

# A Multi-Resolution Approach for Spinal Metastasis Detection using Deep Siamese Neural Networks

Juan Wang<sup>a</sup>, Zhiyuan Fang<sup>b</sup>, Ning Lang<sup>c</sup>, Huishu Yuan<sup>c</sup>, Min-Ying Su<sup>d</sup>,  
Pierre Baldi<sup>a,\*</sup>

<sup>a</sup>*Institute for Genomics and Bioinformatics and Department of Computer Science,  
University of California, Irvine, CA, 92697 USA*

<sup>b</sup>*Department of Computer Science, Southern University of Science and Technology,  
Shenzhen, Guangdong 518055, China*

<sup>c</sup>*Department of Radiology, Peking University Third Hospital, Beijing, 10019, China*

<sup>d</sup>*Department of Radiological Sciences, University of California, Irvine, CA 92697 USA*

---

\*Corresponding author. Email: pfbaldi@ics.uci.edu, Phone: (+1)9498245809, Fax: (+1)9498249813.

## Abstract

Spinal metastasis, a metastatic cancer of the spine, is the most common malignant disease in the spine. In this study, we investigate the feasibility of automated spinal metastasis detection in magnetic resonance imaging (MRI) by using deep learning methods. To accommodate the large variability in metastatic lesion sizes, we develop a Siamese deep neural network approach comprising three identical subnetworks for multi-resolution analysis and detection of spinal metastasis. At each location of interest, three image patches at three different resolutions are extracted and used as the input to the networks. To further reduce the false positives (FPs), we leverage the similarity between neighboring MRI slices, and adopt a weighted averaging strategy to aggregate the results obtained by the Siamese neural networks. The detection performance is evaluated on a set of 26 cases using a free-response receiver operating characteristic (FROC) analysis. The results show that the proposed approach correctly detects all the spinal metastatic lesions while producing only 0.40 FPs per case. At a true positive (TP) rate of 90%, the use of the aggregation reduces the FPs from 0.375 FPs per case to 0.207 FPs per case, a nearly 44.8% reduction. The results indicate that the proposed Siamese neural network method, combined with the aggregation strategy, provide a viable strategy for the automated detection of spinal metastasis in MRI images.

*Keywords:* deep learning, Siamese neural network, multi-resolution analysis, spinal metastasis, magnetic resonance imaging

## 1. Introduction

Spinal metastasis, a metastatic cancer to the spine, is a malignant process in the spine. It is 25 to 35 times more common than any other malignant diseases in the spine [1] and affects more than 100,000 individuals in the U.S. annually [2]. The spine is the third most common site for cancer cells to metastasize [3], following the lung and the liver [4]. More than 80% of spinal metastasis in adults came from the primary tumors, including breast (72%), prostate (84%), lung (31%), thyroid (50%), kidney (37%), and pancreas (33%) [5]; and 30%–90% of cancer patients who die are found to have spinal metastasis in cadaver studies [6]. In addition, spinal metastases can also have a huge impact on quality of life, with complications including pain, fracture, and spinal cord and nerve root compression [7]. Therefore, the detection, diagnosis, and treatment of spinal metastases is clinically important both to save patients' lives and to improve their quality of life.

Due to its excellent soft tissue resolution, magnetic resonance imaging (MRI) is the most sensitive imaging modality for evaluating spinal lesions [8, 9]. Various studies have shown that early stages of spinal metastasis in the bone marrow can be detected with MRI before any bone deterioration [5]. In MRI images, neoplastic involvement in the vertebral body typically shows focal bone marrow replacement with tumorous tissue, resulting in lower T1 signal than adjacent skeletal muscle and accompanying high T2 signal. Therefore, MRI images acquired with different pulse sequences (denoted as MRI sequences) can be used to locate lesions and evaluate the extent of the disease (e.g. involving single or multiple segments). However, despite the advantages mentioned above, the manual detection of spinal metastasis in

26 MRI is time-consuming and tedious considering the large number of slices in  
27 each MRI sequence, as well as the large number of MRI sequences usually  
28 acquired for each patient. Therefore, it is now becoming essential to develop  
29 computerized algorithms for automated detection of spinal metastases in  
30 MRI sequences.

31 Especially, as image acquisition speed improves, many more images with a  
32 higher spatial resolution can be acquired during an examination, and as such  
33 developing computer-aided analysis methods is essential to assist radiologists  
34 in making a thorough evaluation of the entire image set within a reasonable  
35 reading time. This is a trend for all imaging modalities (and all patholo-  
36 gies), and particularly true for MRI because of the multiple sets of images  
37 acquired using different sequences. Although at the present time computer-  
38 aided analysis cannot yet replace visual inspection by trained radiologists,  
39 nonetheless it can already provide an important tool for displaying the most  
40 critical information from hundreds of images, in a convenient way, to assist  
41 radiologists during diagnosis. In time, computerized methods may become as  
42 good, or even better than human experts and lead to significant economies of  
43 scale and accuracy. Thus, in short, it is important to develop computerized  
44 methods to analyze MRI and other imaging modalities in medicine.

45 Given the importance of automated spinal metastasis detection, a few  
46 approaches have been developed in the literature. For example, Roth *et al.*  
47 [10] used a deep convolutional neural network (CNN) as the 2nd tier of a two-  
48 tiered, coarse-to-fine, cascade framework to refine the candidate lesions from  
49 the first tier for sclerotic spine metastasis detection in computer tomography  
50 (CT) images. Wiese *et al.* [11] developed an automatic method based on a

51 watershed algorithm and graph cut for detecting sclerotic spine metastases in  
52 CT images. And Yao *et al.* [12] applied a support vector machine to refine  
53 the initial detections produced with a watershed algorithm for lytic bone  
54 metastasis detection in CT images. However, as can be seen, these studies  
55 are based on CT images and do not use spinal MRI sequences.

56 On the MRI side, efforts for analyzing MRI sequences have focused on  
57 different problems. For example, Carballido-Gamio *et al.* [13] developed a  
58 normalized cut method for vertebra segmentation in spinal MRI. Huang *et al.*  
59 [14] proposed an AdaBoost method for vertebra detection and an iterative  
60 normalized cut algorithm for vertebra segmentation in spinal MRI. Neubert  
61 *et al.* [15] designed an automatic method using statistical shape analysis  
62 and registration of gray level intensity profiles for 3D intervertebral disc and  
63 vertebral body segmentation in MRI. However, in spite of these efforts and to  
64 the best of our knowledge, there is no study in the literature that is focused  
65 on detecting spinal metastases in MRI sequences.

66 Automated and accurate spinal metastasis detection in MRI is a difficult  
67 task, in large part because of the considerable variability in the size of ver-  
68 tebrae. Spinal metastases usually grow in the vertebrae, which are divided  
69 into five regions: cervical, thoracic, lumbar, sacrum, and coccyx. The size of  
70 the vertebrae varies considerably both within an individual, as well as across  
71 individuals. For example, Zhou *et al.* [16] investigated the lumbar vertebrae  
72 from 126 CT images and found that the upper vertebral width is  $40.9 \pm 3.6$   
73 mm in females and  $46.1 \pm 3.2$  mm in males at L3,  $46.7 \pm 4.7$  mm in females  
74 and  $50.8 \pm 3.7$  mm in males at L4, and  $50.4 \pm 4.4$  mm in females and  $54.5 \pm 4.9$   
75 mm in males at L5.

76 In recent years [17], neural networks and deep learning have been used to  
77 successfully tackle a variety of problems in engineering, ranging from com-  
78 puter vision [18, 19, 20] to speech recognition [21], as well as in the natural  
79 sciences, in areas ranging from high energy physics [22, 23], to chemistry  
80 [24, 25], and to biology [26, 27]. Thus it is natural to consider applying  
81 deep learning methods also to biomedical images. For example, Cirean *et*  
82 *al.* applied a deep learning architecture to each pixel for addressing prob-  
83 lems of membrane segmentation in electron microscopy images [28]; Shen *et*  
84 *al.* developed multi-scale convolutional neural network for lung nodule de-  
85 tection in CT images [29]; Wang *et al.* adopted a GoogLeNet-based method  
86 for automated detection and diagnosis of metastatic breast cancer in whole  
87 slide images of sentinel lymph node biopsies [30]; and Wang *et al.* devised  
88 a 12-layer convolutional neural network for cardiovascular disease detection  
89 in mammograms [31]. Therefore it is natural to hypothesize that neural net-  
90 works and deep learning methods can be harnessed for the effective detection  
91 of spinal metastases in MRI sequences.

92 Thus, in short, the purpose of this study is to develop an accurate comput-  
93 erized method to locate metastatic cancer in the spine using deep learning  
94 methods. Specifically, a multi-resolution analysis is proposed to deal with  
95 the large variability of vertebral sizes, and a Siamese neural network [32, 33]  
96 is developed to incorporate the multi-resolution representation of each MRI  
97 slice. More precisely, for each location under consideration, a set of image  
98 patches centered at this location and at different resolutions are extracted  
99 from the MRI slice, resulting in a multi-resolution representation of the in-  
100 put. Then the multi-resolution image patches are fed into a Siamese neural

101 network to predict the probability that the corresponding location corre-  
102 sponds to a metastatic lesion. The Siamese neural network is comprised of  
103 several identical twin networks, each dealing with patches of a different reso-  
104 lution. To further remove false positives (FPs) in spinal metastasis detection,  
105 we consider the structural similarity of neighboring slices in MRI sequences  
106 and aggregate the outputs of the Siamese neural networks using a weighted  
107 averaging approach.

108 The rest of paper is organized as follows. The proposed Siamese neural  
109 network and slice-based aggregation methods for spinal metastasis detection  
110 are described in Section 2 together with other methodological aspects. The  
111 data collection and experiments are described in in Section 3. Finally, the  
112 results are presented in Section 4 and discussed in Section 5.

## 113 2. Methods

### 114 2.1. Motivation

115 As previously mentioned in Section 1, there is considerable variability in  
116 vertebral size. Such variability results in great variations in the sizes of the  
117 metastatic lesions, thus yielding the difficulty in the metastatic lesion detec-  
118 tion. In Figure 1, we provide examples of metastatic lesions with different  
119 sizes, in which each image window represents a region of  $55 \times 79 \text{ mm}^2$ . As  
120 can be seen, the size of the metastatic lesion varies hugely, in which the first  
121 lesions cover almost  $55 \text{ mm}$  in width, while the last lesions only has less than  
122 half of  $55 \text{ mm}$  in width.

123 In real application, such variability as mentioned above poses a major  
124 challenge for metastatic lesion detection. This is because most detectors (if

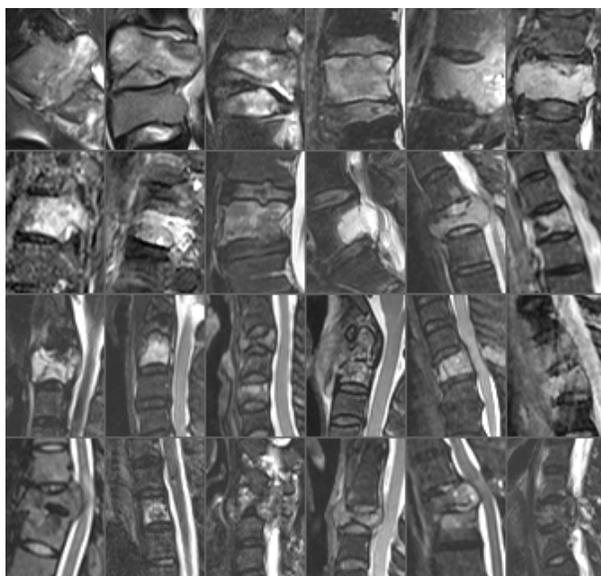


Figure 1: Example of the variability of the lesion size. Each image window represents a region of  $55 \times 79 \text{ mm}^2$ .

125 not all) are usually designed to examine only a relatively small image patch  
126 around the location under consideration. Having local detectors can be com-  
127 putationally efficient. However, in our case, if the local image patch is too  
128 small and the vertebra is large, then the detector sees only a fraction of the  
129 vertebra and cannot accurately distinguish between lesion and normal. In  
130 contrast, if the local image patch is big and the vertebra is small, then most of  
131 the information contained in the patch is not relevant for the discrimination.  
132 Therefore, to accommodate for large variations in vertebral size, we propose  
133 a multi-resolution approach for spinal metastasis detection. Specifically, for  
134 each location of interest, we extract three image patches with same size in  
135 pixels from each MRI slice, corresponding to three different resolutions, yield-  
136 ing three different local representations of the location under consideration.

137 The goal is to make sure that when metastatic lesions are present, they are  
138 salient in at least two image patches. Furthermore, in order to control the  
139 number of parameters and regularize the overall model, we use the same  
140 parameters in the networks associated with each resolution.

141 More precisely, to incorporate the input of the three image patches into  
142 a unified classification framework, we use a Siamese neural network. The  
143 Siamese neural network consists of three identical subnetworks, one for each  
144 image patch. Each subnetwork, is a multi-layer convolutional neural network,  
145 where the lower convolutional layers are used to learn and extract features  
146 which in turn are used to produce a classification by several the higher, fully  
147 connected (FC) layers. The output of the Siamese neural network can be  
148 interpreted as the probability of a metastatic lesion being associated with  
149 the central pixel in the input patch. The advantages of such an architecture  
150 are the ability to automatically learn and extract relevant features, and the  
151 combination of multi-resolution features for classification. In addition, by the  
152 Siamese neural network design, the three subnetworks share the same weight  
153 parameters. Thus the number of trainable parameters remains manageable  
154 and independent of the number of resolution patches.

## 155 *2.2. Siamese neural network architecture*

156 The subnetworks of the overall Siamese architecture comprise stacks of  
157 convolutional layers (Conv), batch normalization layers, nonlinearity layers,  
158 and max-pooling layers (Pooling). Each convolutional layer is followed by  
159 a batch normalization layer and a non-linear transformation. The convolu-  
160 tional layers are feature extractors, and the max-pooling layers enable the  
161 combination of low-level features into high-level features. In Figure 2, we

162 display the Siamese neural network architecture used in this study. Each  
 163 subnetwork includes five convolutional layers. To avoid cluttering the figure,  
 164 the batch normalization layer and the non-linearity layer, associated with  
 165 each convolutional layer, are not shown. Finally, the features learned by the  
 166 subnetworks are concatenated and fed into two fully connected layers for the  
 167 final classification. The details of each layer are as follows.

168 In the development phase, we experimented with subnetworks by using  
 169 many other architectures and in Figure 2 we only report the one with the  
 170 best performance. The architecture of the subnetworks has a sequence of  
 171 1, 2, and 2 convolutional layers before the first three max-pooling layers,  
 172 respectively. In one experiment, for instance, we started from an architecture  
 173 with a sequence of 1, 1, and 1 convolutional layers for subnetworks, and then  
 174 we gradually increased the number of convolutional layers and stopped when  
 175 no further improvement was observed.

176 *Convolutional layers* are the core layers for feature learning and extrac-  
 177 tion. Each convolutional layer produces a feature map by convolving its input  
 178 with a set of convolutional kernels. Mathematically, let  $\mathbf{x}$  be the input,  $\mathbf{y}_k$   
 179 be the  $k$ th feature map in output, and  $\mathbf{w}_k$  ( $k = 1, 2, \dots, K$ , where  $K$  is the  
 180 number of convolutional kernels) be the  $k$ th convolutional kernel, then the  
 181 convolutional layer can be described by:

$$\mathbf{y}_k = \mathbf{x} * \mathbf{w}_k + b_k \tag{1}$$

182 where  $*$  denotes the convolution operation and  $b_k$  is the bias. Usually, there  
 183 are several convolutional kernels in each convolutional layer (i.e.  $K > 1$ ).  
 184 In Figure 2, the number of convolutional kernels corresponds to the number  
 185 preceding the symbol “@” in the line at the top. For example, 16 convolu-

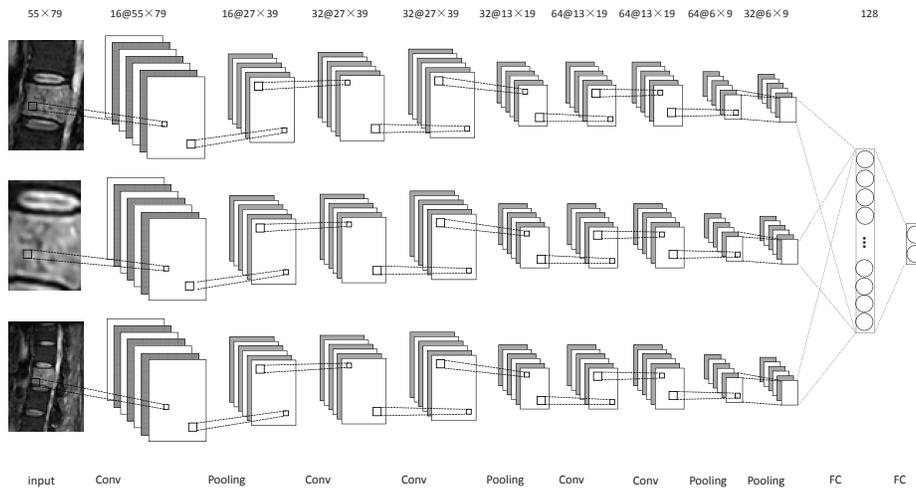


Figure 2: Diagram of the Siamese architecture used in this study. There are three identical subnetworks, each comprising five convolutional layers, five batch normalization layers, five nonlinearity layers, and four max-pooling layers for feature learning. The combination of the resulting features is subsequently input into two fully connected layers for classification. Each convolutional layer is followed by a batch normalization layer and a nonlinearity layer (not shown).

186 tional kernels are used in the first convolutional layer in Figure 2. In this  
187 study, the size of all of the convolutional kernels is set as  $3 \times 3$ .

188 *Batch normalization layers* independently normalize the feature values  
189 to zero mean and unit standard deviation in each training batch. These are  
190 introduced to deal with the internal covariate shift, which refers to the phe-  
191 nomenon that the distribution of each layer’s inputs changes during training  
192 when the parameters of the previous layers change [34]. This normalization  
193 step can speed up learning and improve classification accuracy [34]. Batch  
194 normalization layers preserve both the number and the size of the feature  
195 maps; to save space, they are omitted in Figure 2.

196 *Nonlinearity layers* consist of units which apply a non-linear activation  
197 function to their input, producing in the end a non-linear classifier. In this  
198 study, we use rectified linear units (ReLU), with a non-linear activation func-  
199 tion:

$$f(x) = \max(0, x) \tag{2}$$

200 While similar results could be obtained with sigmoidal transfer functions,  
201 ReLU units can lead to faster training [18, 35] and yield sparse representa-  
202 tions [35]. Just like the batch normalization layers, the nonlinearity layers  
203 also preserve both the number and the size of the feature maps; to save space,  
204 they are omitted in Figure 2.

205 *Max-pooling layers* serve as a method of non-linear down-sampling, by  
206 providing a summary of the outputs of a set of neighboring elements in the  
207 corresponding feature maps, simply by retaining the maximum value of the  
208 elements in each pool [36]. In this study, two types of max-pooling layers  
209 are considered. The first type of max-pooling layer generates its output by

210 considering the  $3 \times 3$  neighborhood region of the same feature map, thus  
211 reducing the size of the feature maps by nearly 50%; the second type of  
212 max-pooling layer produces its output by considering a feature map and its  
213 next feature map at the same location, thus reducing the number of feature  
214 maps by half. In this study, the first type is used for the first three max-  
215 pooling layers, and the second type is used for the last max-pooling layer.  
216 In Figure 2, the size of the feature maps is displayed in the top line, right  
217 after the “@” sign. By comparing the number and size of the feature maps  
218 between the input and the output at each stage, one can determine the type  
219 of max-pooling layer in Figure 2. For example, it can be seen that for the  
220 first max-pooling layer, the number of feature maps does not change from  
221 input to output, but the size of the feature maps decreased from  $55 \times 79$  to  
222  $27 \times 39$ , thus it is the first type of max-pooling layer.

223 *Fully connected layers* correspond to pairs of consecutive layers with full  
224 connectivity between their units. The input of the first fully connected layers  
225 are concatenation of all of the features from the three subnetworks.

226 After the last fully connected layer, a softmax activation function [19] is  
227 introduced, the output of which can be interpreted as the probability that  
228 the central pixel in the input image patch is associated with a metastatic  
229 lesion.

### 230 2.3. Model training

231 In total, the Siamese neural network in Figure 2 has around 733.4K train-  
232 able parameters. The parameters are learnt by stochastic gradient descent  
233 in order to minimize the cross entropy between true class labels and pre-  
234 dicted outputs, together with standard L2 regularization. Let  $(\mathbf{x}_i, y_i), i =$

235  $1, 2, \dots, M$  ( $M$  is the number of training samples) be the input and true  
 236 label training pairs, and  $p_i$  the corresponding predicted output for input  $\mathbf{x}_i$ ,  
 237 then the objective function can be written as:

$$J(\mathbf{w}) = - \sum_{i=1}^m [y_i \log p_i + (1 - y_i) \log(1 - p_i)] + \lambda \|\mathbf{w}\|_2^2 \quad (3)$$

238 where  $\mathbf{w}$  denotes all the weight parameters in the model and  $\lambda$  is a constant  
 239 which controls the trade-off between the classification error on the training  
 240 samples and the complexity of the model.

241 To further reduce the potential for overfitting, we use dropout [37, 38],  
 242 which can be viewed as a regularization technique. Dropout randomly drops  
 243 neural units during training to prevent any unit from being too reliant on  
 244 any other unit (unit co-adaptation). In this study, dropout with probability  
 245 0.5 is applied to the first fully connected layer.

#### 246 *2.4. Training samples extraction*

247 As input to the proposed Siamese neural network architecture, we choose  
 248 image patches with size of  $55 \times 79$  pixels from the MRI slices at three different  
 249 resolutions, i.e. 0.5 mm/pixel, 1 mm/pixel, and 2 mm/pixel, as shown for  
 250 the input layers in Figure 2. Such choice ensures that at least two out of  
 251 the three image patches cover more than one vertebra. It is based on two  
 252 observations: 1) vertebrae have roughly a square shape, and the width of  
 253 the majority of the lumbar vertebrae is in the range of 37.3 mm to 59.4 mm  
 254 [16]; and 2) during the diagnostic process, radiologists usually compare the  
 255 vertebra under consideration with other vertebrae in the same MRI slice.  
 256 More importantly, when single image resolution is considered, the image  
 257 patch size  $55 \times 79$  achieves the best detection performance.

258 As examples, in the first column of Figure 3(a), we show the multi-  
 259 resolution image patches from a big metastatic lesion. The image resolutions  
 260 for the image patches in the first, second, and third rows are 1 mm/pixel,  
 261 0.5 mm/pixel, and 2 mm/pixel, respectively. As can be seen, the metastatic  
 262 lesion is salient in the image patches with resolutions of 1 mm/pixel and 2  
 263 mm/pixel. Similarly, in the first column of Figure 3(b), we also show the  
 264 multi-resolution image patches from a small metastatic lesion. It can be seen  
 265 that the metastatic lesion is salient in the image patches with resolutions of  
 266 1 mm/pixel and 0.5 mm/pixel.

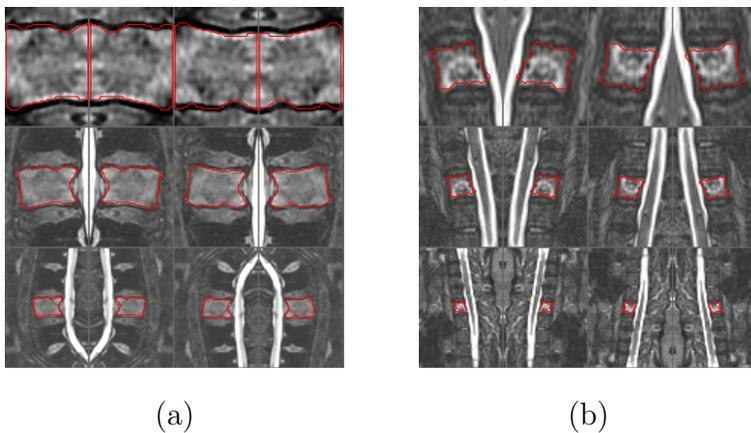


Figure 3: Examples of multi-resolution image patches from a big metastatic lesion (a) and those from a small metastatic lesion (b). The resolutions of the image patches from the first to the third rows are 0.5 mm/pixel, 1 mm/pixel, and 2 mm/pixel. In each image patch, the boundary of the metastatic lesion is marked by a red contour. For demonstration purposes, the image patches after data augmentation are shown as well.

267 The training of the Siamese neural network requires a large number of  
 268 training samples. For this purpose, we extracted training samples from the  
 269 MRI sequences in the training set, described in Section 3.1, as follows: for an

270 MRI slice under consideration, to maximize the number of metastatic lesion  
271 samples, we extract image patches from each pixel in the spinal metastatic  
272 region. Then to obtain corresponding normal samples, we randomly extract  
273 the same number of image patches from the normal region in the slice, to  
274 prevent imbalance between the two classes. To reduce the variation among  
275 the samples, a z-score normalization is applied to each image patch such that  
276 each component has zero mean and unit standard deviation.

277 Note the samples associated with the normal class mainly include two  
278 types of samples, the ones from the normal vertebral regions, and the ones  
279 from the non-vertebral background regions. For spinal metastasis detec-  
280 tion, it is much more difficult to discriminate between metastatic lesions and  
281 normal vertebrae than to discriminate between metastatic lesions and back-  
282 ground. For this reason, we select more normal samples from the normal  
283 vertebral regions. In this study, 70% of the normal samples are randomly  
284 selected from the normal vertebral regions, while the remaining 30% normal  
285 samples are randomly selected from the background regions.

286 In Figure 4, we show an MRI slice and its spinal metastatic region marked  
287 by a radiologist with a red contour together with vertebral regions marked  
288 with blue contours. Specially, for this MRI slice, all of the spinal metastasis  
289 samples are extracted from the region indicated by the red contour, 70% of  
290 normal samples are extracted from the region outside the red contour but  
291 inside the blue contours, and 30% of normal samples are extracted from the  
292 region outside the blue and red contours. It has to be noted that both the  
293 spinal metastatic and vertebral regions are only used for sample extraction,  
294 they are unknown at test time.



Figure 4: Example of lesion region, denoted by the red contour, and vertebral region, denoted by the blue contour in an MRI slice.

295 Finally, to increase the number of training samples, we use a data aug-  
 296 mentation strategy. Data augmentation has been shown to be effective in  
 297 many cases [18, 39]. To preserve the resolution of the image patches, we  
 298 consider the following two data augmentation strategies for the samples in  
 299 the training set: 1) flipping image windows from left to right and 2) flipping  
 300 image windows from top to bottom. In combinations, these strategies can  
 301 lead to a four-fold increase in the size of the training set. Examples of aug-  
 302 mented data are shown in the second to fourth columns of Figure 3 for two  
 303 representative metastatic lesions.

## 304 2.5. Testing

### 305 2.5.1. Likelihood map

306 During testing, the trained model is applied to each pixel of an MRI slice  
 307 in the testing set. The output is a likelihood map, denoted by  $l$ , with the

308 same size as the MRI slice under consideration. The value of each pixel in  
 309 the likelihood map is the estimated probability, according to the network,  
 310 that the pixel corresponds to a metastatic lesion. When the locations are  
 311 close to the boundary of the slice, the image patches extend beyond the slice  
 312 boundaries and symmetric padding is used to fill any region of the patches  
 313 falling outside the boundaries.

### 314 2.5.2. Slice-based aggregation

315 Spinal metastases can be recovered by processing the likelihood map of  
 316 each MRI slice and finding regions of high probability. However, using a single  
 317 likelihood map tends to yield many FPs when all the true positives (TPs)  
 318 are detected. To further reduce the number of FPs, we take advantage of the  
 319 3D information associated with MRI sequence and generate an aggregated  
 320 likelihood map, denoted by  $L$ , for each slice.

321 Examination of MRI sequences reveals that slice may vary greatly but  
 322 neighboring slices tend to be very similar. Based on this observation, for the  
 323  $i$ th slice and location  $(x, y)$  in an MRI sequence, we generate an aggregated  
 324 likelihood using a weighted convex combination as follows:

$$L_i(x, y) = \sum_{j \in \mathcal{N}(i)} \alpha_j l_j(x, y) \quad (4)$$

325 where  $\mathcal{N}(i)$  denotes the neighboring slices of the  $i$ th slice and  $\alpha_j$  is the weight  
 326 of the  $j$ th slice. The weights must be positive and satisfy  $\sum_{j \in \mathcal{N}(i)} \alpha_j = 1$  and  
 327  $\alpha_i > \alpha_j, j \neq i$ . The constraint  $\alpha_i > \alpha_j, j \neq i$  indicates that when obtaining  
 328 the aggregated likelihood map for the  $i$ th slice, more emphasis should be put  
 329 on its likelihood map  $l_i$  as compared to the likelihood maps  $l_j$  of its neighbors.  
 330 In this study, for a slice under consideration, we consider its previous and

331 next slices as neighboring slices used for aggregation. When the slice is the  
332 first or last slice in the MRI sequence, its next two slices or previous two slices  
333 are considered as its neighboring slices. In the experiments, we considered  
334 identical weights for both of the neighboring slices. We did a grid search on  
335 the weights of the slice on the values  $1/3$ ,  $0.4$ ,  $0.5$ , and  $0.6$ . In the end, the  
336 best performance was obtained at  $0.4$ , thus the weight of the slice and its  
337 two neighboring slices were set to be  $0.4$ ,  $0.3$ , and  $0.3$  respectively.

338 Finally, a threshold is applied to the aggregated likelihood map for spinal  
339 metastasis detection. To further reduce the number of FPs, detected regions  
340 of size smaller than  $5 \text{ cm}^2$  are considered to be FPs and are removed from  
341 the list of positives. The use of  $5 \text{ cm}^2$  is based on two observations: 1) the  
342 average abnormal segment area in the dataset is  $5.07 \text{ cm}^2$ , as described in  
343 Section 3.1; and 2) the proposed method tends to produce detected regions  
344 that are larger than those marked by radiologists, as shown in Figures 6 and  
345 8.

### 346 3. Experiments

#### 347 3.1. Dataset

348 In this study we made use of sagittal magnetic resonance images (MRIs)  
349 of the spines from 26 cases, including 14 males and 12 females, with an age  
350 range of  $58 \pm 14$  years (mean  $\pm$  standard deviation). They were obtained  
351 from the clinical database of the Peking University Third Hospital. MR scans  
352 were performed on a 3.0 Tesla Siemens Trio scanner. Only the set of sagittal  
353 images acquired by using the fat-suppressed T2-weighted inversion recovery  
354 pulse sequence, in which the metastatic cancer was most clearly visible, was

355 analyzed. The imaging parameters were: TR = 4,780 ms, TE = 64 ms,  
356 TI = 200 ms, Echo Train length = 9, Field of View = 340 mm × 340 mm,  
357 slice thickness = 3 mm. The resolution of the images is 0.88 mm/pixel, 0.94  
358 mm/pixel, or 1.13 mm/pixel.

359 In this dataset, the primary cancers were: 15 lung, 5 thyroid, two liver,  
360 one breast, one prostate, one esophagus, one urinary tract. All patients  
361 were symptomatic with spinal pain but without any known cause; thus they  
362 were referred for diagnosis by MRI. Biopsies were subsequently taken from  
363 all patients and the diagnosis of metastatic cancer was confirmed through  
364 pathological examination of all biopsy specimens. For this initial study, all  
365 the cases under consideration were carefully selected based on confined dis-  
366 ease in only one or two vertebral body segments. Among these cases, the area  
367 of the smallest abnormal segment was  $1.65 \text{ cm}^2$ , the area of the largest abnor-  
368 mal segment was  $9.54 \text{ cm}^2$ , and the average area of the abnormal segments  
369 was  $5.07 \text{ cm}^2$ .

370 The metastatic lesions in each MRI slice were identified and manually  
371 traced by an experienced radiologist, and the boundary of the vertebral bod-  
372 ies were marked manually by two of the authors of this study. In each case  
373 there are a total of 13 sagittal images covering the entire spine and, depend-  
374 ing on the size of the lesion, the radiologist selected 3-7 slices that contained  
375 the lesion and manually outlined the metastatic lesions.

376 To facilitate the spinal metastasis detection, all the MRI slices are trans-  
377 formed into resolutions of 0.5 mm/pixel, 1 mm/pixel, and 2 mm/pixel for  
378 multi-resolution analysis. In the development phase, we also considered the  
379 multi-resolution analysis in the resolutions of 0.25 mm/pixel, 1 mm/pixel,

380 and 1.25 mm/pixel and the resolutions of 0.5 mm/pixel, 1 mm/pixel, and  
381 1.5 pixel, but got worse detection performance. The detection performance  
382 is evaluated at the resolution of 1 mm/pixel. To speed up the spinal metas-  
383 tasis detection, only the half center of the MRI slices along the x-axis of the  
384 images, containing the vertebral bodies, are considered [13].

### 385 3.2. Experimental setup

386 In our experiments, to get an overall evaluation performance for the whole  
387 dataset, we applied a case-based 10-fold cross-validation procedure as follows.  
388 The dataset was first randomly divided into 10 equally-sized subsets; in each  
389 run, a subset is hold out for performance evaluation (i.e. “testing subset”)  
390 and the remaining data is randomly partitioned into two subsets, one con-  
391 taining 75% of the examples for training (i.e. “training subset”) and the  
392 other containing 25% of the examples for validation (i.e. “validation sub-  
393 set”). To avoid any potential bias, the MRI slices from one case were always  
394 assigned together to either the training subset, or the validation subset, or  
395 the testing subset, but never spread across multiple subsets.

396 The Siamese neural network model was implemented by a Theano-based  
397 deep learning framework – Lasagne. It was trained by stochastic gradient  
398 descent (SGD) [40], with a batch size of 128, momentum of 0.9, and learning  
399 rate of 0.06. On average, there were 85,503 training examples in each run  
400 based on the sample extraction procedure described in Section 2.4.

401  $\lambda$  in (3) is an important parameter for controlling model complexity and  
402 the risk of overfitting during training. To determine the optimal parameter  
403  $\lambda$ , a grid search procedure was considered by evaluating the classification  
404 error on the samples in the validation set. Values of 0.001, 0.01, and 0.1 were

405 considered for  $\lambda$  in the grid search and a value of 0.01 is used in the final sys-  
406 tem. To obtain the samples for validation, the sample extraction procedure  
407 described in Section 2.4 was applied to the MRI slices in the validation set  
408 (described in Section 3.1).

### 409 3.3. Performance evaluation

410 For performance evaluation, we conduct a free-response receiver operating  
411 characteristic (FROC) analysis on the detection results. FROC analysis is  
412 widely used for detection performance evaluation in medical imaging [10, 41],  
413 and it provides an overall spinal metastasis detection performance evaluation  
414 with respect to the ground truth provided by the radiologists across all possi-  
415 ble decision thresholds. In particular, we compute and plot the FROC curve  
416 corresponding to the true-positive (TP) rate on the y-axis versus the number  
417 of FPs per case on the x-axis, in which the TP rate is calculated as the ratio  
418 of the number of TPs over the number of true spinal metastatic regions.

419 Very often we find cases where the cancer has spread outside the verte-  
420 brae resulting in soft tissue lesions with similar signal intensity as the cancer  
421 lesions inside the vertebrae. In this case, the automatic detection method de-  
422 veloped in this work would segment the entire lesion, both inside and outside  
423 the vertebrae. Therefore, in the FROC analysis, a detection is considered to  
424 be a TP if its center of gravity falls within 20 mm from the center of gravity  
425 of the region outlined by the radiologist; otherwise, it is considered to be a  
426 FP.

427 While the spinal metastasis detection is conducted for each MRI slice, the  
428 FROC analysis is done on a per-case basis as follows. For each case, its TP  
429 rate and the number of FPs is obtained as the average of the TP rate and

430 the average of the number of FPs, computed over the corresponding MRI  
431 sequence. For each MRI sequence, its TP rate and the number of FPs are  
432 obtained by averaging the TP rate and the number of FPs of all its slices.  
433 The purpose of such case-based FROC analysis is to accommodate for the  
434 fact that it is the same lesions that are under evaluation for the different  
435 slices of the different MRI sequences associated with a given case. The case-  
436 based analysis is used in general to avoid any potential bias introduced by  
437 differences in the number of MRI slices and MRI sequences, although in this  
438 particular study, only one MRI sequence is available for each case.

439 To accommodate for the variations associated with the distribution of  
440 the cases and to facilitate statistical comparison, we apply a bootstrapping  
441 procedure in the FROC analysis [41]. In our experiments, a total of 20,000  
442 bootstrap samples were used, based on which the p-value was reported on  
443 the detection results.

#### 444 **4. Results**

445 In Figure 5, we show the FROC curve obtained by the proposed approach,  
446 both with and without the aggregation procedure for comparison purpose.  
447 As can be seen, the FROC curve is noticeably improved when information  
448 from neighboring slides is aggregated. A statistical comparison between the  
449 proposed approach with and without the use of aggregation yields  $p$ -value  
450 of 0.0014 for FPs over the range of  $[0, 0.5]$  per case. Specifically, at TP rate  
451 of 90%, the FP rate is reduced from 0.375 per case without aggregation to  
452 0.207 per case with aggregation, nearly a 44.8% reduction. Moreover, with  
453 a FP rate of 0.20 per case, the sensitivity is improved from 72.7% without

454 aggregation to 89.1% with aggregation. Finally, it can be seen that the  
 455 proposed approach yields only 0.40 FPs per case when correctly detecting all  
 456 the spinal metastasis.

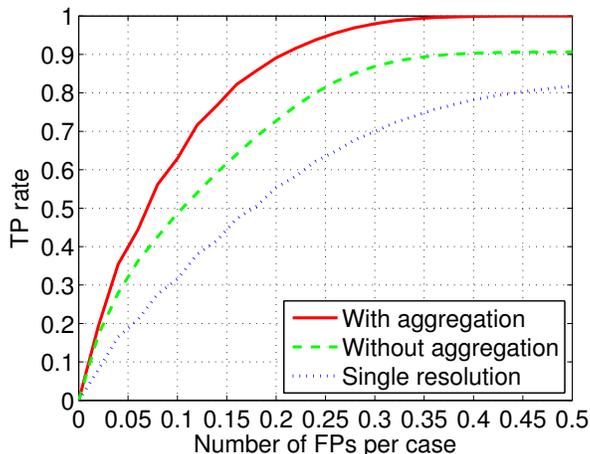


Figure 5: The detection performance achieved by the proposed approach (With aggregation). To demonstrate the effectiveness of the aggregation, the detection performance without the use of the aggregation is shown for comparison (Without aggregation). To demonstrate the effectiveness of the multi-resolution method, the detection performance achieved by using only one resolution (i.e. 1 mm/pixel) is shown as well.

457 In Figure 6, we show an example of three consecutive slices from an MRI  
 458 sequence (top) and their corresponding aggregated likelihood maps (bottom).  
 459 To demonstrate the corresponding detections, we apply a threshold of 0.6,  
 460 and the detection boundaries are marked by blue contours. For comparison  
 461 purposes, the boundaries of the metastatic lesions provided by the radiologist  
 462 are marked by red contours. As can be seen, the proposed approach using  
 463 the aggregated likelihood map correctly detects all the metastatic lesions.

464 For comparison, in Figure 7, we show the exact same example of three

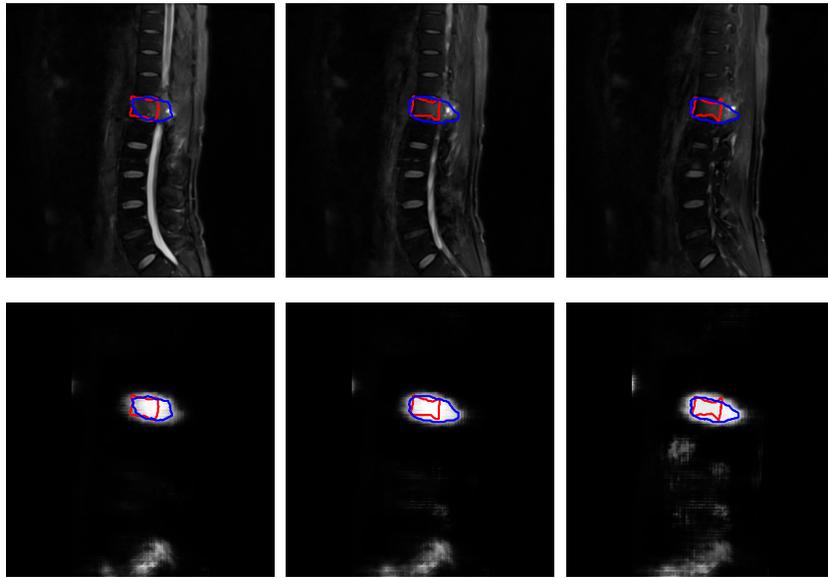


Figure 6: Examples of three consecutive slices in an MRI sequence (top) and their corresponding aggregated likelihood maps (bottom). The spinal metastasis boundaries provided by the radiologist are marked by red contours, while the boundaries of the detections obtained with a threshold of 0.6 are marked by blue contours.

465 consecutive slices, as in Figure 6 (top) and the corresponding likelihood maps  
 466 without using the aggregation strategy (bottom). As in Figure 6, the bound-  
 467 aries of the metastatic lesions provided by the radiologist are marked by red  
 468 contours, while the boundaries of the regions detected using a threshold of  
 469 0.6 are marked by blue contours. Compared with Figure 6, one FP con-  
 470 tour appears in the middle slice, clearly exemplifying how the aggregation  
 471 procedure can indeed reduce the FP rate in spinal metastasis detection.

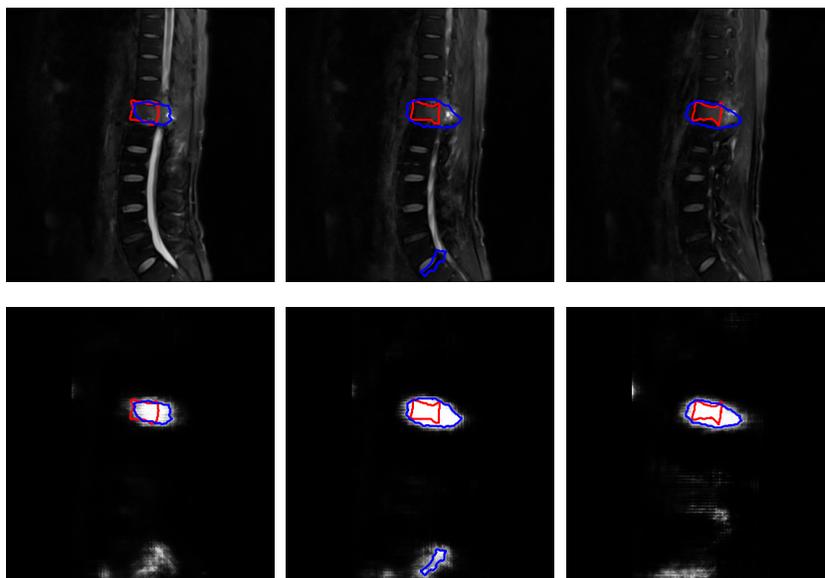


Figure 7: Example of three consecutive slices in an MRI sequence (top) and their corresponding likelihood maps without the use of aggregation (bottom). The spinal metastasis boundaries provided by the radiologist are marked by red contours, while the boundaries of the detections obtained with a threshold of 0.6 are marked by blue contours. In this case, the lack of aggregation results in the false-positive blue contour detection in the middle slide.

472 Finally, to demonstrate the effectiveness of the multi-resolution approach,  
 473 in Figure 5, for comparison purposes, we also show the best detection perfor-

474 mance achieved by the single resolution method, using the patch size  $55 \times 79$   
475 at image resolution 1 mm/pixel. The slice-based aggregation method was  
476 applied as well. As can be seen, the FROC curve of the proposed multi-  
477 resolution approach is noticeably higher than the single resolution method.  
478 A statistical comparison between them yields  $p$ -value of 0.0083 for FPs over  
479 the range of  $[0, 0.5]$  per case. Specifically, with a FP rate of 0.20 per case,  
480 the sensitivity is improved from 55.4% for single resolution method to 89.1%  
481 for multi-resolution approach.

## 482 5. Discussion

### 483 5.1. FP analysis

484 To better understand the occurrence of FPs in spinal metastasis detection,  
485 in Figure 8 we show another example of three consecutive slices from an MRI  
486 sequence (top) and the corresponding aggregated likelihood maps (bottom).  
487 In these plots, the boundaries of the spinal metastatic lesions provided by the  
488 radiologist are marked by red contours, while the boundaries of the regions  
489 detected using a threshold of 0.6 are marked by blue contours. As can be seen,  
490 there is one FP contour in the right slice of Figure 8. This FP corresponds  
491 to the cerebellum which typically has a moderate T2 signal intensity similar  
492 to that of a spinal lesion. Based on this example, we manually examined all  
493 the FPs detected at a threshold of 0.6, and found that most FPs (6 out of  
494 8 cases) occur in the cerebellum. Such FPs which have nothing to do with  
495 the spine can easily be removed if an algorithm for vertebral region detection  
496 is applied before spinal metastasis detection. Indeed, we found that only  
497 one FP in one slice of one MRI sequence happens at the vertebral region,

498 indicating that the proposed approach is indeed accurate in spinal metastasis  
499 detection.

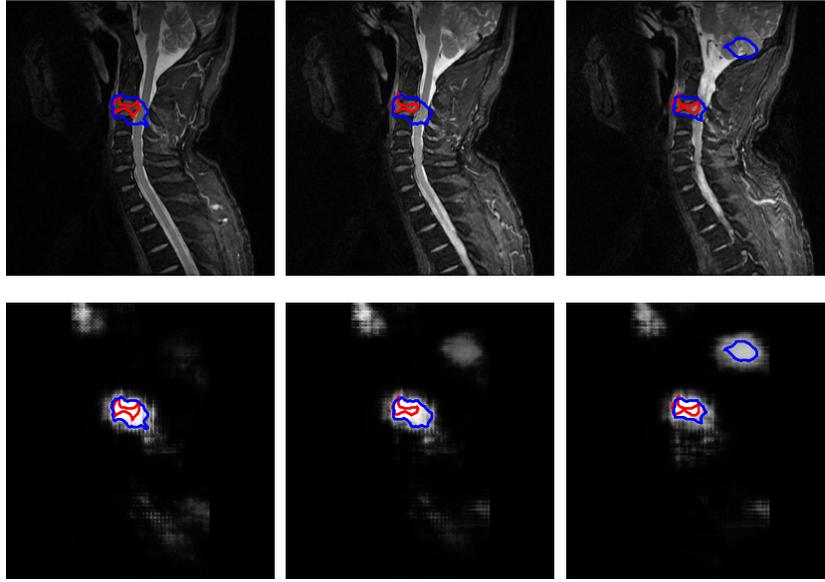


Figure 8: Example of three consecutive slices in an MRI sequence (top) and their corresponding aggregated likelihood maps (below). The spinal metastasis boundaries provided by the radiologist are marked by red contours, while the boundaries of the detections obtained with a threshold of 0.6 are marked by blue contours. In this case, there is a false-positive contour in the right slide associated with the cerebellum.

500 *5.2. Limitations*

501 The results from this study indicate that the proposed approach can de-  
502 tect spinal metastasis in MRI accurately. It has to be noted that all the cases  
503 considered in this study contain metastatic lesions. In the future, it might be  
504 desirable to include some normal cases (i.e. cases without metastatic lesions)  
505 as well. However, we note that in our dataset, the majority of local regions  
506 in the MRI sequences do not have any metastatic lesions, thus can be viewed

507 as substitutes for normal cases in terms of FP detection. Therefore, the re-  
508 ported number of FPs per case will likely remain very similar when normal  
509 cases are included. Nevertheless, the inclusion of normal cases should be  
510 helpful for more comprehensive evaluations, in particular for estimating pos-  
511 itive predictive values (PPV) and negative predictive values (NPV) in image  
512 screening.

513 The dataset used in this study consisted of only 26 cases. Although it  
514 is labor intensive, it would be obviously desirable in the future to curate  
515 larger data sets, containing more cases, to evaluate the proposed method  
516 more comprehensively. Furthermore, having larger training sets may also  
517 allow one to train more accurate Siamese networks with similar, or even more  
518 complex, architectures which may ultimately improve detection performance.

519 From the results in Section 4, the proposed method can accurately detect  
520 spinal metastases in MRI sequences. However, from Figures 6, 7 and 8, it can  
521 be seen that the boundaries of the regions detected by the proposed approach  
522 tend to be less accurate. Nevertheless, if the need were to arise in applica-  
523 tions for accurate metastatic boundary segmentation, any semi-automatic  
524 segmentation method—such as level set segmentation [42] or active contour  
525 model [43, 44]—could be included, as a post-processing step for refining the  
526 segmentation of the spinal metastatic lesions. These segmentation methods  
527 would benefit from the initialization of seed points provided by the proposed  
528 approach. It must be noted, however, that the development and evaluation  
529 of accurate segmentation methods would face even greater data challenges,  
530 as not only it would require more curated data, but the curation process itself  
531 would pose great challenges (e.g. a reader study). This is because different

532 radiologists may easily agree on the presence or absence of a lesion but, when  
533 present, may disagree on its precise boundaries.

## 534 **6. Conclusion**

535 In this study, we have investigated the feasibility of automatic spinal  
536 metastasis detection in MRI by using deep learning methods. For this pur-  
537 pose, we developed and implemented a multi-resolution approach using a  
538 Siamese convolutional neural network to accommodate for the large vari-  
539 ability in vertebral body size. The output of the Siamese neural network  
540 is further aggregated across neighboring slices in an MRI sequence to fur-  
541 ther reduce the FP rate. We have evaluated the detection performance on  
542 a set of 26 cases by FROC analysis. The results show that the proposed  
543 approach is effective and can correctly detect all the spinal metastatic le-  
544 sions while obtaining only 0.40 FPs per case. At TP rate of 90%, the use  
545 of the aggregation reduces the FPs from 0.375 FPs per case to 0.207 FPs  
546 per case, a nearly 44.8% reduction. Taken together, these results show that  
547 the proposed Siamese neural network with the aggregation strategy has the  
548 potential for providing the basis for an automated accurate spinal metastasis  
549 detection system that can be clinically deployed. The approach and its eval-  
550 uation will greatly benefit in the future from the aggregation and curation  
551 of larger datasets.

## 552 **Acknowledgment**

553 This research was in part supported by National Science Foundation grant  
554 IIS-1550705 and a Google Faculty Research Award to PB. In addition, the

555 work of MS was in part supported by NIH/NCI grant R01 CA127927 and  
556 P30 CA62203, the work of HY was in part supported by the National Natural  
557 Science Foundation of China (81471634), and the work of NL was in part sup-  
558 ported by the Beijing Natural Science Foundation (7164309). We also wish  
559 to thank Yuzo Kanomata for computing support and NVIDIA Corporation  
560 for a hardware donation.

## 561 **References**

- 562 [1] G. R. Mundy, Metastasis to bone: causes, consequences and therapeutic  
563 opportunities, *Nature Reviews Cancer* 2 (8) (2002) 584–593.
- 564 [2] M. J. Murphy, S. Chang, I. Gibbs, Q.-T. Le, D. Martin, D. Kim, Image-  
565 guided radiosurgery in the treatment of spinal metastases, *Neurosurgical*  
566 *focus* 11 (6) (2001) 1–7.
- 567 [3] T. F. Witham, Y. A. Khavkin, G. L. Gallia, J.-P. Wolinsky, Z. L.  
568 Gokaslan, Surgery insight: current management of epidural spinal cord  
569 compression from metastatic spine disease, *Nature clinical practice Neu-*  
570 *rology* 2 (2) (2006) 87–94.
- 571 [4] P. Klimo, M. H. Schmidt, Surgical management of spinal metastases,  
572 *The Oncologist* 9 (2) (2004) 188–196.
- 573 [5] L. M. Shah, K. L. Salzman, Imaging of spinal metastatic disease, *Inter-*  
574 *national journal of surgical oncology* 2011.
- 575 [6] D. M. Sciubba, R. J. Petteys, M. B. Dekutoski, C. G. Fisher, M. G.  
576 Fehlings, S. L. Ondra, L. D. Rhines, Z. L. Gokaslan, Diagnosis and man-

- 577        agement of metastatic spine disease: a review, *Journal of Neurosurgery:*  
578        *Spine* 13 (1) (2010) 94–108.
- 579 [7] T. Pope, H. L. Bloem, J. Beltran, W. B. Morrison, D. J. Wilson, *Mus-*  
580        *culoskeletal imaging*, Elsevier Health Sciences, 2014.
- 581 [8] P. R. Algra, J. L. Bloem, H. Tissing, T. Falke, J. Arndt, L. Verboom,  
582        *Detection of vertebral metastases: comparison between MR imaging and*  
583        *bone scintigraphy.*, *Radiographics* 11 (2) (1991) 219–232.
- 584 [9] G. J. O’Sullivan, F. L. Carty, C. G. Cronin, *Imaging of bone metastasis:*  
585        *An update*, *World journal of radiology* 7 (8) (2015) 202.
- 586 [10] H. R. Roth, J. Yao, L. Lu, J. Stieger, J. E. Burns, R. M. Summers, *De-*  
587        *tection of sclerotic spine metastases via random aggregation of deep*  
588        *convolutional neural network classifications*, in: *Recent Advances in*  
589        *Computational Methods and Clinical Applications for Spine Imaging*,  
590        Springer, 2015, pp. 3–12.
- 591 [11] T. Wiese, J. Yao, J. E. Burns, R. M. Summers, *Detection of sclerotic*  
592        *bone metastases in the spine using watershed algorithm and graph cut*,  
593        in: *SPIE Medical Imaging*, International Society for Optics and Pho-  
594        *tonics*, 2012, pp. 831512–831512.
- 595 [12] J. Yao, S. D. O’Connor, R. Summers, *Computer aided lytic bone metas-*  
596        *tasis detection using regular CT images*, in: *Medical Imaging*, Interna-  
597        *tional Society for Optics and Photonics*, 2006, pp. 614459–614459.
- 598 [13] J. Carballido-Gamio, S. J. Belongie, S. Majumdar, *Normalized cuts in*

- 599 3-d for spinal MRI segmentation, *IEEE transactions on medical imaging*  
600 23 (1) (2004) 36–44.
- 601 [14] S.-H. Huang, Y.-H. Chu, S.-H. Lai, C. L. Novak, Learning-based verte-  
602 bra detection and iterative normalized-cut segmentation for spinal mri,  
603 *IEEE transactions on medical imaging* 28 (10) (2009) 1595–1605.
- 604 [15] A. Neubert, J. Fripp, C. Engstrom, R. Schwarz, L. Lauer, O. Salvado,  
605 S. Crozier, Automated detection, 3D segmentation and analysis of high  
606 resolution spine MR images using statistical shape models, *Physics in*  
607 *medicine and biology* 57 (24) (2012) 8357.
- 608 [16] S. H. Zhou, I. D. Mccarthy, A. H. Mcgregor, R. R. Coombs, S. P. Hughes,  
609 Geometrical dimensions of the lower lumbar vertebrae—analysis of data  
610 from digitised CT images, *European Spine Journal* 9 (3) (2000) 242–248.
- 611 [17] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural*  
612 *Networks* 61 (2015) 85–117.
- 613 [18] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with  
614 deep convolutional neural networks, in: *Advances in neural information*  
615 *processing systems*, 2012, pp. 1097–1105.
- 616 [19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Er-  
617 han, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in:  
618 *Proceedings of the IEEE Conference on Computer Vision and Pattern*  
619 *Recognition*, 2015, pp. 1–9.
- 620 [20] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image  
621 recognition, arXiv preprint arXiv:1512.03385.

- 622 [21] A. Graves, A.-r. Mohamed, G. Hinton, Speech recognition with deep  
623 recurrent neural networks, in: Acoustics, Speech and Signal Processing  
624 (ICASSP), 2013 IEEE International Conference on, IEEE, 2013, pp.  
625 6645–6649.
- 626 [22] P. Baldi, P. Sadowski, D. Whiteson, Searching for exotic particles in  
627 high-energy physics with deep learning, Nature Communications.
- 628 [23] P. Sadowski, J. Collado, D. Whiteson, P. Baldi, Deep learning, dark  
629 knowledge, and dark matter, Journal of Machine Learning Research,  
630 Workshop and Conference Proceedings 42 (2015) 81–97.
- 631 [24] M. A. Kayala, P. Baldi, Reactionpredictor: Prediction of complex chem-  
632 ical reactions at the mechanistic level using machine learning, Journal  
633 of chemical information and modeling 52 (10) (2012) 2526–2540.
- 634 [25] A. Lusci, G. Pollastri, P. Baldi, Deep architectures and deep learning  
635 in chemoinformatics: the prediction of aqueous solubility for drug-like  
636 molecules, Journal of chemical information and modeling 53 (7) (2013)  
637 1563–1575.
- 638 [26] P. Di Lena, K. Nagata, P. Baldi, Deep architectures for protein contact  
639 map prediction, Bioinformatics 28 (19) (2012) 2449–2457.
- 640 [27] J. Zhou, O. G. Troyanskaya, Predicting effects of noncoding variants  
641 with deep learning-based sequence model, Nature methods 12 (10)  
642 (2015) 931–934.
- 643 [28] D. Cireş, A. Giusti, L. M. Gambardella, J. Schmidhuber, Deep neural  
644 networks segment neuronal membranes in electron microscopy images,

- 645 in: Advances in neural information processing systems, 2012, pp. 2843–  
646 2851.
- 647 [29] W. Shen, M. Zhou, F. Yang, C. Yang, J. Tian, Multi-scale convolutional  
648 neural networks for lung nodule classification, in: International Confer-  
649 ence on Information Processing in Medical Imaging, Springer, 2015, pp.  
650 588–599.
- 651 [30] D. Wang, A. Khosla, R. Gargeya, H. Irshad, A. H. Beck, Deep  
652 learning for identifying metastatic breast cancer, arXiv preprint  
653 arXiv:1606.05718.
- 654 [31] J. Wang, H. Ding, F. Azamian, B. Zhou, C. Iribarren, S. Molloy, P. Baldi,  
655 Detecting cardiovascular disease from mammograms with deep learning,  
656 IEEE Transactions on Medical Imaging.
- 657 [32] J. Bromley, J. W. Bentz, L. Bottou, I. Guyon, Y. LeCun, C. Moore,  
658 E. Säcker, R. Shah, Signature verification using a "siamese" time  
659 delay neural network, International Journal of Pattern Recognition and  
660 Artificial Intelligence 7 (04) (1993) 669–688.
- 661 [33] P. Baldi, Y. Chauvin, Neural networks for fingerprint recognition, Neural  
662 Computation 5 (3) (1993) 402–418.
- 663 [34] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep net-  
664 work training by reducing internal covariate shift, arXiv preprint  
665 arXiv:1502.03167.
- 666 [35] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks.,  
667 in: Aistats, Vol. 15, 2011, p. 275.

- 668 [36] B. Graham, Fractional max-pooling, arXiv preprint arXiv:1412.6071.
- 669 [37] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting.,  
670 Journal of Machine Learning Research 15 (1) (2014) 1929–1958.  
671
- 672 [38] P. Baldi, P. Sadowski, The dropout learning algorithm, Artificial intel-  
673 ligence 210 (2014) 78–122.
- 674 [39] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, Return of the  
675 devil in the details: Delving deep into convolutional nets, arXiv preprint  
676 arXiv:1405.3531.
- 677 [40] L. Bottou, Stochastic gradient descent tricks, in: Neural Networks:  
678 Tricks of the Trade, Springer, 2012, pp. 421–436.
- 679 [41] J. Wang, R. M. Nishikawa, Y. Yang, Improving the accuracy in detec-  
680 tion of clustered microcalcifications with a context-sensitive classifica-  
681 tion model, Medical physics 43 (1) (2016) 159–170.
- 682 [42] C. Li, C. Xu, C. Gui, M. D. Fox, Distance regularized level set evolution  
683 and its application to image segmentation, IEEE transactions on image  
684 processing 19 (12) (2010) 3243–3254.
- 685 [43] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active contour models,  
686 International journal of computer vision 1 (4) (1988) 321–331.
- 687 [44] S. Lankton, A. Tannenbaum, Localizing region-based active contours,  
688 IEEE transactions on image processing 17 (11) (2008) 2029–2039.