

# SCIENTIFIC REPORTS



OPEN

## Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning

Xinggong Wang<sup>1,2</sup>, Wei Yang<sup>3</sup>, Jeffrey Weinreb<sup>4</sup>, Juan Han<sup>5</sup>, Qiubai Li<sup>6</sup>, Xiangchuang Kong<sup>7</sup>, Yongluan Yan<sup>2</sup>, Zan Ke<sup>1</sup>, Bo Luo<sup>8</sup>, Tao Liu<sup>8</sup> & Liang Wang<sup>1,9</sup>

Prostate cancer (PCa) is a major cause of death since ancient time documented in Egyptian Ptolemaic mummy imaging. PCa detection is critical to personalized medicine and varies considerably under an MRI scan. 172 patients with 2,602 morphologic images (axial 2D T2-weighted imaging) of the prostate were obtained. A deep learning with deep convolutional neural network (DCNN) and a non-deep learning with SIFT image feature and bag-of-word (BoW), a representative method for image recognition and analysis, were used to distinguish pathologically confirmed PCa patients from prostate benign conditions (BCs) patients with prostatitis or prostate benign hyperplasia (BPH). In fully automated detection of PCa patients, deep learning had a statistically higher area under the receiver operating characteristics curve (AUC) than non-deep learning ( $P = 0.0007 < 0.001$ ). The AUCs were 0.84 (95% CI 0.78–0.89) for deep learning method and 0.70 (95% CI 0.63–0.77) for non-deep learning method, respectively. Our results suggest that deep learning with DCNN is superior to non-deep learning with SIFT image feature and BoW model for fully automated PCa patients differentiation from prostate BCs patients. Our deep learning method is extensible to image modalities such as MR imaging, CT and PET of other organs.

Prostate cancer (PCa) is a cause of death since ancient time documented in Egyptian Ptolemaic mummy imaging<sup>1</sup>. In the western countries, PCa is the most prevalent cancer and the third leading cause of cancer death in men<sup>2</sup>. In East Asian countries, the morbidity and mortality of PCa have increased dramatically, making it the fastest rising cancer among male malignancies. The accurate detection of PCa is an interesting but challenging task for physicians<sup>3</sup>. The distinction of PCa from prostate benign conditions (BC) including prostate benign hyperplasia (BPH) and prostatitis, is critical to personalized medicine<sup>4</sup>. Currently, MR Images of the prostate are manually evaluated by

<sup>1</sup>Department of Radiology, Tongji Hospital, Huazhong University of Science and Technology, Jiefang Road 1095, 430030, Wuhan, China. <sup>2</sup>School of Electronics Information and Communications, Huazhong University of Science and Technology, Luoyu Road 1037, Wuhan, Hubei, 430074, China. <sup>3</sup>Department of Nutrition and Food Hygiene, MOE Key Lab of Environment, Hubei Key Laboratory of Food Nutrition and Safety, Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Hangkong Road 13, 430030, Wuhan, China. <sup>4</sup>Department of Radiology and Biomedical Imaging, Yale University School of Medicine, New Haven, 208042, Connecticut, USA. <sup>5</sup>Department of Maternal and Child and Adolescent & Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Hangkong Road 13, 430030, Wuhan, China. <sup>6</sup>Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA, 02115, USA. <sup>7</sup>Department of Radiology, Union Hospital, Huazhong University of Science and Technology, Jiefang Road 1277, 430022, Wuhan, China. <sup>8</sup>School of mechanical science and engineering, Huazhong University of Science and Technology, Luoyu Road 1037, 430074, Wuhan, China. <sup>9</sup>Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science & Technology, Jie-Fang-Da-Dao 1095, Wuhan, 430030, P.R. China. Correspondence and requests for materials should be addressed to L.W. (email: [wang6@tjh.tjmu.edu.cn](mailto:wang6@tjh.tjmu.edu.cn))

radiologists. However, the detection of PCa using MR images varies considerably. In a recent meta-analysis for PCa detection the specificity was 0.73 (95% confidence interval [CI] 0.60–0.83) with sensitivity of 0.89 (95% CI 0.86–0.92)<sup>5</sup>.

Machine learning (ML) is a branch of artificial intelligence (AI) that adopts various optimization, probabilistic, and statistical tools to learn from past examples and to then employ that prior training to classify new data, predict novel trends or identify new patterns<sup>6–8</sup>. Imaging recognition technology is one of the core technologies of ML<sup>9</sup>.

Traditional image recognition methods, such as Support Vector Machine (SVM), Artificial Neural Networks (ANNs), Bayesian Networks (BNs), Decision Trees (DTs), k-nearest neighbors (k-NN), and Adaboost<sup>9–11</sup>, use hand crafted image features, such as texture, shape, density of pixels, and off-shelf classifiers. We name these traditional imaging recognition methods as non-deep-learning methods in this paper. The non-deep-learning methods have been contributed to medicine, such as breast cancer, neurodegenerative diseases, psychiatric diseases, and others<sup>7,8,11–14</sup>. However, the main limitation of non-deep-learning methods is that these methods depend on the feature-extraction step and it is hard to find and extract suitable image features for a specific medical-image recognition problem. If the feature is not effective enough, the off-shelf classifier is not able to obtain satisfactory results.

The use of AI technologies in medical imaging has becoming popular, especially with the development of deep learning<sup>15–27</sup>. To deal with image data, deep convolutional neural network (DCNN) is used<sup>28–33</sup>. Deep learning methods enable to learn adaptive image features and perform image classification simultaneously. Considering significant achievement of deep learning on image recognition, it is potential to apply deep learning to classify MR images for automated disease assistant-diagnosis. To the best of our knowledge, this is not yet a study to compare deep-learning and non-deep-learning methods to medical imaging for the automatic classification of PCa or BCs patients with MR images.

The objective of the study is to search for PCa patients through MR imaging classification using deep learning versus non-deep learning. The choice of non-deep-learning algorithm is one challenge of this study. As a matter of fact, there have been many computer-aided diagnosis (CAD) algorithms for the prostate in MRI<sup>7</sup>. However, most of the CAD algorithms do not have source code and implementation details<sup>7</sup>. It is impossible to make comparison between studies without source code and implementation details. Thus, for a fair comparison we chose SIFT image feature with bag-of-words (BoW) model, a representative method in image recognition and analysis. In computer vision domain, the BoW was the representative method for CAD before the era of deep learning. Please refer to the image-net 2012 results<sup>34</sup>.

We compare a deep learning with DCNN algorithm and a non-deep learning with SIFT image feature and BoW model algorithm on building classifiers to distinguish PCa patients with prostate BCs patients using morphologic images (axial 2D T2-weighted imaging) of the prostate. We carry out a patient-based PCa and BCs classification, which only require patient-level labeling of the PCa or BCs patients and the corresponding images rather than segmenting the lesions on images in the learning algorithm. The gold standard used to assign the labels to the training images was the pathological diagnosis. The deep learning with DCNN method obtains significantly better than the non-deep learning with SIFT image feature and BoW model method. The deep learning with DCNN method may ultimately enable imaging interpretation to be easier and assist novices or general physicians to increase diagnostic accuracy, efficiency and consistency, and contribute to precision medicine.

## Results

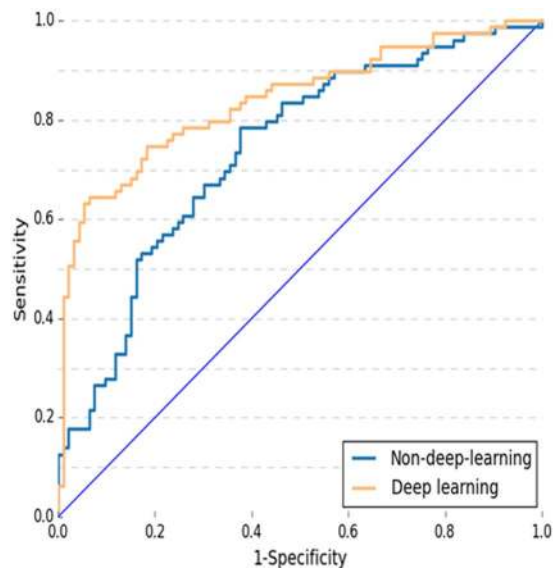
To validate the effectiveness of the proposed deep learning method, comparison was made between deep learning with DCNN versus non-deep learning with SIFT image feature and BoW model in area under receiver operating characteristic curve (AUC), sensitivity, specificity, PPV and NPV. A  $P < 0.05$  was considered statistically significant. Figure 1 and Table 1 provide overall PCa patients differentiation statistics of deep learning with DCNN versus non-deep learning with SIFT image feature and BoW model. The AUCs in deep learning with DCNN and non-deep learning with SIFT image feature and BoW model were 0.84 (95% CI 0.78–0.89) and 0.70 (95% CI 0.63–0.77) respectively. There was a significant difference between the AUCs in deep learning with DCNN and non-deep learning with SIFT image feature and BoW model ( $P = 0.0007 < 0.001$ ). At the cut-off value above 0.5, sensitivity was 69.6%, specificity was 83.9%, positive prediction value (PPV) was 78.6%, and negative prediction value (NPV) was 76.5% for deep learning with DCNN. At the same cut-off value, sensitivity was 49.4%, specificity was 81.7%, PPV was 69.6%, and NPV was 65.5% for non-deep learning with SIFT image feature and BoW model.

## Discussion

It is one of the first studies applying deep learning with DCNN as well as non-deep-learning to medical imaging. The novel study is a patient-based formulation for the automatic classification of PCa or prostate BCs patients with morphologic images (axial 2D T2-weighted imaging), in which a deep learning method and a non-deep learning method were compared. We develop a structure and workflow of deep learning that classifies PCa and prostate BCs patients based on the prostate morphologic images (axial 2D T2-weighted imaging), and compare deep learning classifiers with DCNN versus non-deep-learning with SIFT image feature and BoW model in the detection of PCa or prostate BCs patients.

Non-deep-learning methods<sup>9–11</sup>, such as SVM, ANNs, BNs, DTs, k-NN, and Adaboost, basically depend on a feature extraction step which usually describes images using texture, gradient and Gabor filters, etc. In contrast, it is a merit of our method that the proposed deep-learning method learns image features automatically within deep networks.

For a fair comparison, we chose SIFT features and BoW model as the representative of the so-called non-deep learning methods for image recognition. We extracted encoded the SIFT features using a pre-trained dictionary



**Figure 1.** The ROC curves for PCa and prostate BCs patients differentiation of non-deep learning with SIFT image feature and BoW model versus deep learning with deep convolutional neural network (DCNN). Note: ROC curve: receiver operating characteristic curve. AUC: area under ROC. PCa = prostate cancer prostate BCs = prostate benign conditions BPH = benign prostatic hyperplasia.

| Predictive test   | AUC mean (95% CI)* | N (%) of PCa patient correctly classified |               |               |                |
|-------------------|--------------------|---|---------------|---------------|----------------|
|                   |                    | Sensitivity                               | Specificity   | PPV           | NPV            |
| Deep learning     | 0.84 (0.78–0.89)   | 69.6% (55/79)                             | 83.9% (78/93) | 78.6% (55/70) | 76.5% (78/102) |
| Non-deep-learning | 0.70 (0.63–0.77)   | 49.4% (39/79)                             | 81.7% (76/93) | 69.6% (39/56) | 65.5% (76/116) |

**Table 1.** Overall PCa patient differentiation statistics of deep learning with DCNN versus non-deep learning with SIFT image feature and BoW model method. Note: \* $P = 0.0007 < 0.001$ . Criterion: cut-off value  $> 0.5$ . CI: confidence Interval. PPV: positive prediction value. NPV: negative prediction value. ROC curve: receiver operating characteristic curve. AUC: area under ROC. PCa = prostate cancer.

to generate a vector image representation. We applied a linear SVM classifier on the vector representation to obtain image classification. The SIFT features were extracted from the augmented images, which were generated using the same data augmentation method as we used in the deep learning method.

In our study, the training data and test data were better partitioned in terms of “subject”, i.e., patient. In this case, there were 172 subjects. In comparison to the recent ML papers on prostate studies, there were 147 subjects<sup>9</sup> and 66 T2-weighted prostate images<sup>35</sup>, respectively. Even though this study was based on the ImageNet model, incorporating the features from the natural images, it should be noted that MR images are different from the nature images. It means that it may require a much larger sample size than 172. To make the subject sufficient for the study, our experiments were carried out according to 10-fold cross validation method published by Beheshti I<sup>36</sup> and Ying J<sup>37</sup>. The method can ensure the robustness of our classifiers and the independence between the training data and test data. The pre-trained ImageNet model with augmented data was fine-tuned. As shown in the results, our deep learning classifier obtained robust performance. From the perspective of deep learning mechanism, the DCNN may work as a feature extractor, which provided a better feature for MR images. The final layer is a linear classifier which is equivalent to the classifiers in the non-deep learning method. As shown in the result, our method of deep learning from 172 subjects is feasible and interesting.

Deep learning that classifies image requires small-size images as the input of DCNN. A DCNN model that takes small-size image enables faster training and testing speed. We used the method published by Samala<sup>33</sup>, Rajkomar<sup>32</sup> to resize each image with cubic interpolation. While preserving the original label during the training of deep learning classifier, we cropped each training image of  $360 \times 360$  into multiple sub-images of  $288 \times 288$ . The method is popular in computer vision domain.

In this study, we showed that convolution features learnt from morphologic images (axial 2D T2-weighted imaging) of the prostate were used to classify PCa or prostate BCs patients. These convolution features were basically unrecognized by human inspection. However, the machine could efficiently and effectively recognize the convolution features and classified PCa or prostate BCs patients (Supplementary).

The deep learning system proposed in this paper relies on high-efficiency computational resources. Our experiments were carried out on a workstation with two Nvidia Titan X GPUs, an 8-cores Intel i7 CPU and 32GB memory. With these powerful computational resources, it took only 10 minutes to train a DCNN model for 1000 iterations and took less than 0.5 second to test a patient.

| Characteristics              | Summary         |
|------------------------------|-----------------|
| PCa patients                 | N = 79          |
| Age (years) (average, range) | 67.9(50–88)     |
| Number of MR imaging *       | N = 1164        |
| Prostate BCs patients        | N = 93          |
| Age (years) (average, range) | N = 66.5(47–91) |
| Number of MR imaging*        | N = 1438        |
| BPH                          | N = 75          |
| BPH + prostatitis            | N = 18          |

**Table 2.** Patient Characteristics. Note: PCa = prostate cancer. prostate BCs = prostate benign conditions. BPH = benign prostatic hyperplasia. \*Morphologic images (axial 2D T2-weighted imaging). The study is a patient-based to compare deep-learning with DCNN versus non-deep-learning with SIFT image feature and BoW model for the automatic classification of PCa or prostate BCs patients with morphologic images (axial 2D T2-weighted imaging).

As described in the result section, the AUCs of deep learning with DCNN and non-deep learning with SIFT image feature and BoW model were 0.84 (95% CI 0.78–0.89) and 0.70 (95% CI 0.63–0.77) respectively. There was a significant difference between the AUCs in deep learning with DCNN and non-deep learning with SIFT image feature and BoW model ( $P = 0.0007 < 0.001$ ). The sensitivity was 69.6%, specificity was 83.9%, positive prediction value (PPV) was 78.6%, the negative prediction value (NPV) was 76.5% for deep learning with DCNN. The sensitivity was 49.4%, specificity was 81.7%, PPV was 69.6%, NPV was 65.5% for non-deep learning with SIFT image feature and BoW model. Thus our deep learning with DCNN is able to produce better differentiation performance than non-deep learning with SIFT image feature and BoW model. Deep learning with DCNN has better capability in image recognition of prostate MRI than non-deep-learning with SIFT image feature and BoW model.

There are the limitations of the study. First, we use more than one image from each patient assuming independence among them, and classify PCa or prostate BCs patients with the images individually. Currently, in the research of computer vision and deep learning, understanding 3D data still lacks of effective methods. Analyzing 3d data in a frame by frame manner is the most popular way. In the future work, we will work on developing more effective deep learning methods by utilizing the dependence of the images from each patient for better PCa detection. Second, our study is a patient-based to compare deep-learning with non-deep-learning for the automatic classification of PCa or prostate BCs patients with morphologic images (axial 2D T2-weighted imaging). There was no manual work such as lesion-based label annotations or segmentation in our study. We believe that it is a merit of this paper. It has a clear advantage and makes sense in computer vision and deep learning. The advantage is that it does not require image label annotations or segmentations manually. Nevertheless, the patient-based study cannot provide information regarding tumor location on images. Third, our study merely used morphologic images (axial 2D T2-weighted imaging) of the prostate. The use of adding quantitative features from functional images including DWI and DCE-MRI remains to be investigated. Fourth, we merely focus on the comparison between deep learning with DCNN and non-deep-learning with SIFT image feature and BoW model. The comparison between deep learning and radiologist remains to be investigated. Nevertheless, the objective of our study is to determine which one of two methods, deep learning with DCNN versus non-deep learning with SIFT image feature and BoW model, is better in the detection of PCa. We found deep learning with DCNN is able to produce better differentiation performance than non-deep learning with SIFT image feature and BoW model in the detection of PCa patients.

In summary, we elucidated that prostate MR image machine learning classifiers in terms of morphological information can successfully search for PCa patients. This deep learning with DCNN method outperforms non-deep-learning with SIFT image feature and BoW model for the automatic classification of MR images of PCa or BCs patients. Our prostate MR images obtained from routine clinical practice, our machine learning classifiers could be efficiently implemented into current clinical practice. Accurate diagnosis provided by our deep learning with the DCNN method can contribute to precision medicine.

## Methods

**Subject.** The research protocol for this retrospective study was approved by the institutional review board of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology. All methods were performed in the principles of the Declaration of Helsinki. Written informed consent was obtained from the subject. 172 patients with 2,602 morphologic images (axial 2D T2-weighted imaging) of the prostate were obtained from Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, with pathologically confirmed PCa and prostate BCs including BPH and prostatitis. The patient's cohort included 93 prostate BCs and 79 PCa patients. Patient characteristics of the cohort are summarized in Table 2.

**MR Image Acquisition.** All MR images were performed by using a 3.0 Tesla (T) whole-body unit (MAGNETOM Skyra, Siemens Medical Solutions, Erlangen, Germany) followed by pathology reports and one year's follow-up. The acquisition protocol included T2-weighted imaging (T2WI), T1-weighted imaging (T1WI), dynamic contrast-enhanced MRI (DCE-MRI) and diffusion-weighted imaging (DWI). The final cohort size was

| No. of group                                 |     | 1          | 2          | 3          | 4          | 5          | 6          | 7          | 8          | 9          | 10         |
|--|-----|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Training (No. of patients and No. of images) | All | 155 (2342) | 155 (2352) | 154 (2310) | 155 (2342) | 155 (2318) | 155 (2372) | 155 (2371) | 154 (2324) | 155 (2350) | 155 (2337) |
|  | BC  | 84 (1295)  | 82 (1272)  | 84 (1283)  | 81 (1252)  | 86 (1314)  | 83 (1298)  | 88 (1370)  | 81 (1252)  | 82 (1276)  | 86 (1330)  |
|  | PCa | 71 (1047)  | 73 (1080)  | 70 (1027)  | 74 (1090)  | 69 (1004)  | 72 (1074)  | 67 (1001)  | 73 (1072)  | 73 (1074)  | 69 (1007)  |
| Testing (No. of patients and No. of images)  | All | 17 (260)   | 17 (250)   | 18 (292)   | 17 (260)   | 17 (284)   | 17 (230)   | 17 (231)   | 18 (278)   | 17 (252)   | 17 (265)   |
|  | BC  | 9 (143)    | 11 (166)   | 9 (155)    | 12 (186)   | 7 (124)    | 10 (140)   | 5 (68)     | 12 (186)   | 11 (162)   | 7 (108)    |
|  | PCa | 8 (117)    | 6 (84)     | 9 (137)    | 5 (74)     | 10 (160)   | 7 (90)     | 12 (163)   | 6 (92)     | 6 (90)     | 10 (157)   |

**Table 3.** Training data, testing data of 10-fold cross validation on PCa and BCs patients differentiation experiments of deep learning with deep convolutional neural network (DCNN) versus non-deep-learning with SIFT image feature and BoW model. Note: PCa = prostate cancer, prostate BCs = prostate benign conditions.

172 patients with 2,602 morphologic images (axial 2D T2-weighted imaging) of the prostate used in our analysis (Table 2). The acquisition parameters for obtaining the 2-dimensional transverse T2WI turbo spin-echo images were set as follows: repetition time (TR) 6750 ms, echo time (TE) 104 ms, echo train length 16, section thickness 3 mm, no intersection gap, field of view (FOV) 180 mm × 180 mm, image matrix 384 × 384. Most CNN networks require square images. In the paper, we resize the image to 360 × 360 before input the image into deep networks. The network we designed requires input image in this size. The common practice in deep learning helps to improve the speed of deep learning classifier.

The method published by Samala<sup>33</sup>, Rajkomar<sup>32</sup> was applied to resize each image with cubic interpolation. The method could ensure accelerated our training and testing DCNN, while preserving the original label during the training of deep learning classifier. Each image of 360 × 360 was cropped into multiple sub-images of 288 × 288. Imaging processing and imaging analysis used all the images. For convenience, we set the images to have equal weight. The final classification result was obtained via a voting strategy formatted in Equation  $P_i = \frac{1}{M_i} \sum_{j=1}^{M_i} P_{ij}$ .

**Clinical variables and pathology reports.** The clinical variables and pathology reports of each PCa patients and prostate BCs patients were acquired from our research database (Table 2).

**Cross-Validation.** For the robustness of our classifiers, a 10-fold cross validation method published by Beheshti I<sup>36</sup> and Ying J<sup>37</sup> was applied to randomly partition the 172 patients into 10 equal sized groups to ensure the independence between the training data and test data, as described in Table 3.

**Formulation of Image Classification using Deep Learning with DCNN.** To describe the MR image classification using deep learning in the patient-based study, we gave the formulations in this section. At first, we denoted a patient as  $X_i = \{x_{ij}, j \in [1, \dots, M_i]\}$ , where  $X_i, i \in [1, \dots, N]$  denoted the  $i$ -th patient, there were  $N$  patients in total, the  $i$ -th patient contains  $M_i$  images, and  $x_{ij}$  was the  $j$ -th image of the  $i$ -th patient. Then, each patient had a clinical label. Accordingly, each patient was denoted as  $X_i$  which means that we treated all images of a patient as a whole, each lesion of the patient will be either used for training or for testing in one fold of the experiments.

For the patient  $X_i$ , its clinical label was denoted as  $Y_i$ . We defined that if the  $i$ -th patient was PCa,  $Y_i = 1$  and if the  $i$ -th patient was BCs,  $Y_i = 2$ .

Training a deep learning model required the label of each image, which meant we needed the label of  $x_{ij}$  denoted as  $y_{ij}$ . To address the issue, we set the image label as its patient label, denoted as  $y_{ij} = Y_i$ .

After obtaining the labels of all training images, we trained a deep CNN model using back-propagation in the Caffe deep learning toolbox. The learned deep CNN model was denoted as  $f$ .

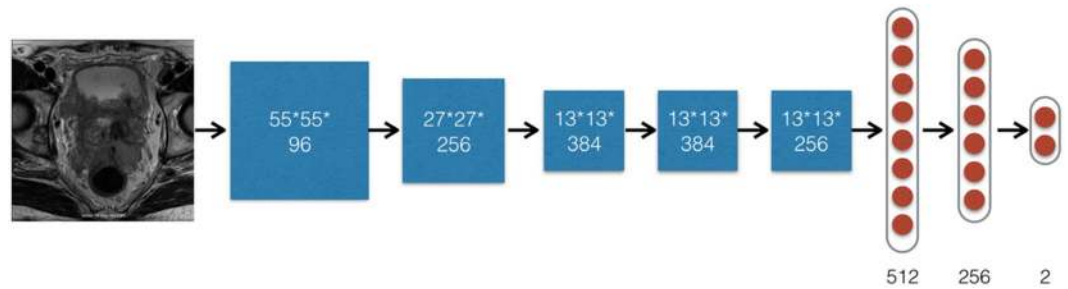
In testing phase, we used the learned deep CNN model  $f$  to compute the classification probability  $p_{ij}$  of image  $x_{ij}$  denoted as  $f(x_{ij})$ .  $P_{ij}$  was a vector denoted as  $P_{ij} \in R^{K \times 1}$ , where  $K$  was the number of classes. In this study,  $K = 2$ ,  $p_{ij}(1)$  was the probability  $x_{ij}$  belonged to prostate PCa, and  $p_{ij}(2)$  was the probability  $x_{ij}$  belonged to BCs, and the sum of  $p_{ij}$  was 1 which was restricted in by the deep CNN structure.

After obtaining the classification result of each PCa or prostate BC patient with images, we could easily obtain the classification result of a patient by averaging its images classification results, denoted as  $P_i = \frac{1}{M_i} \sum_{j=1}^{M_i} P_{ij}$ .  $P_i$  was a  $K = 2$  dimensional vector and its sum was equal to 1. Finally, we could get classification label of the patient  $X_i$  by checking which dimension had larger probability. If the first dimension had large value, it was classified as PCa patient; otherwise, it was classified as BC patient.

**Deep Learning Toolbox.** In this study, a popular deep learning toolbox created by Berkeley Vision and a learning center, named Caffe, was utilized for implementing the automated PCa and BCs patient differentiation system (Supplementary). Python for data preparation, analysis and visualization was used. The hyper-parameters of our DCNN were set as follows: gamma 0.1, momentum 0.9, weight decay 0.1, and maximum training iteration 1000 (Supplementary).

**Training Deep Learning Classifier.** The method published by Samala<sup>33</sup>, Rajkomar<sup>32</sup> to resize each image with cubic interpolation was applied. A pre-trained ImageNet model was used for fine-tuning our model with augmented data. In data augmentation, each training image of 360 × 360 was cropped into multiple sub-images





**Figure 2.** The structure of deep learning with deep convolutional neural network (DCNN) for the automatic classification of a PCa or BCs patient with morphologic images (axial 2D T2-weighted imaging). A  $288 \times 288 \times 3$  MR image was input. Five convolution layers and two inner product layers with sizes were shown in the figure. A max-pooling layer and non-linear ReLU layer following each convolution layer. A max-pooling layer downsize feature map gradually as demonstrated. Finally, an output layer specified PCa patient probability on input image. Note: PCa = prostate cancer. Prostate BCs = prostate benign conditions. BPH = benign prostatic hyperplasia.

of  $288 \times 288$  while preserving the original label during the training of deep learning classifier. The deep learning classifiers were used to conduct supervised machine learning. The inputs to the classification algorithms were the prostate morphologic images (axial 2D T2-weighted imaging). There were 5 convolutional layers and 2 inner product layers. After every convolutional layer, there was a max-pooling layer and a non-linear ReLU layer, respectively. Thus, different from the non-deep learning, which depends on the feature extraction step, DCNN automatically learnt from image features using the image labels given by the pathological diagnosis. The learned features were adaptive and more suitable for solving the problem without of human design. The outputs specified the predicted diagnosis groups. The hyper-parameters were set based on our experiences of tuning deep networks for image recognitions are given in Fig. 2.

**Formulation of Image Classification using Non-deep-learning Method.** We used a non-deep-learning method named as bag-of-words (BoW) with SIFT image feature, a representative method for image recognition. SIFT image feature is popular for its scale and rotation invariance, it builds local image representation based on image gradient. For each image, a set of SIFT features were extracted and encoded using a pre-trained dictionary. The BoW model aggregated the encoded SIFT features into a vector representation for each image. Image classification was done with a linear SVM classifier on the vector representation. In addition, for a fair comparison between deep learning and non-deep-learning, the SIFT features were extracted from the augmented images, which were generated using the same data augmentation method as we used in the deep learning method. Thus, for deep learning and non-deep-learning, the data augmentation method was the same. Then, we gave the formulation of BoW using SIFT feature as follows. We denoted the SIFT features extracted from MR image  $x_{ij}$  as  $S_{ij}$ .  $S_{ij}$  were a set of feature vectors. Then we applied the Fisher vector coding<sup>38</sup> to aggregate the local image descriptors into a vector representation of image  $x_{ij}$ , denoted as  $F_{ij} = FV(S_{ij})$ .

Based on the vector representations, we used SVM as the classifier of the studied non-deep-learning method, which was learned from training images and applied to classify testing images. For each training image, SVM classifier gave its probability of being PCa patient. We subsequently used the same equation for deep learning method (in Eq.  $P_i = \frac{1}{M_i} \sum_{j=1}^{M_i} p_{ij}$ ) to calculate the probability of a PCa patient.

**Non-Deep-Learning Toolbox.** To implement the proposed non-deep-learning method, the VLFeat toolbox was used to extract SIFT feature and Fisher vector coding and the LibLinear toolbox to do SVM training and testing. We used the same data augmentation method used in the deep learning method to process the training images. The hyper-parameters were given as follows: the number of clustering center of Fisher vector coding was set to 64. The bias term of linear SVM was set to be true.

**Training data, testing data of 10-fold cross validation on the PCa and prostate BCs patients differentiation experiments with deep learning versus non-deep-learning.** For the robustness of our classifiers, we applied a 10-fold cross validation method published by Beheshti<sup>36</sup>, Ying<sup>37</sup> to randomly partition the 172 patients into 10 equal sized groups. Specifically, as described in Table 3, the 10-fold cross-validation method ensured the independence between the training data and test data for deep learning with DCNN versus non-deep-learning with SIFT image feature and BoW model. Please refer to the source code in the Supplementary.

**Assessment.** For diagnostic classification to differentiate PCa patients from prostate BCs patients based on deep learning with DCNN versus non-deep-learning with SIFT image feature and BoW model, ROC curves were performed and AUCs of every classifiers based on deep learning with DCNN versus non-deep-learning with SIFT image feature and BoW model were generated.

## References

- Prates, C., Sousa, S., Oliveira, C. & Ikram, S. Prostate metastatic bone cancer in an Egyptian Ptolemaic mummy, a proposed radiological diagnosis. *International Journal of Paleopathology* **1**, 98–103 (2011).
- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics, 2017. *CA: a cancer journal for clinicians* **67**, 7–30 (2017).
- Weinreb, J. C. *et al.* PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *European urology* **69**, 16–40 (2016).
- Herold, C. J. *et al.* Imaging in the Age of Precision Medicine: Summary of the Proceedings of the 10th Biannual Symposium of the International Society for Strategic Studies in Radiology. *Radiology* **279**, 226–238 (2016).
- Woo, S., Suh, C. H., Kim, S. Y., Cho, J. Y. & Kim, S. H. Diagnostic Performance of Prostate Imaging Reporting and Data System Version 2 for Detection of Prostate Cancer: A Systematic Review and Diagnostic Meta-analysis. *European urology* (2017).
- Mitchell, T. Machine Learning. *McGraw* (1997).
- Trebeschi, S. *et al.* Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric MR. *Sci Rep-Uk* **7** (2017).
- Oakden-Rayner, L. *et al.* PrecisionRadiology: Predicting longevity using feature engineering and deep learning methods in a radiomics framework. *Sci Rep-Uk* **7** (2017).
- Fehr, D. *et al.* Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E6265–E6273 (2015).
- Erickson, B. J., Korfiatis, P., Akkus, Z. & Kline, T. L. Machine Learning for Medical Imaging. *Radiographics: a review publication of the Radiological Society of North America, Inc* **37**, 505–515 (2017).
- Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G. & Mechelli, A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev* **36**, 1140–1152 (2012).
- Cruz, J. A. & Wishart, D. S. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform* **2**, 59–77 (2007).
- Salvatore, C., Battista, P. & Castiglioni, I. Frontiers for the Early Diagnosis of AD by Means of MRI Brain Imaging and Support Vector Machines. *Curr Alzheimer Res* **13**, 509–533 (2016).
- Doyle, S. H. M., Shah, K., Madabhushi, A., Feldman, M., Tomas, J. Automated grading of prostate cancer using architectural and textural image features. (Biomedical Imaging: From Nano to Macro, 2007 ISBI 2007 4th IEEE. International Symposium on: IEEE, 2007, p1284-7).
- Wang, J. *et al.* Detecting Cardiovascular Disease from Mammograms with Deep Learning. *IEEE transactions on medical imaging* (2017).
- Shen, D., Wu, G. & Suk, H. I. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng* (2017).
- Lee, H. *et al.* Fully Automated Deep Learning System for Bone Age Assessment. *J Digit Imaging* (2017).
- Gao, X. W., Hui, R. & Tian, Z. Classification of CT brain images based on deep learning networks. *Comput Methods Programs Biomed* **138**, 49–56 (2017).
- Forsberg, D., Sjoblom, E. & Sunshine, J. L. Detection and Labeling of Vertebrae in MR Images Using Deep Learning with Clinical Annotations as Training Data. *J Digit Imaging* (2017).
- Zhang, Q. *et al.* Deep learning based classification of breast tumors with shear-wave elastography. *Ultrasonics* **72**, 150–157 (2016).
- Sirinukunwattana, K. *et al.* Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. *IEEE transactions on medical imaging* (2016).
- Ortiz, A., Munilla, J., Gorriz, J. M. & Ramirez, J. Ensembles of Deep Learning Architectures for the Early Diagnosis of the Alzheimer's Disease. *Int J Neural Syst* **26**, 1650025 (2016).
- Nie, D., Zhang, H., Adeli, E., Liu, L. & Shen, D. 3D Deep Learning for Multi-modal Imaging-Guided Survival Time Prediction of Brain Tumor Patients. *Med Image Comput Assist Interv* **9901**, 212–220 (2016).
- Gulshan, V. *et al.* Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **316**, 2402–2410 (2016).
- Cheng, J. Z. *et al.* Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. *Sci Rep* **6**, 24454 (2016).
- Ithapu, V. K. *et al.* Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment. *Alzheimers Dement* **11**, 1489–1499 (2015).
- Plis, S. M. *et al.* Deep learning for neuroimaging: a validation study. *Front Neurosci* **8**, 229 (2014).
- Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A. & Mougiakakou, S. Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE transactions on medical imaging* **35**, 1207–1216 (2016).
- Cha, K. H. *et al.* Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets. *Medical physics* **43**, 1882 (2016).
- Ghafoorian, M. *et al.* Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin. *Neuroimage Clin* **14**, 391–399 (2017).
- Lekadir, K. *et al.* A Convolutional Neural Network for Automatic Characterization of Plaque Composition in Carotid Ultrasound. *IEEE J Biomed Health Inform* **21**, 48–55 (2017).
- Rajkomar, A., Lingam, S., Taylor, A. G., Blum, M. & Mongan, J. High-Throughput Classification of Radiographs Using Deep Convolutional Neural Networks. *J Digit Imaging* **30**, 95–101 (2017).
- Samala, R. K. *et al.* Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. *Medical physics* **43**, 6654 (2016).
- Deng, J. E. A. Imagenet large scale visual recognition challenge 2012. [www.image-net.org/challenges/LSVRC/2012](http://www.image-net.org/challenges/LSVRC/2012). 1–42 (2012).
- Guo, Y., Gao, Y. & Shen, D. Deformable MR Prostate Segmentation via Deep Feature Learning and Sparse Patch Matching. *IEEE transactions on medical imaging* **35**, 1077–1089 (2016).
- Beheshti, I. *et al.* Structural MRI-based detection of Alzheimer's disease using feature ranking and classification error. *Comput Methods Programs Biomed* **137**, 177–193 (2016).
- Ying, J. *et al.* Classification of Exacerbation Frequency in the COPD Gene Cohort Using Deep Learning with Deep Belief Networks. *IEEE J Biomed Health Inform* (2016).
- Sánchez, J. P. F. & Mensink, T. *et al.* Image classification with the fisher vector: Theory and practice. *Int. J. Comput. Vis* **105**, 222–245 (2013).

## Acknowledgements

This work was supported in part by grants from the National Natural Science Foundation of China (grant # 81171307, 81671656).

## Author Contributions

All the authors, X.W., W.Y., J.W., J.H., Q.L., X.C.K., Y.L.Y., Z.K., B.L., T.L., L.W. conceived, designed, performed the analyses, interpreted the results. All the authors, X.W., W.Y., J.W., J.H., Q.L., X.C.K., Y.L.Y., Z.K., B.L., T.L., L.W. wrote the main manuscript text and X.G.W., J.W., L.W. T.L., Q.L., prepared figures. L.W., X.G.W. and T.L. supervised the work.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-15720-y>.

**Competing Interests:** The authors declare that they have no competing interests.

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



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

© The Author(s) 2017