Zhongxiao Wang, Tianyu Xiong, Mingxin Jiang, Yun Cui, Xiaosong Qian, Yao Su, Xiaolei Zhang, Shiqi Xu, Dong Wen, Xianling Dong*, Minfu Yang* and Yinong Niu*

Automatic prediction model of overall survival in prostate cancer patients with bone metastasis using deep neural networks

https://doi.org/10.1515/oncologie-2023-0115
Received March 16, 2023; accepted June 12, 2023; published online June 28, 2023

Abstract

Objectives: Bone is the most common site of metastasis in prostate cancer (PCa) patients and is correlated with poor prognosis and increasing economic burden. Few studies have analyzed the prognostic prediction for metastatic PCa patients with the assistance of neural networks.

Methods: Four convolutional neural network (CNN) models are developed and evaluated to predict the overall survival (OS) of PCa patients with bone metastasis. All the CNN models are first trained with 64 samples and evaluated with 10 samples; two models use only bone scan images and two models use both bone scan images and clinical parameters (CPs). The predictions of the best models are compared with those by two urology surgeons on 20 test samples.

Results: Our best models can predict OS of PCa patients with bone metastasis with AUC=0.8022 by using only bone scan images and AUC=0.8132 by using both bone scan images and CPs on 20 test samples. The best Youden indexes of the two models are 0.6263 and 0.7142, respectively, which are 0.3077 and 0.3131 higher than that of the urologists’ average Youden index, which indicate that CNN models exhibit significant advantages.

Conclusions: CNN models are suitable to predict OS in PCa patients with bone metastasis using bone scan images and CPs. Our models show better performance in terms of accuracy and stability than urology surgeons.

Keywords: bone metastasis; bone scan; convolutional neural network; overall survival; prostate cancer

Introduction

Prostate cancer (PCa) is the second largest cause of malignant tumors in men worldwide, accounting for over 370,000 death annually [1]. PCa has a strong proclivity to metastasize to bone, which may be partly explained by the “dependence of the seed on a fertile soil” theory [2]. According to a previous study, based on autopsies, bone is the most common metastatic site of PCa and is involved in approximately 90 % of patients with metastatic PCa [3]. The progression of bone metastasis is a comprehensive multi-step process that includes the colonization of circulating cancer cells and the reconstruction of bone structure [4, 5]. Bone metastasis has been confirmed to be an adverse prognostic factor of PCa that imposes clinical and economic burdens on patients, especially those with skeletal-related events such as pathologic fracture, spinal cord compression, or palliative treatment [6–9].

For PCa patients with bone metastasis, bone scanning is the most prevalent and cost-effective method for diagnosis and follow-up [10]. It is performed using an intravenous injection of technetium-99m-labeled diphosphonate, a compound that can rapidly accumulate in the bones. Imaging is typically performed 2–6 h after injection, which allows clearance of the radiotracer from the soft tissues and
improves bone visualization [11]. Anterior and posterior whole-body images of the skeleton are obtained and potential sites of bone lesions are identified as hotspots. Bone metastases of PCa often show increased blood flow levels and a high rate of new bone formation, leading to radiotracer accumulation. One major disadvantage of bone scans is non-tumor-specific radiotracer uptake, resulting in limited specificity; this is inferior to the specificity of other molecular imaging methods such as positron emission tomography/computed tomography (PET/CT) [12, 13]. However, owing to its sensitivity, availability, and affordability, bone scanning remains the mainstay method for investigating bone metastasis in PCa patients in most institutions [14].

In recent years, deep neural networks, particularly convolutional neural network (CNN) models, have achieved excellent success in computer vision applications, including image classification, image semantic segmentation, and image object detection. Numerous CNN models have been developed to improve performance on natural image classification [15–21]. Based on the success of natural image recognition, many CNN models have been developed and proven effective for medical image processing [22–29].

In this study, we built CNN models to predict the overall survival (OS) of PCa based on data from bone scan images and clinical parameters (CPs). First, several CNN models have been developed based either on bone scan images or using both bone scan images and CPs. Second, the four best-trained CNN models were selected using a validation set. Finally, the performance of the selected models was compared with that of two urology surgeons using a test set.

**Materials and methods**

**Patient selection**

We retrospectively reviewed the data of patients with PCa and bone metastasis at Beijing Chaoyang Hospital from September 2008 to August 2021. All patients were diagnosed with PCa using ultrasound-guided transrectal prostate biopsy. The highest Gleason score among all biopsy cores was documented as the biopsy Gleason score. Bone metastatic lesions were evaluated using a bone scan prior to treatment. The images were acquired using a conventional SPECT/CT camera (Infinia Hawkeye 4, GE, USA). Patients who received any type of PCa treatment before the bone scan were excluded.

All patients were initially treated with a combined androgen blockade, which was administered with an androgen receptor blocker (bicalutamide) and a GnRH agonist (leuproline or goserelin), or surgical castration followed by abiraterone glucocorticoid and/or docetaxel chemotherapy for subsequent castration-resistant prostate cancer. Cytoreductive prostatectomy was performed by urological surgeons with at least 10 years of experience in laparoscopic surgery. Bone-modifying agents, pain relief, and palliative radiotherapy for terminally ill patients were used as appropriate and were not regulated.

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and approved by the Institutional Review Board of Beijing Chaoyang Hospital, Capital Medical University (No. 2022-Re-55), which waived the requirement for informed consent for this retrospective analysis.

**Data preparation**

Bone scan images and CPs at diagnosis were retrieved by reviewing electronic medical records. The endpoint of this study was all-cause mortality, and the OS time was calculated from the diagnosis of metastatic PCa to the date of death. Ultimately, 94 samples were collected from the dataset. 64 samples from the dataset were randomly selected as the training set and 10 of the remaining samples were used as the validation set to select the best models. To verify the accuracy of our model in an extensive setting and compare it with that of various urology surgeons as human readers, the remaining 20 samples were used as the test set.

A typical series of bone scan images contained four sub-images for each patient, including two grayscale images and two color (RGB) images. The grayscale images included one front scan image (GFSI) and one back scan image (GBSI); the color images also included a front scan image (CFSI) and a back scan image (CBSI). A typical example of a bone scan is shown in Figure 1. The extra CPs corresponding to the bone scan images of each patient were collected at the first diagnosis, as shown in Table 1.

The OS of each PCa patient was obtained during outpatient follow-up. Samples with an OS of less than 5 years were labeled as positive, and the others labeled as negative. The distribution of OS within the datasets is shown in Table 2. The training set comprised 42 positive and 22 negative samples. The validation set contained 7 positive and 3 negative samples. The test set contained 13 positive and 7 negative samples.

**Development of CNN models for PCa OS**

Two CNN models were developed to predict the OS of patients with PCa based solely on bone scan images. Two other CNN models were developed using both bone scan images and CPs. The models were trained to minimize the cross-entropy loss function between the labels and the model predictions. A momentum optimizer was used to train the models. The batch size was 64. The original learning rate was 0.01, which was decayed by 0.9 per epoch.

**Image preprocessing for CNN models:** The bone scan images contained many unusable areas. To obtain the useful area, the central $256 \times 850$ pixels of each sub-image were cropped and concatenated. Two concatenation methods were developed: Figure 2(a) shows one method that concatenates all four sub-images, and Figure 2(b) shows another method that concatenates GFSI with CFSI and concatenates GBSI with CBSI. In the next step, all concatenating images were resized to $256 \times 640$ pixels and fed into the basis CNN models.

**The basis CNN model:** The residual module used in ResNet is based on the basis CNN model. Figure 2(c) shows the structure of the basis CNN model, which contains nine convolutional layers and one global average
Table 1: Baseline characteristics of all patients.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients, n</td>
<td>94</td>
</tr>
<tr>
<td>Age at diagnosis (year), median (IRQ)</td>
<td>72 (69−76)</td>
</tr>
<tr>
<td>Clinical tumor stage, n, %</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>44 (46.8)</td>
</tr>
<tr>
<td>T3a</td>
<td>9 (9.6)</td>
</tr>
<tr>
<td>T3b</td>
<td>20 (21.3)</td>
</tr>
<tr>
<td>T4</td>
<td>21 (22.3)</td>
</tr>
<tr>
<td>Visceral metastasis, n, %</td>
<td>4 (4.3)</td>
</tr>
<tr>
<td>Biopsy gleason score, n, %</td>
<td></td>
</tr>
<tr>
<td>3 + 3</td>
<td>6 (6.4)</td>
</tr>
<tr>
<td>3 + 4</td>
<td>6 (6.4)</td>
</tr>
<tr>
<td>4 + 3</td>
<td>10 (10.6)</td>
</tr>
<tr>
<td>8</td>
<td>15 (16.0)</td>
</tr>
<tr>
<td>9 or 10</td>
<td>30 (31.9)</td>
</tr>
<tr>
<td>Unknown</td>
<td>27 (28.7)</td>
</tr>
<tr>
<td>PSA level at diagnosis, ng/mL, median (IRQ)</td>
<td>67.35 (17.40−390.84)</td>
</tr>
<tr>
<td>PSA level when bone scan was performed, ng/mL, median (IRQ)</td>
<td>40.70 (17.40−390.84)</td>
</tr>
<tr>
<td>Cytoreductive surgery, n, %</td>
<td>33 (35.1)</td>
</tr>
<tr>
<td>Hemoglobin, g/L, median (IRQ)</td>
<td>129 (116−140)</td>
</tr>
<tr>
<td>White blood cell count (10^3), median (IRQ)</td>
<td>5.98 (4.79−7.60)</td>
</tr>
<tr>
<td>Neutrophil ratio, % median (IRQ)</td>
<td>65.4 (58.2−72.0)</td>
</tr>
<tr>
<td>Neutrophil count (10^9), median (IRQ)</td>
<td>3.97 (2.72−5.41)</td>
</tr>
<tr>
<td>Lymphocyte ratio, %, median (IRQ)</td>
<td>24.7 (19.2−31.2)</td>
</tr>
<tr>
<td>Lymphocyte count (10^9), median (IRQ)</td>
<td>1.55 (1.16−1.87)</td>
</tr>
<tr>
<td>Monocyte ratio, %, median (IRQ)</td>
<td>6.7 (5.5−8.2)</td>
</tr>
<tr>
<td>Monocyte count (10^9), median (IRQ)</td>
<td>0.39 (0.32−0.50)</td>
</tr>
<tr>
<td>Platelet count (10^9), median (IRQ)</td>
<td>202 (146−245)</td>
</tr>
<tr>
<td>Albumin, g/L, median (IRQ)</td>
<td>40.7 (35.8−43.7)</td>
</tr>
<tr>
<td>Serum calcium, mmol/L, median (IRQ)</td>
<td>2.26 (2.15−2.37)</td>
</tr>
<tr>
<td>Total cholesterol level, mmol/L, median (IRQ)</td>
<td>4.54 (3.80−5.11)</td>
</tr>
<tr>
<td>Alkaline phosphatase, u/L, median (IRQ)</td>
<td>90 (75−157)</td>
</tr>
</tbody>
</table>

IQR, interquartile range; PSA, prostate-specific antigen.

Table 2: The distributions of the data sets.

<table>
<thead>
<tr>
<th>Data sets</th>
<th>&lt;5 years (positive)</th>
<th>&gt;5 years (negative)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>42</td>
<td>22</td>
<td>64</td>
</tr>
<tr>
<td>Validation set</td>
<td>7</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Test set</td>
<td>13</td>
<td>7</td>
<td>20</td>
</tr>
</tbody>
</table>

Development models using both bone scan images and CPs: To improve the performance of the models further, two other models were developed using both bone scan images and CPs. Figure 3(a) shows one model (model C) concatenating the CPs and the features extracted from one CNN model and feeding all the features into an FC layer + Softmax to obtain the prediction results. Figure 3(b) shows another model (model D) extracting features from two CNN models: one extracting the features of the concatenating GFSI and CFSI sub-images, and the other extracting the features of the concatenating GBSI and CBSI sub-images. The CPs and all CNN extracting features were concatenated and fed into the next step.

Performance comparison to urology surgeons

Two urologists with at least 5 years of experience at Beijing Chaoyang Hospital, Capital Medical University, were recruited for prognosis prediction. The prediction of urology surgeons proceeded in two phases. First, urologists were provided with all originally acquired bone scan images for each patient, including the grayscale front, grayscale back, color front, and color back scan images. In the second phase, urologists were provided with both bone scan images and the CPs for each patient. Patients were presented to urologists in a random order. The urologists were aware that all patients had metastatic PCa but were otherwise blinded to the follow-up results. The urologists were instructed to
predict the 5-year OS with a yes or no answer for each patient based on their expertise and clinical experience. To evaluate the stability of urology surgeons, they were instructed to repeat the prediction procedure on another day using 30 samples to test intra-observer reliability.

We compared the prediction performance of the CNN models and urology surgeons, which was first evaluated by reading only bone scan images and then by reading both bone scan images and CPs. The diagnostic results were compared using the metrics described below.
Analysis of diagnostic performance using metrics methods

Sensitivity and specificity were used to evaluate the performances of urology surgeons and our models. Sensitivity represents the true-positive rate, whereas specificity represents the true-negative rate. In this study, we considered samples with an OS of <5 years as positive samples and other samples as negative samples.

Our models only predicted the positive probability of each sample, therefore the performance was best illustrated by the AUC and the receiver operating characteristic (ROC) curve. An ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied [30, 31]. The ROC curve in this scenario was created by plotting the sensitivity against (1 – specificity) at various threshold settings. The AUC is the area under the ROC curve. The Youden index was used to facilitate a comparison between the performances of the CNNs and of the urologists.

Results

Performance of the models using only bone scan images

The validation set was used to evaluate the models’ performances. Two models with different structures—Model A, shown in Figure 2(a), and Model B, shown in Figure 2(b)—were established. Models with three different network widths were evaluated to select the best ones. Model A exhibited the performance of AUC=0.6190, 0.7619, and 0.7143 for C=8, 16, and 32, respectively. Figure 4(a) shows the best performance of model A, with an AUC of 0.7619, when C=16. Model B exhibited the performance of AUC=0.7143, 0.8571, and 0.8095 for C=4, 8, and 16, respectively. Figure 4(a) shows the best performance of Model B, with AUC=0.8571 when C=8. The best performance of Model B was superior to that of Model A, with the AUC of the former being 9.52 % higher than that of the latter.

Performance of the models using both bone scan images and CPs

To further improve the performance of the models, two models using both bone scan images and CPs, namely Model C, shown in Figure 3(a), and Model D, shown in Figure 3(b), were established and evaluated using the validation set. Figure 4(b) shows the best performance of Model C, with an AUC of 0.7619 when C=16. Figure 4(b) also shows the best performance of Model D, with an AUC of 0.9048 when C=8. When considering both the bone scan images and CPs, Model D performed better than model C, with the AUC of the former being 14.29 % higher, indicating that the model structure used in Models B and D was more suitable for processing bone scan images.

Performance comparison between CNN models and urology surgeons

To comprehensively compare our models with the predictions of the urologists, all four models (A, B, C, and D) and the predictions of the urologists were evaluated. For the same samples, our models obtained the same prediction results at different times; however, the urologists obtained different results. To evaluate the stability of the urologists, the intra-observer agreement between the two urologists was tested. The results are shown below.

Using only bone scan images

Models A and B were compared with the two urologists for predicting PCa OS using bone scan images. To compare our two best models with the predictions of the urologists, the two urology surgeons from Beijing Chaoyang Hospital and the two best models independently scored 20 bone scan images (test set). The best performance was obtained by
Model A (AUC=0.7692) and Model B (AUC=0.8022), with both outperforming the two urology surgeons, as shown in Figure 5(a). When the specificity was set to be equal to the urologists’ average level, the sensitivities of Models A and B were 15.38 and 30.77 % higher, respectively, than the urologists’ average level. The best Youden indices of Models A and B were 0.5604 and 0.6263, respectively, which were 0.2418 and 0.3077 higher than the urologists’ average level. Our models achieved better performance in PCa OS when using only bone scan images.

Using both bone scan images and CPs

Our models were further compared with the predictions of the urologists for PCa OS using both bone scan images and CPs. Figure 5(b) shows the performance of our models and the predictions of the urologists. Model C demonstrated an AUC of 0.7912 and model D exhibited an AUC of 0.8132; thus, Model D showed better performance. Both models showed better performance than the urologists. When the sensitivities were set to be equal to the urologists’ average level, the specificities of both Models C and D were 7.2 % higher than the urologists’ average level. When the specificities were set to the urologists’ average levels, the sensitivities of both Models C and D were 7.7 % higher than the urologists’ average levels. The best Youden indices of Models C and D were 0.6373 and 0.7142, respectively, which were 0.2362 and 0.3131 higher than the urologists’ average level. Overall, our models achieved a better performance for PCa OS when considering both bone scan images and CPs.
The stability of urology surgeons

Each of the two selected urologists predicted 30 samples at one month interval to evaluate the consistency of the human readers. Table 3 shows their performance. When only bone scan images were used, the two urologists obtained consistency rates of 86.7 % and 76.7 %, respectively, with an average of 81.7 %. When using both bone scan images and CPs, the two urologists obtained consistency rates of 83.3 % and 86.7 %, respectively, with an average of 85.0 %. All models had a 100 % consistency rate regardless of the interval; therefore, our model showed better stability than the predictions by the urologists.

Discussion

In recent years, CNN methods have demonstrated many successful applications in medical image processing, proving that they are suitable for this domain. Several previous studies have shown that CNN methods are useful for assisting in the diagnosis and radiotherapy treatment planning for PCa patients [32–34]. However, these studies have been limited to predicting the prognosis of metastatic PCa using neural networks. In addition, models based on a combination of bone scan images and CPs have not yet been reported. In this study, we built prediction models for patients with PCa and bone metastasis using CNN methods based on data from a Chinese population. The bone scan images were typical medical images. Therefore, a CNN was employed in this study to process the bone scan images to predict the OS of PCa patients with bone metastasis.

Prognostic prediction for PCa patients with bone metastasis is essential for determining the appropriate treatment. The increased extent of metastatic lesions has been confirmed to be associated with an increasingly poor prognosis [6, 35]. Several researchers have attempted to quantify the degree of bone metastasis using bone scan images and have obtained different results. The extent of disease (EOD) score is a semi-quantitative parameter useful for the prognostic prediction of PCa [36]. However, this method is subjective, and the results may vary significantly depending on the observer. An index called the “bone scan index” (BSI) was developed to provide an objective and quantitative measurement of the percentage of the skeleton consumed by bone metastases [37]. Although the BSI provides a relatively objective estimation of the metastatic burden of PCa, especially with the assistance of a computer-aided diagnostic system, it is an indirect indicator of disease aggressiveness [38, 39]. Further analysis is required to establish the thresholds for risk stratification of prognosis [40]. In addition, the calculation of BSI does not consider the different effects among metastatic sites, as metastases in the appendicular skeleton, such as the ribs and limbs, have been reported to be significantly correlated with a poorer response to androgen deprivation therapy and a shorter survival time [35]. Additionally, the utility of BSI is limited by its reliance on automated measurement platforms. However, in this study, the bone scan images were directly processed using CNN methods, which simultaneously analyzed both the number and position of metastatic sites. Furthermore, our models were able to generate a direct prediction of OS for PCa patients with bone metastasis, were simple to use, and omitted secondary analyses for prognosis.

Another advantage of our models is their superior performance compared to that of the selected urologists. In our study, we noted that the tumor stage, biopsy Gleason score, and the degree of metastatic burden shown in the bone scan images were the major concerns of clinicians when making their predictions. Generally, patients with higher tumor stages, higher Gleason scores, and more metastatic lesions are considered to have a shorter OS. Similarly, prediction models reported in previous studies were based on these clinical parameters [41, 42]. However, urologists are not sufficiently competent to process large volumes of data. As a result, their predictions hardly benefit from the addition of more CP types. In this study, our models could directly analyze the entire database using CNN methods and provided significantly more accurate predictions than the selected urologists. The results showed that our CNN models performed better than the urologists in predicting OS in patients with PCa with bone metastasis using bone scan images. Therefore, we believe that our model may effectively improve the management of PCa patients with bone metastasis.

Four CNN models were trained and evaluated in this study. Two models (Models A and B) used only bone scan images, and two models (Models C and D) used both bone scan images and CPs. The results showed that the latter achieved better performance than the former. The CPs contain different types of information regarding the features extracted from the images. When considering both the bone

Table 3: The inter-observer results of two urologists.
scan images and the CPs, more information was fed into the last classifier, which resulted in a better performance. Therefore, Model C showed better performance than Model A, and Model D showed better performance than Model B. Two models (Models B and D) independently extracted the features of the front and back sub-images, while two models (Models A and C) fused and extracted the features of the four sub-images. The front and back sub-images of the bone scan provide different clinical information regarding bone invasion. It is better to extract the features of each sub-image and merge them in the last classifier than to merge them in each convolutional layer. Therefore, Model B performed better than Model A and Model D performed better than Model C.

Our CNN models showed better performance in terms of accuracy and stability than the predictions of the urologists. However, this study has some limitations. First, the sample size used in our study was relatively small, and more samples would improve the performance of the models. Second, because our study was based solely on single-center medical samples, it was difficult to verify the generalizability of our CNN models. Therefore, additional multicenter medical samples may help improve and verify the generalizability of the models. Third, the performances of only two urologists were evaluated and compared in our study. The evaluations of prediction by a greater number of urological surgeons with different levels of clinical experience is warranted.

Conclusions

In conclusion, our study is the first to prove that CNN models can be used to predict OS in patients with PCa presenting bone metastasis based on data from bone scan images and CPs. Our models showed better performance in terms of accuracy and stability than the predictions of urological surgeons. Our study provides a useful reference for adopting deep neural networks for prognostic prediction in patients with metastatic PCa.

Research funding: This work was supported by National Natural Science Foundation of China (grant numbers: 82170783), Funded by Science and Technology Project of Hebei Education Department (BJK2023060), Hebei Province Introduced Returned Overseas Chinese Scholars Funding Project (C20220107) and Hebei Natural Science Foundation (C2022406010).

Author contributions: Concept and design: Z.W., and T.X.; Acquisition, analysis of data: Y.S., and Z.W.; Drafting of the manuscript: Z.W., T.X., X.D., and Y.N.; Formal analysis: Y.C. and X.Q.; Administrative, technical, or material support: X.Z., and S.X.; Obtained funding: X.D., and Y.N.; Supervision: D.W., X.D., M.Y., and Y.N.

Competing interests: All the authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential competing of interest.

Informed consent: The requirement of informed consent for this retrospective analysis was waived with the approval of the Institutional Review Board of Beijing Chaoyang Hospital, Capital Medical University.

Ethical approval: Our study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Institutional Review Board of Beijing Chaoyang Hospital, Capital Medical University (NO.: 2022-Ke-55).

References


Supplementary Material: This article contains supplementary material (https://doi.org/10.1515/oncologie-2023-0115).