone scan images and clinical data such as localization of bone pain, medical condition and previous history of trauma. If there was no clear diagnosis in the clinical report, the examination was excluded. Finally, 15,474 annotated examinations from 13,811 patients, including 9595 benign diagnoses and 5879 malignant cases were included in the dataset. The dataset contained 6699 male patients (mean age 61.25 ± 12.58) and 7112 female patients (mean age 54.43 ± 11.58); the age distribution of the patients is depicted in Fig. 2. Tables 2 and 3 list the types and incidences of the primary lesions in the dataset. Differing from previous studies (Sadik et al., 2006; 2008) that manually excluded misleading examples, the dataset constructed in this study followed a real-world distribution without excluding any cases, under the premise that the system trained on this dataset would be more suitable for routine clinical application.

Table 2
Types and incidences of primary tumors among 5879 bone metastasis examinations.

Type	Number	Type	Number
Lung cancer	2475	Pancreatic cancer	22
Prostate cancer	1164	Thyroid cancer	22
Breast cancer	1281	Lymphoma	20
Colorectal-cancer	135	Ureteral carcinoma	15
Nasopharyngeal-carcinoma	150	Laryngeal carcinoma	15
Liver cancer	140	Malignant melanoma	15
Gastric cancer	90	Endometrial cancer	15
Renal cancer	86	Biliary carcinoma	15
Bladder cancer	69	Parotid carcinoma	14
Esophageal carcinoma	58	Ovarian carcinoma	14
Mediastinal malignant tumor	36	Total	5879
Uterine cervix cancer	28		

3.3. Data partition

To give accurate testing results, the dataset was split into three subsets: a training set, validation set, and testing set. Examinations acquired using the two different devices were uniformly distributed across the subsets. Considering that images from the same patient would show similarities, examinations from the same patient were not divided into the different subsets. Finally, 12,274, 1600, and 1600 samples were used for the training, validation and testing sets respectively. The distribution of each subset is shown in Table 4.

Data availability. Currently, anyone can get the validation subset by emailing the corresponding author by stating that the data

Table 3
Types and incidences of primary lesions among the 9595 benign examinations. The "Other Cancers" group contained a relatively small number of diseases, including tongue carcinoma, endometrial cancer, and parotid carcinoma. The "Other Benign lesions" group contained various diseases, including fracture, hemoptysis, and bone pain.

Type	Number	Type	Number
Breast cancer	2478	Thyroid cancer	12
Lung cancer	2347	Benign lung lesions	2051
Prostate cancer	382	Benign prostate lesions	69
Nasopharyngeal-carcinoma	244	Benign breast lesions	47
Esophageal carcinoma	134	Benign liver lesions	30
Colorectal-cancer	132	Benign brain lesions	26
Liver cancer	111	Benign renal lesions	25
Gastric cancer	51	Benign thyroid lesions	16
Renal cancer	37	Benign gastric lesion	12
Mediastinal malignant tumor	44	Other Cancers	154
Bladder cancer	20	Other Benign lesions	1160
Ovarian cancer	13	Total	9595

Table 4
Data distribution in training, validation and testing set.

Subsets	Philips		GE		Total	
	Benign	Malignant	Benign	Malignant	Benign	Malignant
Training	4310	3070	3285	1609	7595	4679
Validation	564	395	436	205	1000	600
Testing	564	395	436	205	1000	600
Total	5438	3860	4157	2019	9595	5879

is used for research purpose only. The whole dataset will be publicly available in the future.

4. Methods

In this section, the proposed architecture for automated diagnosis of WBS images is first illustrated in detail, and is then followed by the image preprocessing method.

4.1. Overall architecture

This work aimed to automatically diagnose bone metastasis by analysis of WBS images. Differing from many existing image classification tasks where a single image is used as a sample, a single sample in our dataset contains two images: a posterior view image and an anterior view image. Generally, a dataset in which a single sample contains multiple images can be formed as $D = \{(X_i, Y_i); i = 1, 2, \dots, N\}$, $X_i = \{x_{i,1}, \dots, x_{i,j}, \dots, x_{i,J}\}$. Here, X_i denotes the ith sample in the dataset that contains J images, and $x_{i,j}$ is the jth image in the ith sample, and Y_i is the corresponding label of the sample X_i. For the dataset used in this study, $J = 2$, $x_{i,1}$ is the posterior view image, $x_{i,2}$ denotes the anterior view image, and $Y_i \in \{malignant,\ benign\}$.

The goal of this study was to learn a robust transform function f: $X_i \rightarrow Y_i$. In this work, we parameterized the transform f using deep neural networks that are well known for their strong nonlinearity. As the inputs to the network are J images instead of a single image, we developed a J-way input network. This network consists of three parts. In the first part, a deep neural network N_{ex} is employed to extract features from the input J images, and it can be formulated as $F_{i,j} = N_{ex}(x_{i,j})$. Here, $x_{i,j}$ is the jth image in the ith examination, and $F_{i,j}$ denotes the high level features of $x_{i,j}$ extracted by network N_{ex}. In the second part of the network, high level features of the ith sample are fused using a feature fusion operator N_{fu}, $S_i = N_{fu}(F_{i,1}, \dots, F_{i,j}, \dots, F_{i,J})$. Here, S_i denotes the fused feature. In the final part, a customized classification neural network N_{cl} is employed to output a prediction of the true label, with $P_i = N_{cl}(S_i)$, P_i being the model output after applying a softmax function corresponding to the $i$the sample. The proposed architecture is trained with a backpropagation algorithm by minimizing a cross-entropy cost function over the training set, defined as:

$$\mathcal{L} = -\sum_{i=1}^{N} Y_i^T \ln(P_i). \tag{1}$$

The overall architecture of the proposed network is depicted in Fig. 3, and the detail of each part is presented in the following subsection.

4.2. Part one: feature extraction network

In the first part of the proposed networks, a CNN is employed to automatically extract high-level features from the input images. CNNs have shown great success in the computer vision field, benefitting from their powerful feature extraction abilities. Generally, a deep CNN is a feedforward network with a stack including three types of layers: convolutional (LeCun et al., 1995), pooling (Lin et al., 2013), and fully-connected layers. The convolutional layers are the main component of CNNs, consisting of several filters to extract the stationary local attributes of the inputs. The operation in the convolution layer can be formulated as:

$$\begin{cases} z_{i,j}^{l+1} = \sum_{p=0}^{P_l-1} \sum_{q=0}^{Q_l-1} \sum_{k=0}^{K_l-1} w_{p,q,k}^{l} \cdot a_{i+p,j+q,k}^{l}, \\ a_{i,j}^{l+1} = f(z_{i,j}^{l+1}), \end{cases} \tag{2}$$

Here $w_{p,q,k}^{l}$ are the parameters of the convolution filter in the lth convolutional layer, with P, Q and K denoting the width, height and channel of the filter respectively. i and j define the spatial location of one element in the feature map z^{l+1}, f is a non-linear activation function, which is usually a rectified linear unit (ReLU) function.

Recently, transfer learning has achieved great success in image recognition tasks, and the fine-tuning with pretrained CNNs has become a natural choice for classification task (Shin et al., 2016). In this study, the CNNs were first trained on ImageNet, and the learned parameters $\tilde{\theta}$ were then used to initialize the parameter θ of the first part of our network. Three state-of-the-art

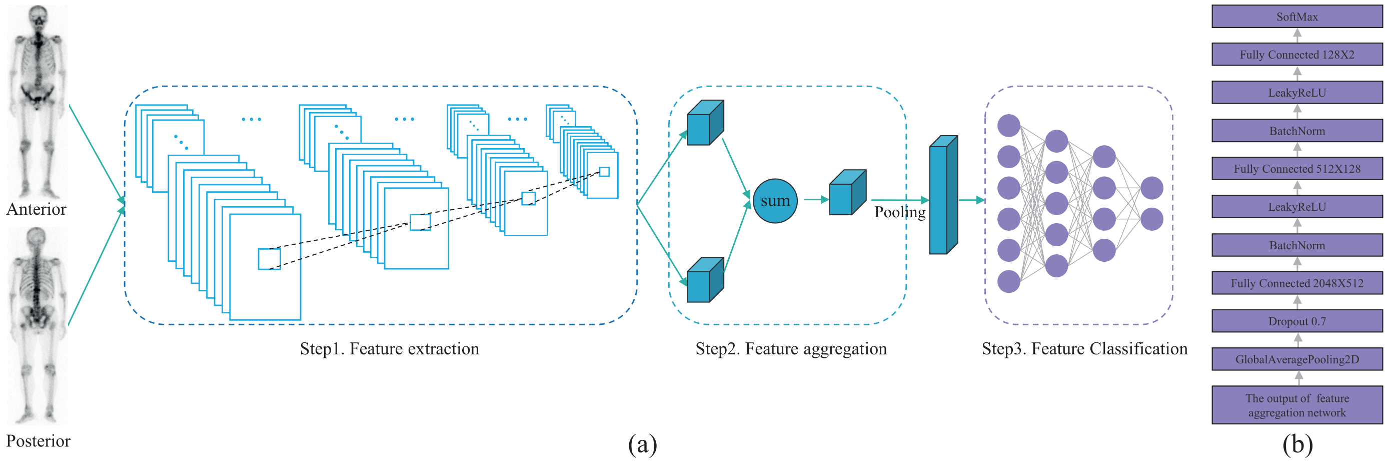


Fig. 3. Overview of the overall architecture (a) Overview of the proposed three parts of the network. In part one, a feature extraction network is employed to extract high level features from input images. Then, in part two, these high-level features are fused by a feature aggregation network. Finally, a classification network is employed to classify the fused features. (b) The detailed architecture of the final classification network N_{cl}.

CNN models were explored, Inception-V3 (Szegedy et al., 2016), DenseNet (Huang et al., 2017), and SENet (Hu et al., 2018). The modules before the first fully connected layer in these CNNs are used as the feature extraction network. The Inception-V3 network consists of several ”Inception modules”, with each module being a composition of pooling layers and convolutional layers with different kernel sizes. By using these multi-scale kernels, the Inception-V3 network shows a better ability to capture features with different sizes and shapes. DenseNet uses a dense connectivity mechanism to solve gradient vanish and ”wash out”, and requires fewer parameters than traditional CNNs. Differing from residual connections, DenseNet combines features by concatenating them instead of summing them, and it can be formulated as $a^l = f([a^0, a^1, \dots, a^{l-1}])$, where $[a^0, a^1, \dots, a^{l-1}]$ denotes the concatenation of the feature maps produced in previous layers. SENet proposes a ”Squeeze-and-Excitation”(SE) block that adaptively recalibrates channel-wise feature responses by explicitly modelling interdependencies between channels.

4.3. Part two: feature fusion operator

Differing from traditional image classification tasks that perform decisions based on a single image, the diagnosis of WBS images must be based on the joint analysis of anterior and posterior view images. This requires the model to perform a diagnosis using multiple images instead of a single one. To achieve this goal, the model should learn to aggregate features from both of the images, and then perform the recognition using the aggregated features. Thus, the aggregation of features forms a key problem that heavily influences the performance of the model.

In this study, several feature aggregated strategies were explored. The max and mean feature aggregating operators are two classical fusion operators (Feichtenhofer et al., 2016) that can be formulated as:

$$s^i_{c,w,h} = \max_{j=1,\dots,J} (f^{i,j}_{c,w,h}). \tag{3}$$

and

$$s^i_{c,w,h} = \frac{1}{J} \sum_{j=1}^{J} (f^{i,j}_{c,w,h}). \tag{4}$$

where $s^i_{c,w,h}$ is one element of the aggregated feature, S_i, $f^{i,j}_{c,w,h}$ is one element of the $F_{i,j}$, c, w, h indicating the spatial location, and $F_{i,j}$ is the high-level feature of the jth image in ith examination.

Hot spots in WBS images are the key feature for diagnosing bone metastasis. To increase the focus of the model on hot spot areas, thereby rendering a higher performance, a spatial attention feature aggregation operator is proposed, which is inspired by Hu et al. (2018). High-level features are weighted across spatial locations and aggregated by a sum operator. The proposed mechanism is described as follows. First, the high-level features $F_{i,j}$ extracted by the feature extraction network are passed through a convolutional layer that produces a spatial location descriptor $M_{i,j}, m^{i,j}_{1,w,h} \in \Re^{1\times W\times H}$. Here, the kernel size of the convolutional layer is 1 × 1, and the input channels and output channels are equal to C and 1 respectively. Then, a sigmoid function is applied to $M_{i,j}$ producing a weight descriptor $Q_{i,j}, q^{i,j}_{1,w,h} \in \Re^{1\times W\times H}$ across the spatial location:

$$q^{i,j}_{1,w,h} = \frac{1}{1 + e^{(-m^{i,j}_{1,w,h})}}. \tag{5}$$

Finally, $Q_{i,j}$ is multiplied by $F_{i,j}$ and the sum operator is used to aggregate the scaled embedding:

$$s^i_{c,w,h} = \sum_{j=1}^{J} (q^{i,j}_{1,w,h} \cdot f^{i,j}_{c,w,h}). \tag{6}$$

The detail of this spatial attention block is depicted in Fig. 4.

4.4. Part three: classification network

Based on the characteristics of the WBS images, a customized standard deep neural network (SDNN) was designed as a component classifier for the final part of our network. The fused feature maps produced by part two are input to the SDNN to produce the final prediction. A global pooling layer is first employed to normalize the spatial size of the feature maps; thus, a fully connected layer can be added behind it. A dropout layer (Srivastava et al., 2014) is added after the global pooling layer to alleviate network overfitting. The drop probability of this dropout layer was set to 0.7 for all experiments in this study. Following the dropout layer are several modules, each of which is a composite of three consecutive operations: a fully connected layer, a batch normalization (BN) (Ioffe and Szegedy, 2015), and a leaky rectified linear unit (LeakyReLU). The architecture of the proposed SDNN is depicted in Fig. 3.

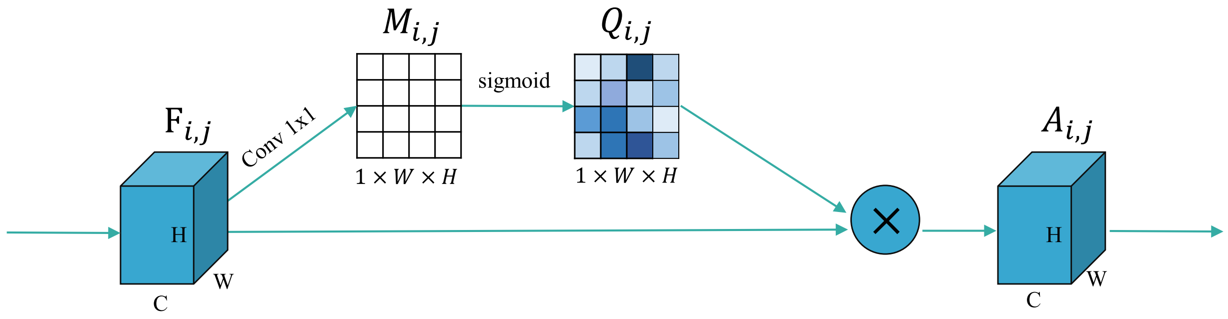


Fig. 4. A spatial attention block. High-level features are first passed through a convolutional layer, which produces a spatial location descriptor $M_{i,j}$. Then $M_{i,j}$ is applied a sigmoid function producing a weight descriptor $Q_{i,j}$. Finally, $F_{i,j}$ is multiplied by $Q_{i,j}$. C, W, H denote the channel, width, and height of features respectively.

4.5. Image pre-processing

The original data of the WBS images are in Hounsfield units (HU) format and are stored in DICOM files. Before input to the model, the HU values were converted to grayscale images with a range of [0, 255], using a windowWidth setting equal to 47 and a windowCenter equal to 23.5. Note that the bone area is white in color with a black background in the original WBS images, while the color of all WBS images presented in this study is reversed to improve the presentation. The size of the valid area in the images captured using the Philips device is the same as that with the GE device. The images captured by both the GE and Philips devices have large meaningless black edges, and to save computational resources and render a higher performance, a region of interest (ROI) extraction algorithm based on a thresholding segmentation method was developed to extract the valid area from the original images. The extracted images exhibited resolutions in the range of 201 × 690 to 975 × 253 pixels, with height-width ratios in the range of 2.7 to 4.1. We standardized the resolution by padding all images to a uniform height-width ratio R_{hw}, and resized them to a uniform size so that the smaller edge of the image was equal to 256. The total pre-processing method is presented in Algorithm 1, and all functions used in the algorithm were built-in functions of the Open Source Computer Vision Library (Bradski, 2000). The *getContours* function implements the algorithm proposed by Suzuki et al. (1985). In this study, the ratio R_{hw} was fixed to 3.4, which is the average ratio across the whole dataset, and the threshold *th* was fixed at 10. After preprocessing, all images exhibited a resolution of 256 × 846, with nearly no black border.

Algorithm 1: ROI extraction.

```
Input : Original anterior view image $I_{ant}$ and corresponding
        posterior view image $I_{post}$, height-width ratio $R_{hw}$,
        thresh $th$.
Output: Extracted ROI image of both two view, $R_{ant}$ and $R_{post}$.
begin
    for $i$ in height of $I_{ant}$ do
        for $j$ in width of $I_{ant}$ do
            if $I_{ant}[i, j] > th$ then
                $I_{ant}[i, j] = 255$
            end
            else
                $I_{ant}[i, j] = 0$
            end
        end
    end
    Find contours of all the continuous points on $I_{ant}$, get a
    list $L_{cnts}$ that contains all contours.
    $L_{cnts} = getContour(I_{ant})$
    $area_{\max} = 0, cnt_{\max} = None$
    for $cnt$ in $L_{cnts}$ do
        area = getContourArea(cnt)
        if $area > area_{\max}$ then
            $area_{\max} = area$
            $cnt_{\max} = cnt$
        end
    end
    Find the bounding Rectangle of $cnt_{\max}$, get $x, y, h, w$. Here
    x, y is the coordinates of the upper left point of the
    rectangle, w, h denote the width and height of the
    rectangle.
    $x, y, w, h = boundingRect(cnt_{\max})$
    $center_x = x + w/2, center_y = y + h/2$
    if $h/w > R_{hw}$ then
        $w = h/R_{hw}$
    end
    else
        $h = w * R_{hw}$
    end
    $R_{ant} = I_{ant}[center_y - h/2 : center_y + h/2, center_x - w/2 :$
    $center_x + w/2]$
    $R_{post} = I_{post}[center_y - h/2 : center_y + h/2, center_x - w/2 :$
    $center_x + w/2]$
    return $R_{ant}, R_{post}$.
end
```

5. Experimental and results

In this section, we first describe the experimental configurations in detail, including the implementation of the proposed method, evaluation strategy, and evaluation metric. We then present the experimental results and analysis.

5.1. Experimental configuration

Implementation: All feature extraction networks are pre-trained on the ImageNet dataset, then fine-tuned on our dataset using adadelta (Zeiler, 2012) as the optimizer with a learning rate of 0.1 with a weight decay rate of 10^{-4} for 200 epochs. The mini-batch size is fixed at 12. The proposed methods were implemented using PyTorch (Paszke et al., 2017), an open-source deep learning platform. All experimental trials were performed using a workstation with Ubuntu operating system, four Nivdia Tesla P100 GPUs, and 64 GB of RAM.

Evaluation Metrics: The overall performance of each model is assessed on the testing set, using the model which achieves the high-

Table 5
Comparison of different input ways. All experiments are using Inception-V3 as the feature extraction network with max feature aggregating operator.

Input ways	F1	Accuracy	Sensitivity	Specificity
A	0.909	93.31%	89.50%	95.60%
B	0.923	**94.19%**	92.83%	95.00%
C	0.914	93.69%	89.50%	96.20%

Table 6
Comparison of different feature aggregation operators. Different feature extraction networks are also explored.

Aggregation	Feature extraction networks		
methods	Inception-V3	DenseNet-169	SE-ResNet-50
max	94.19%	93.94%	93.94%
mean	94.44%	94.13%	94.06%
attention	**95.00%**	**94.44%**	**94.56%**

est accuracy on the validation set. In this study, we use the Sensitivity, Specificity, Accuracy, and F_1-score as evaluation criteria.

Evaluation Strategy: Several experiments are conducted to verify the advance of proposed methods. First, we explored the influence of preprocessing methods for models; second, different state-of-the-art ImageNet pretrained networks used as feature extraction network were compared; third, the effectiveness the proposed spatial attention block was analyzed; Lastly, we compared our model with three experienced physicians, further validated the effectiveness of our methods.

5.2. Input methods

CNNs have acquired many successes in the image recognition field. However, there is an obvious problem with the state-of-the-art CNN models: these models require a fixed input image size. For example, SENet uses an input size of 224 × 224, and DenseNet uses images with a size of 299 × 299. In this study, the WBS images had a resolution of 256 × 876 after preprocessing, and direct resizing of such high aspect ratio images to a square size could result in geometric distortion, which would impede the diagnosis of bone metastasis. Moreover, differing from most medical imaging tasks which use a single image for diagnosis, the diagnosis of bone scan images must involve the joint analysis of both anterior and posterior views. In this study, we propose a methodology that analyses a two-view bone scan examination in a holistic manner. To explore the best input method for bone scan images, several experiments were conducted, and these are described in this section.

These three input methods were:

- A: Direct resizing of the pre-processed anterior and posterior images to 256 × 256. The input in this model is two images with a resolution of 256 × 256.
- B: Directly inputting the pre-processed anterior and posterior images into the model with a change in the kernel size of the final pooling layer so that the output of the pooling layer has a width and height equal to 1.
- C: Directly inputting the pre-processed anterior and posterior images into the model with replacement of the final pooling layer in the model with a Spatial Pyramid Pooling (SPP) layer (He et al., 2015). SPP is famous for its strong ability to tackle input images with arbitrary sizes.

All three experiments used Inception-V3 as the feature extraction network with a max feature aggregating operator.

The results of the experiments are shown in Table 5. The model using input method B showed the highest performance, achieving an F1 score of 0.923, an accuracy of 94.19%, a sensitivity of 92.83%, and a specificity of 95.00% on the testing set. The accuracy of method A was a bit lower, which was mainly caused by the geometric distortion. The original SPP layer used in input method C also showed a lower accuracy than method B, indicating it did not work well with WBS images.

5.3. Feature aggregation methods

The performance of the proposed feature aggregation operator is described below. To explore the effect of the feature extraction network architecture on the bone scan image analysis tasks, several state-of-the-art classification networks were tested, including Inception-V3 (Szegedy et al., 2016), DenseNet-169 (Huang et al., 2017), and SE-ResNet-50 (Hu et al., 2018). Table 6 presents the comparison of the different architectures of the feature extraction networks and different feature aggregation methods. As shown in the table, Inception-V3 combined with the spatial attention operator showed the best performance, which was mainly a result of the multi-scale kernels. The spatial attention feature aggregation operator achieved better performance than the max and mean operator for almost all feature extraction networks, indicating the proposed attention method effectively helped the models to focus on the key areas in the image. We used the model with an Inception-V3 feature extraction network with a spatial attention feature aggregation operator as our final architecture, as it achieved the best performance with an accuracy of 95.00%.

5.4. Multi-view versus single view

We compared the performance of the network with or without multiple view fusions. The network trained on single view is Inception-V3. The proposed method was further compared with two recent multi-view studies for diagnosing mammography images (Carneiro et al., 2017; Khan et al., 2019). The input size of all networks is 256 × 876. The results are depicted in Table 7. The network using posterior view as input shows higher performance than anterior view, indicating that posterior view could contain more information than anterior view. Furthermore, the network with multiple view fusion shows a large improvement in performance compared to the network with only anterior or posterior view, and our method outperforms both the methods described above.

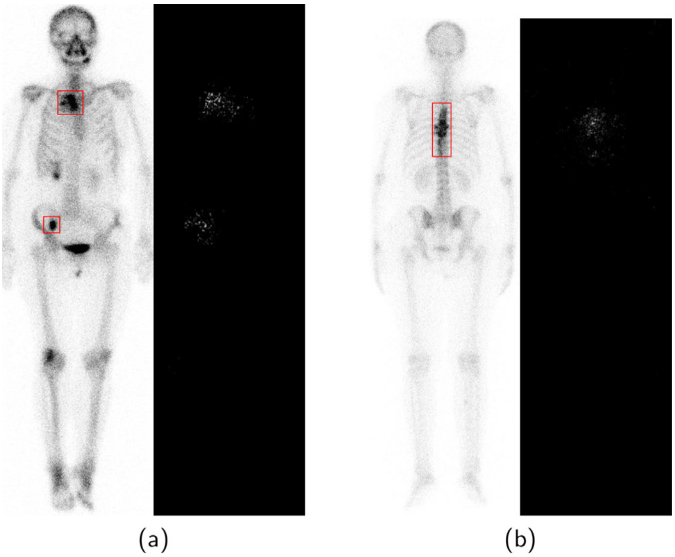
5.5. Visualization

The visualization results of the proposed models are presented in Fig. 5, using a guided backpropagation algorithm (Springenberg et al., 2014) for the classification of benign and malignant bone scans. The guided backpropagation algorithm computes the gradient with the most activation in the output layer with regard to the inputs.

The first two images are of true positive cases and their corresponding visualization results. By comparing these two images it can be clearly observed that the maximum output neuron in the network is highly correlated with the areas of hot spots in the input images. This is consistent with the guidelines for physicians, in which the hot spots on the bones are important for the analysis of WBS images, and indicates that the proposed model learned to extract the essential features for diagnosing the absence or presence of bone metastasis from the WBS images. The latter two images are of false positive cases and their corresponding visualization results. Most false positive cases were atypical samples, and although the predictions were wrong, the model was still capable of focusing on hot spot areas within the images.

Table 7
Comparison of network with single- or multi-view inputs. The proposed method was further compared with two recent multi-view studies.

		F1	Accuracy	Sensitivity	Specificity
Single View	anterior	0.86	90.25%	79.67%	**96.60%**
	posterior	0.9	92.75%	86.67%	96.40%
Multi-view	Carneiro et al. (2017)	0.902	92.88%	87.50%	96.10%
	Khan et al. (2019)	0.912	93.56%	89.00%	96.30%
	our	**0.933**	**95.00%**	**93.17%**	96.10%



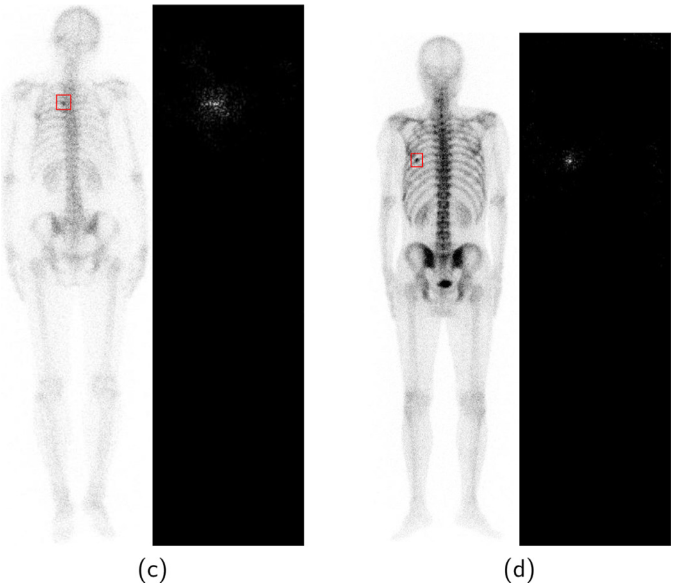


Fig. 5. Visualization examples using guided backpropagation algorithm. (a) and (b) are images of the true positive. (c) and (d) are images of the false positive.

5.6. Model ensemble and clinical test

The ensemble learning is known to be an effective approach for enhancing the performance for which individual models were independently trained. In this section, we explored the ensemble performance of the proposed networks. The final prediction score of ensemble models was the averaged softmax scores of all models. The results were acquired on the testing set and shown in the Table 8. The ensembled model showing a high performance, with accuracy of 95.25%, sensitivity of 92.67% and specifity of 96.80%.

Table 8
Performance of each single network and the ensembled model.

Model	F1	Accuracy	Sensitivity	Specificity
0:Inception-V3	0.933	95.00%	93.17%	96.10%
1:DenseNet-169	0.925	94.44%	91.33%	96.30%
2:SE-ResNet-50	0.927	94.56%	91.50%	96.40%
Ens(012)	0.936	95.25%	92.67%	96.80%

Table 9
Results of the proposed model and human experts. The model was using Inception-V3 as the feature extraction network. 150 benign and 150 malignant cases are involved and three human experts were instructed to perform diagnosis based on the WBS images. The performance of our model is comparable to human experts.

	F1	Accuracy	Sensitivity	Specificity	Time
Inexperienced	0.731	75.00%	68.00%	82.00%	135mins
Moderately	0.767	80.00%	66.00%	94.00%	49mins
Experienced	0.847	86.00%	77.33%	94.66%	160mins
Model	0.893	89.00%	92.00%	86.00%	24s

Table 10
The significance test results about the accuracy, specificity, and sensitivity among the proposed method and human experts. P values were calculated by using the 2-sided Pearson's chi-squared test. Two-tailed $P < 0.05$ indicates a significant difference.

Group	Accuracy	Sensitivity	Specificity
Junior,Intermediate	0.143	0.373	0.001
Junior,Senior	0.001	0.203	0.001
Intermediate,Senior	> 0.05	0.029	0.803
Junior,Model	< 0.001	< 0.001	0.345
Intermediate,Model	0.002	< 0.001	0.021
Senior,Model	0.267	< 0.001	0.011

5.7. Comparisons between the model and experts

To further evaluate the performance of our model, its performance was compared with three nuclear medicine physicians. According to a previous study (Sadik et al., 2009), these three experts can be divided into three levels: inexperienced ($<$ 800 WBS interpretations), moderately experienced (800-5000 WBS interpretations), and experienced ($>$ 5000 WBS interpretations). A total of 1201 examinations performed at the Department of nuclear medicine of West China Hospital, Sichuan University from Jan to Feb 2016, were collected for this comparison. Examinations that were easy to classify were excluded by a senior expert, and the final dataset constructed for the expert-model comparison, contained 150 benign and 150 malignant examinations randomly chosen from the collected examinations. The gold standard for the examinations in the dataset was based on the clinical reports and was validated by medical follow-up. In the diagnosis process, the physicians were blinded to the ground truth diagnosis, the distribution of patients with bone metastasis in the dataset. Both our

Table 11
Recall of the proposed model on each primary tumor type among bone metastasis examinations in the testing set.

Type	Recall	Type	Recall
Lung cancer	99.60%(246/247)	Pancreatic cancer	66.67%(2/3)
Prostate cancer	98.28%(114/116)	Thyroid cancer	66.67%(2/3)
Breast cancer	99.22%(127/128)	Lymphoma	33.33%(1/3)
Colorectal-cancer	78.57%(11/14)	Ureteral carcinoma	66.67%(2/3)
Nasopharyngeal-carcinoma	86.67%(13/15)	Laryngeal carcinoma	66.67%(2/3)
Liver cancer	85.71%(12/14)	Malignant melanoma	33.33%(1/3)
Gastric cancer	66.67%(6/9)	Endometrial cancer	33.33%(1/3)
Renal cancer	66.67%(6/9)	Biliary carcinoma	66.67%(2/3)
Bladder cancer	57.14%(4/7)	Parotid carcinoma	0.00%(0/2)
Esophageal carcinoma	50.00%(3/6)	Ovarian carcinoma	50.00%(1/2)
Mediastinal malignant tumor	75.00%(3/4)	Total	93.17%(559/600)
Uterine cervix cancer	0.00%(0/3)		

Table 12
Recall of the proposed model on each primary lesion type among benign examinations in the testing set.

Type	Number	Type	Recall
Breast cancer	98.40% (246/250)	Thyroid cancer	100.00%(3/3)
Lung cancer	99.58%(239/240)	Benign lung lesions	96.67%(203/210)
Prostate cancer	95.00%(38/40)	Benign prostate lesions	75.00%(6/8)
Nasopharyngeal-carcinoma	88.00%(22/25)	Benign breast lesions	80.00%(4/5)
Esophageal carcinoma	93.33%(14/15)	Benign liver lesions	75.00%(3/4)
Colorectal-cancer	73.33%(11/15)	Benign brain lesions	100.00%(3/3)
Liver cancer	100.00%(15/15)	Benign renal lesions	66.67%(2/3)
Gastric cancer	83.33%(5/6)	Benign thyroid lesions	0.00%(0/3)
Renal cancer	100.00%(5/5)	Benign gastric lesion	100.00%(3/3)
Mediastinal malignant tumor	80.00%(4/5)	Other Cancers	87.50%(14/16)
Bladder cancer	66.67%(2/3)	Other Benign lesions	96.67%(116/120)
Ovarian cancer	100.00%(3/3)	Total	96.1%(961/1000)

automated system and the physicians performed the diagnoses using only the WBS images, without extra information such as clinical data of the patients.

The sensitivity and specificity of our system and the three experts are shown in Table 9. The accuracy of the experts was highly dependent on their level of experience, with the expert with higher experience showing higher accuracy. The proposed model achieved an accuracy of 89% and outperformed all three experts, especially the junior expert. Furthermore, the total time taken by our model was 24 s, while it was around 135 min for the inexperienced physician, 49 min for the moderately experienced physician, and 160 min for the experienced human physician. The diagnosis of the proposed model was very efficient. The high diagnostic accuracy achieved on this clinical data demonstrates the effectiveness of the proposed system for the diagnosis of WBS images, and that it can be applied as a clinical decision support tool. Moreover, the 2-sided Pearson's chi-squared test was conducted to evaluate whether there are significant differences in specificity, sensitivity as well as accuracy among the proposed method and human experts. Table 10 shows *P* values of the 2-sided Pearson's chi-squared test. The proposed method achieved a significantly larger sensitivity than three human experts, which is most meaningful in clinical diagnosis.

5.8. Analytic experiment

The dataset constructed in this study followed a real-world distribution without excluding any cases, under the premise that a system with good performance on this dataset would be more suitable for routine clinical application. Tables 2 and 3 list the types and incidences of the primary tumors among bone metastasis examinations and the primary lesions among benign examinations in the dataset respectively. In this sub-section, we present the performance of the proposed method for each type of primary lesions in detail. All results are assessed using Inception-V3 as the feature extraction network on the testing set. Table 11 shows the performance of the proposed method for each type of primary tumors among bone metastasis examinations. Table 12 shows the performance of the proposed method for each type of primary lesions among benign examinations.

It can be seen that the proposed method shows a good and similar performance among the types with a large number of examinations. For types with a small number of examinations, the model still shows a noticeable recall rate. Thus our model could be a strong system for routine clinical bone scans diagnosis.

6. Conclusion

In this paper, we demonstrate the effectiveness of a deep neural network model for recognizing the absence or presence of bone metastasis on WBS images. The developed architecture was based on deep convolutional neural networks and consisted of three parts: a feature extraction network, a feature aggregation network, and a feature classification network. Several methods of data input into the model were compared. Three state-of-the-art ImageNet-pretrained networks were explored for use as feature extraction networks: Inception-V3, DenseNet-169, and SE-ResNet-50. We constructed a large-scale annotated WBS image dataset containing 15,474 examinations to train and evaluate the proposed model. Benefitting from the novel architectures and the large-scale dataset, our results show that our model demonstrated superior performance. To the best of our knowledge, this is the first work exploring deep neural networks for the automated diagnosis of bone metastasis directly on WBS images. The constructed WBS image dataset is larger than the datasets used in all previous research by an order of magnitude. Furthermore, previous studies usually excluded cases that could be misleading during the training process, such as patients with a urinal catheter, large bladder, sternotomy, or fracture. Our dataset differs in that it follows a natural distribution without exclusion of any atypical cases, yet the developed system still demonstrated high performance, especially with regard to its high sensitivity, which is helpful for reducing the false negative rate of human experts.

The visualization results demonstrate that the proposed method learned to locate abnormal areas, which are the features of most concern to physicians. Compared with three experienced nuclear medicine physicians, the developed system obtained comparable performance and made the diagnosis very efficiently, rendering it feasible for clinical applications. Furthermore, it provides the same diagnosis on a given image every time, which is difficult for a human expert. Our future studies will focus on the following: 1) verifying our model on a larger real-world dataset in a clinical setting; 2) collecting more data to improve the model's performance; 3) detecting the abnormalities in images and directly characterizing the lesions; and 4) combining CT images for difficult to diagnose samples, potentially reducing the false positive rate.

Declaration of Competing Interest

None.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 61432012 and U1435213) and 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (ZYGD18016)

References

Belcher, L., 2017. Convolutional neural networks for classification of prostate cancer metastases using bone scan images. Student Paper

Bombardieri, E., Aktolun, C., Baum, R.P., Bishof-Delaloye, A., Buscombe, J., Chatal, J.F., Maffioli, L., Moncayo, R., Mortelmans, L., Reske, S.N., 2003. Bone scintigraphy: procedure guidelines for tumour imaging. Eur. J. Nucl. Med. Mol. Imaging 30 (12), B99–B106. doi:10.1007/s00259-003-1347-2.

Bradski, G., 2000. The OpenCV Library. Dr. Dobb's J. Softw. Tools.

Carneiro, G., Nascimento, J., Bradley, A.P., 2017. Automated analysis of unregistered multi-view mammograms with deep learning. IEEE Trans. Med. Imaging 36 (11), 2355–2365.

Coleman, R., 2001. Metastatic bone disease: clinical features, pathophysiology and treatment strategies. Cancer Treat. Rev. 27 (3), 165–176.

Erdi, Y.E., Humm, J.L., Imbriaco, M., Yeung, H., Larson, S.M., 1997. Quantitative bone metastases analysis based on image segmentation. J. Nucl. Med. 38 (9), 1401.

Feichtenhofer, C., Pinz, A., Zisserman, A., 2016. Convolutional two-stream network fusion for video action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1933–1941.

Geng, S., Jia, S., Qiao, Y., Yang, J., Jia, Z., 2015. Combining CNN and mil to assist hotspot segmentation in bone scintigraphy. In: International Conference on Neural Information Processing. Springer, pp. 445–452.

Geng, S., Ma, J., Niu, X., Jia, S., Qiao, Y., Yang, J., 2016. A mil-based interactive approach for hotspot segmentation from bone scintigraphy. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 942–946.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37 (9), 1904–1916.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. In: European Conference on Computer Vision. Springer, pp. 630–645.

Hu, J., Chen, Y., Yi, Z., 2019. Automated segmentation of macular edema in OCT using deep neural networks. Med. Image Anal. 55, 216–227.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

Huang, J.-Y., Kao, P.-F., Chen, Y.-S., 2007. A set of image processing algorithms for computer-aided diagnosis in nuclear medicine whole body bone scan images. IEEE Trans. Nucl. Sci. 54 (3), 514–522.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167.

Jouirou, A., Baâzaoui, A., Barhoumi, W., 2019. Multi-view information fusion in mammograms: acomprehensive overview. Inf. Fusion 52, 308–321.

Khan, H.N., Shahid, A.R., Raza, B., Dar, A.H., Alquhayz, H., 2019. Multi-view feature fusion based four views model for mammogram classification using convolutional neural network. IEEE Access 7, 165724–165733.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105.

LeCun, Y., Bengio, Y., et al., 1995. Convolutional networks for images, speech, and time series. Handb. Brain Theory Neural Netw. 3361 (10), 1995.

Lin, M., Chen, Q., Yan, S., 2013. Network in network. arXiv:1312.4400.

Liu, X., Hou, F., Qin, H., Hao, A., 2018. Multi-view multi-scale CNNs for lung nodule type classification from CT images. Pattern Recognit. 77, 262–275.

Mundy, G.R., 1997. Mechanisms of bone metastasis. Cancer 80 (S8), 1546–1556.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in PyTorch. NIPS-W.

Qi, X., Zhang, L., Chen, Y., Pi, Y., Chen, Y., Lv, Q., Yi, Z., 2019. Automated diagnosis of breast ultrasonography images using deep neural networks. Med. Image Anal. 52, 185–198.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 115 (3), 211–252.

Sadik, M., Hamadeh, I., Nordblom, P., Suurkula, M., Höglund, P., Ohlsson, M., Edenbrandt, L., 2008. Computer-assisted interpretation of planar whole-body bone scans. J. Nucl. Med. 49 (12), 1958–1965.

Sadik, M., Jakobsson, D., Olofsson, F., Ohlsson, M., Suurkula, M., Edenbrandt, L., 2006. A new computer-based decision-support system for the interpretation of bone scans. Nucl. Med. Commun. 27 (5), 417–423.

Sadik, M., Suurkula, M., Höglund, P., Järund, A., Edenbrandt, L., 2009. 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 (3), 368–375.

Shin, H.-C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging 35 (5), 1285–1298.

Springenberg, J. T., Dosovitskiy, A., Brox, T., Riedmiller, M., 2014. Striving for simplicity: the all convolutional net. arXiv:1412.6806.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Suzuki, S., et al., 1985. Topological structural analysis of digitized binary images by border following. Comput. Vis. Graph. Image Process. 30 (1), 32–46.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.

Wang, S., Zhou, M., Gevaert, O., Tang, Z., Dong, D., Liu, Z., Tian, J., 2017. A multi-view deep convolutional neural networks for lung nodule segmentation. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 1752–1755.

Wu, Q., Yang, R., Zhou, F., Hu, Y., 2013. Comparison of whole-body MRI and skeletal scintigraphy for detection of bone metastatic tumors: a meta-analysis. Surg. Oncol. 22 (4), 261–266.

Yin, T.-K., Chiu, N.-T., 2004. A computer-aided diagnosis for locating abnormalities in bone scintigraphy by a fuzzy system with a three-step minimization approach. IEEE Trans. Med.Imaging 23 (5), 639–654.

Zeiler, M.D., 2012. ADADELTA: An adaptive learning rate method. Comput. Sci..

Zhang, W., Zhong, J., Yang, S., Gao, Z., Hu, J., Chen, Y., Yi, Z., 2019. Automated identification and grading system of diabetic retinopathy using deep neural networks. Knowl. Based Syst. 175, 12–25.

Zhao, Z., Li, L., Li, F.-l., 2009. Radiography, bone scintigraphy, SPECT/CT and MRI of fibrous dysplasia of the third lumbar vertebra. Clin. Nucl. Med. 34 (12), 898–901.