Annals of Nuclear Medicine (2020) 34:824–832
https://doi.org/10.1007/s12149-020-01510-6

ORIGINAL ARTICLE

Development of Convolutional Neural Networks to identify bone metastasis for prostate cancer patients in bone scintigraphy

Nikolaos Papandrianos[1] · **Elpiniki I. Papageorgiou**[1,3] · **Athanasios Anagnostis**[2,3]

Received: 27 April 2020 / Accepted: 4 August 2020 / Published online: 24 August 2020

Abstract

Objective The main aim of this work is to build a robust Convolutional Neural Network (CNN) algorithm that efficiently and quickly classifies bone scintigraphy images, by determining the presence or absence of prostate cancer metastasis.

Methods CNN, widely applied in medical image classification, was used for bone scintigraphy image classification. The retrospective study included 778 sequential male patients who underwent whole-body bone scans. A nuclear medicine physician classified all the cases into 3 categories: (1) normal, (2) malignant, and (3) degenerative, which were used as the gold standard.

Results An efficient CNN architecture was built, based on CNN exploration performance, achieving high prediction accuracy. The results showed that the method is sufficiently precise when it comes to differentiating a bone metastasis from other either degenerative changes or normal tissue (overall classification accuracy = 91.42% ± 1.64%). To strengthen the outcomes of this study the authors further compared the best performing CNN method to other popular CNN architectures for medical imaging, like ResNet50, VGG16 and GoogleNet, as reported in the literature.

Conclusions The prediction results reveal the efficacy of the proposed CNN-based approach and its ability for an easier and more precise interpretation of whole-body images in bone metastasis diagnosis for prostate cancer patients in nuclear medicine. This leads to marked effects on the diagnostic accuracy and decision-making regarding the treatment to be applied.

Keywords Bone metastasis · Prostate cancer · Nuclear imaging · Bone scintigraphy · Deep learning · Image classification · Convolutional Neural Networks

Introduction

Bone metastasis is one of the most frequent cancer complications and emerges mainly in patients with certain types of primary tumors, especially of the breast, prostate and lung [1]. These types of cancer have great avidity for bone, causing painful and untreatable symptoms; thus, an early diagnosis is a crucial factor for making treatment decisions as well as possibly having a significant impact on the progress of the disease and patient quality of life. Lymph nodes and bones are the sites where metastatic prostate cancer mainly appears. By implementing a carefully selected diagnostic imaging, the number of metastatic foci in the skeletal system can be determined [2].

Different modern imaging techniques have been used in clinical practice, for achieving rapid diagnosis of bone metastases such as scintigraphy, positron emission tomography (PET) and whole-body MRI [3, 4]. Bone scintigraphy (BS) is the most common imaging method in nuclear medicine and is regarded as the gold standard [3, 5]. BS that uses single-photon emission computed tomography (SPECT) imaging, is particularly important for the clinical diagnosis of metastatic cancer, both in men and women [1], [6].

✉ Nikolaos Papandrianos
npapandrianos@uth.gr

Elpiniki I. Papageorgiou
elpinikipapageorgiou@uth.gr; e.papageorgiou@certh.gr

Athanasios Anagnostis
athananagno@uth.gr; a.anagnostis@certh.gr

[1] Department of Energy Systems, Faculty of Technology, University of Thessaly, Geopolis Campus, Larissa - Trikala Ring Road, 41500 Larissa, Greece

[2] Computer Science & Telecommunications Department, University of Thessaly, 35131 Lamia, Greece

[3] Institute for Bio-economy and Agri-technology, Center for Research and Technology Hellas, Thessaloníki, Greece

In the medical image analysis domain, machine learning and especially deep learning algorithms have been investigated and implemented, due to their marked applicability in diagnostic medical images interpretation [7, 8]. The contribution of deep learning in medical imaging is attained through the use of Convolutional Neural Networks (CNNs) [8, 9]. The learning process is an advantageous feature of CNNs as they can learn useful representations of images and other structured data. CNNs are proven to be powerful deep learning models in medical image analysis [7–9]. In this field, the most popular CNNs methods are AlexNet (2012) [10], ZFNet (2013) [11], VGGNet16 (2014) [12, 13], GoogleNet (2014) [14], ResNet (2015) [15] and DenseNet (2017) [16]. The advantageous features of CNNs have a significant impact on their performance, making them so far, the most efficient methods for medical image analysis and classification.

CNNs have been applied in nuclear medicine mainly for classification and segmentation tasks showing some promising preliminary results [17]. Two research works (master theses) regarding the CNN application in bone metastasis diagnosis, were devoted to image analysis and classification of bone scans using the Exini software [25, 26]. The dataset in both works was delivered by Exini Diagnostics AB and contained image patches of hotpots that had been already detected. With regard to the produced results, the testing accuracy for the first study was 89% [18], whereas the evaluation indicator of the area under receiver operating characteristic (ROC) score was 0.9739 for the second case, being significantly higher than the respective ROC of 0.9352, obtained by methods reported in the literature for the same test set [26].

The remaining research works that concern the same imaging modality (BS), are devoted to the introduction of computer-aided diagnosis (CAD) systems with the use of ANN and other machine learning (ML) methods for detection of bone metastasis in bone scintigraphy images. The first CAD system for BS was introduced in 1997 [19]. It was semi-automated and allowed quantification of bone metastases from whole-body scintigraphy through an image segmentation method. Fully automated CAD systems were later examined by Sajn et al. [20], who introduced a machine learning-based expert system. Recently, a new parallelepiped classification method was proposed to automatically detect bone metastasis in BS images [21]. From the produced results, it emerges that this method is quite accurate in differentiating metastatic bone from normal tissue (overall classification accuracy = 87.58 and κ coefficient = 0.8367).

Focusing on PET/CT imaging, a deep CNN for automatically differentiating benign and malignant lesions in NaF PET/CT images of patients with metastatic prostate cancer has been implemented [22]. The proposed method followed the VGG19 architecture, by employing 16 3×3 convolutional layers followed by 2 fully connected layers and a final softmax layer. Next, a CNN-based system was examined to detect malignant findings in FDG PET–CT examinations in a retrospective study [23]. The study included 3485 sequential patients who underwent whole-body FDG PET–CT. All the respective cases were classified into 3 categories: (1) benign, (2) malignant and (3) equivocal. The same research team has lately developed, a CNN-based system of similar configuration to ResNet24, that predicts the location of the malignant uptake and also evaluates the accuracy of predictions [24], classifying whole-body FDG PET images into the above-mentioned three categories.

With respect to this research study, we focus only on the classification task of metastatic prostate cancer in bones and not on whole-body scans segmentation in skeletal scintigraphy images. Reviewing the relevant literature, a couple of research works ([18] and [25]) devoted to BS have been carried out, which presented some preliminary results only, while the advantageous and outstanding capabilities of CNNs have not been yet fully investigated. The main challenge in BS is to build an extremely accurate algorithm that automatically identifies whether a patient has bone metastasis or not, just by looking at whole-body scans.

This paper presents innovative ideas and contributes mainly to the following issues:

- The development of a robust CNN algorithm which can automatically identify metastatic patients of prostate cancer by examining whole-body scans.
- The efficacy of RGB mode that enhances the classification accuracy compared to other state-of-the-art CNN models, like ResNet50, VGG16, Inception V3, and Xception.

After a thorough CNN investigation regarding the architecture and configuration of hyperparameters, the authors proposed a robust CNN which can identify the right category of bone metastasis.

This paper is structured as follows: "Material and methods" section presents the materials and methods that are utilized in this research study. This is followed by the section that reports on the proposed CNN architecture for bone metastasis classification, whereas "Results" section contains an exploration analysis, examining different parameters and configurations for the best CNN model, and a comparative analysis with other state-of-the-art CNNs, along with the produced results. "Discussion of results and conclusion" section provides a rigorous discussion on analysis of results, the overall conclusions of this study, and future steps.

Materials and methods

Prostate cancer patient images

This study was approved by the Director of the Diagnostic Medical Center "Diagnostico-Iatriki A.E." who also gave his consent due to the retrospective nature of this study. All procedures in this study were in accordance with the Declaration of Helsinki. The retrospective review that was performed, comprised 778 whole-body scintigraphy images from 817 different male patients who visited the Nuclear Medicine Department of the Diagnostic Medical Center "Diagnostico-Iatriki A.E.", Larissa, Greece, between January 2013 and June 2018. The selected images came from prostate cancer patients with suspected bone metastatic disease, who had undergone whole-body scintigraphy.

Out of 778 bone scan images, 328 bone scans concern patients with bone metastasis (see Fig. 1), 271 benign cases that include degenerative changes and 179 normal patients without bone metastasis. A nuclear medicine physician, with 15 years' experience in bone scan interpretation, classified these cases into 3 categories: (1) healthy (metastasis absent), (2) degenerative (no metastasis, but image includes degenerative lesions/changes) and (3) malignant (metastasis present), which were used as the gold standard. The metastatic images were confirmed after further examinations performed by CT/MRI.

A Siemens gamma camera Symbia S series SPECT System (by a dedicated workstation and software Syngo VE32B) with two heads of low-energy high-resolution (LEHR) collimators was used for patient scanning. Two types of radionuclide were used for BS, 99m-Tc-HDP (TechneScan®) and 99-Tc-MDP (PoltechMDP 5 mg). Whole-body scintigraphy was acquired approximately 3 h after an intravenous injection of 600–740 MBq of a radiopharmaceutical agent, depending on the patient's body type.

In total, 778 planar bone scan images from patients with known P-Ca were retrospectively reviewed. A whole-body field was used to digitally record anterior and posterior views with a resolution of 1024 × 256 pixels. Images represent counts of detected gamma decays in each spatial unit with 16-bit grayscale depth.

Main aspects of Convolutional Neural Networks

A Convolutional Neural Network is a type of ANN that uses connections between layers to maintain the spatial relationships within the data. It is structured with layers that perform convolutions, followed by layers that perform pooling, several times over. In particular, the main elements/components of a CNN architecture are: (i) convolutional layers, (ii) activation layer, (iii) pooling, (iv) dropout regularization and (v) batch normalization. Gradient descent is used for optimization and backpropagation for updating the weights after an iteration [26, 27]. The last part of a CNN is one or more fully connected layers, which "flatten" the output of the previous layer. By definition, this means that it is transformed into a vector that can be used as the output layer. Based on the outcome vector values, the algorithm gives a label to each image [26]. In most CNN models, the ReLU function is the activation function that is used in all convolutional and fully connected (dense) layers, whereas the sigmoid function is the final activation function usually used in output nodes [26].

Methodology

In this study, an approach based on CNN model, as a common method in medical image analysis, is suggested for a three-class bone metastasis classification. Figure 2 illustrates the preprocessing, training and testing phase for the

Fig. 1 Image samples of metastatic prostate cancer (from our dataset)

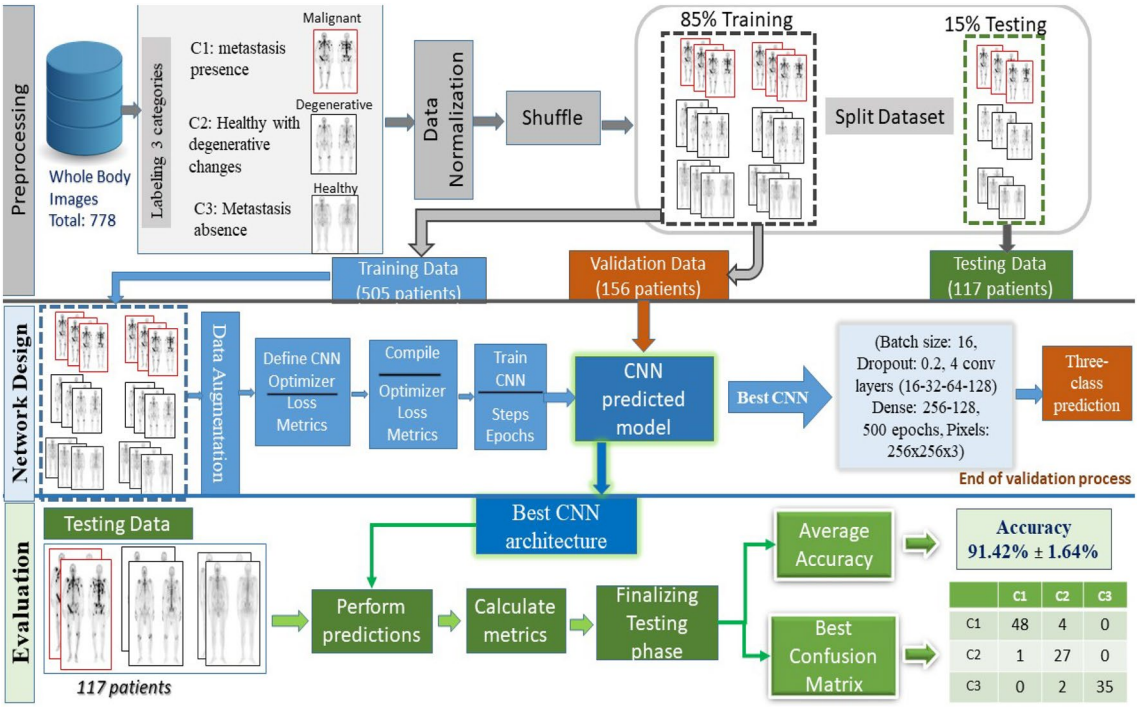

Fig. 2 The three stages of the proposed methodology for bone metastasis classification

examined dataset of 778 patients, classifying them into three categories: (i) malignant (bone metastasis present), (ii) healthy (no bone metastasis) and (iii) degenerative (healthy, presenting degenerative changes).

Preprocessing stage

In this stage, images given by a nuclear medicine doctor are first loaded in RGB mode and stored in a PC memory. Data normalization follows, which is a useful and necessary method in a machine-learning process that rescales the data values within 0 and 1. In order to avoid choosing wrong samples for training and testing, the shuffling method comes to give a random order to data. In the case of a small number of data, data augmentation takes over and applies the techniques of rescale, rotation range, zoom range, and flip in order to create variant images without generating new ones and avoid overfitting. Data augmentation is used only in the training phase that follows next. Finally, the original dataset is split into three segmented parts: testing (15%) and the rest (85%) is split into 80% for training and 20% for validation.

Network design stage

In our effort to determine which is the most proper architecture for the CNN, considering image classification, we first need to perform an exploration process, defining the most significant parameters. The experimentation and trial phase can help the researchers to explore and then choose the most effective network architecture with respect to the number and type of convolutional layers, number of nodes and pooling layers, filters, drop rate and batch size, and number of dense nodes. Next, activation and loss functions are defined based on bibliographic research.

Evaluation/testing stage

The testing data that were initially split and completely unknown to the model are used during the testing process. Predictions on each image class are then carried out by the classifier and finally, a comparison between the calculated predicted class and the true class is performed. In the next step, common performance metrics, such as testing accuracy, precision, recall, F1-score, sensitivity and specificity of the model are computed to evaluate the classifier [8, 28, 29]. An error/confusion matrix is also employed to further evaluate the performance of the model and check if it is biased over one or the other class.

Proposed CNN architecture for bone metastasis classification

In this research study, a CNN architecture is proposed to precisely identify bone metastasis from whole-body scans of men with prostate cancer. In this context, a deep-layer

network with 4 convolutional—pooling layers, 2 dense layers followed by a dropout layer, as well as a final output layer with three nodes, is built (see Fig. 3).

Following the structure of the CNN, the first (input) convolutional layer consists of 3 × 3 filters (kernels) of size 3 × 3, followed by a max-pooling layer of size 2 × 2 and a dropout layer with 0.2 as dropout rate. Next, a flattening operation transforms the 2-dimensional matrices to 1-dimensional arrays, in order to run through the hidden fully connected (dense) layer with 64 nodes. To avoid overfitting, a dropout layer was suggested to drop randomly 20% of the learned weights. The final layer is a three-node layer (output layer).

In the suggested CNN architecture, the ReLU function is used in all convolutional and fully connected (dense) layers, whereas the categorical cross-entropy function is the final activation function used in output nodes. Two performance metrics, concerning accuracy and loss, are used. As regards loss, the categorical cross-entropy function is calculated with an ADAM optimizer, which is an adaptive learning rate optimization algorithm [30].

Results

This section reports the findings of this work, based upon the results produced from the methodology applied. For the classification accuracy calculations, each process was repeated 10 times.

Three-class classification problem using CNN in RGB mode

The experiments were performed in a collaborative environment, called Google Colab [31], which is a free Jupyter notebook environment in the cloud. The software that was used in this study also included OpenCV, which was used for loading and manipulating images, Glob, for reading filenames from a folder, Matplotlib, for plot visualizations and finally, Numpy, for all mathematical and array operations. Moreover, Python was used for coding, Keras (with Tensorflow [32]) for programming the CNN, with data normalization, data splitting, confusion matrices and classification reports carried out with Sci-Kit Learn. The computations ranged between 6 and 8 s per training (epoch) for RGB images (256 × 256 × 3), depending on the input.

It is worth mentioning that the original images, as acquired from the scanning device, are in RGB format, containing 3-channel color information. Although being in RGB structure, these images appear as a grayscale due to the absence of color components in the image. As regards the type of CNN, the authors applied a 2D CNN for each color channel of every image, and then aggregated the inputs into the final part of the network, which comprised a dense layer.

Following the steps described in "Methodology" section, a meticulous CNN exploration process was accomplished, in which experiments with various convolutional layers, drop rates, epochs, number of dense nodes, pixel sizes and batch sizes were conducted. Different values for image pixel sizes were also examined, such as 150 × 150 × 3, 200 × 200 × 3, 256 × 256 × 3, 300 × 300 × 3 and 350 × 350 × 3, and various values for batch sizes, such as 8, 16, 32 and 64 were investigated. In addition, variant drop rate values like 0.2, 0.4, 0.7 and 0.9 were studied, as well as a divergent number of dense nodes, like 16, 32, 64, 128, 256 and 512 was explored. The number of epochs ranged from 200 up to 1200. The number of convolutional and pooling layers was also investigated, whereas early stopping criterion was considered without producing promising classification accuracies.

After a thorough CNN exploration analysis, the best CNN configuration was a CNN with 4 convolutional layers, starting with 16 filters for the first layer, and for each

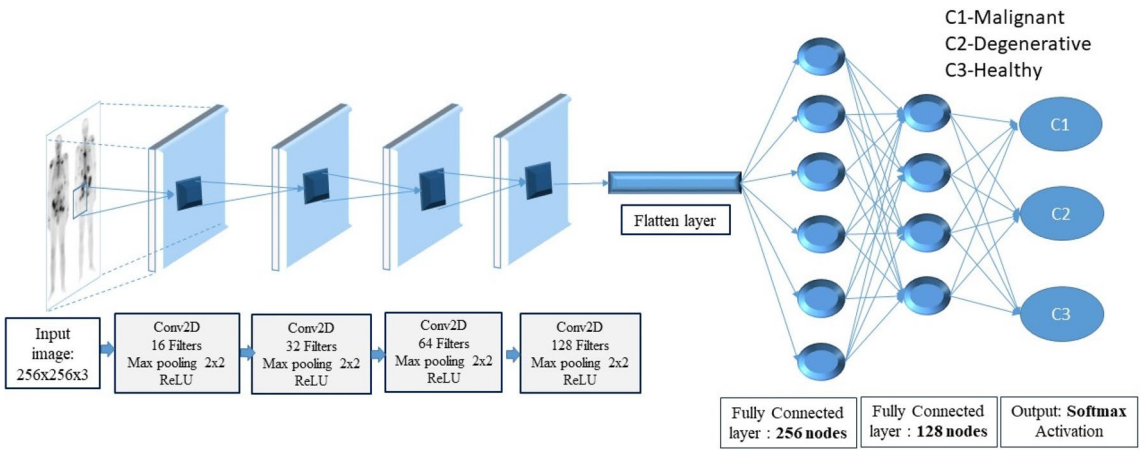

Fig. 3 The CNN framework for bone metastasis classification using BS images

convolutional layer that comes next, the number of filters is doubled (16- > 32- > 64- > 128). This set-up with the best performance parameters is shown in Fig. 3. Table 1 gathers performance analysis and results for the CNN models with the following specifications: 4 convolutional layers (16, 32, 64, 128), epochs = 500, dropout = 0.2, pixel size = 256 × 256 × 3, dense nodes = 256–128 and different batch sizes for 10 runs.

Figure 4a graphically represents the overall classification accuracies for the suggested RGB-CNN, with different dense nodes, concerning the best CNN architecture, whereas Fig. 4b presents the testing accuracy results for different dense nodes and number of convolutional layers. The dense 256–128 provides the highest classification accuracies for dropout = 0.2 and batch size = 16.

Following the exploration analysis, Table 2 depicts the performance analysis and results produced in Google Colab [31], for the suggested RGB-CNN model of 4 convolutional layers (16, 32, 64, 128), epochs = 500, dropout = 0.2, dense nodes = 256–128, for various pixel sizes.

The charts in Fig. 5 represent the (a) accuracy and (b) loss precision curves for the best RGB-CNN model.

Comparison with state-of-the-art CNNs

To further investigate the performance of the proposed CNN architecture, an extensive comparative analysis

Table 1 Classification accuracy and loss values for the examined CNN model (*AV* accuracy validation, *LV* loss validation, *AT* accuracy testing, *LT* loss testing)

	Batch size = 8				Batch size = 16				Batch size = 32			
	AV	LV	AT	LT	AV	LV	AT	LT	AV	LV	AT	LT
Run 1	81,25	0,42	85,72	0,34	91,41	0,23	91,07	0,2	86,72	0,29	92,71	0,18
Run 2	92,97	0,17	86,61	0,53	94,53	0,15	91,96	0,25	93,75	0,27	92,71	0,22
Run 3	83,59	0,41	95,53	0,19	92,19	0,12	93,75	0,15	91,41	0,27	83,30	0,37
Run 4	88,28	0,3	88,39	0,22	91,41	0,22	91,07	0,24	88,28	0,27	92,71	0,21
Run 5	82,03	0,34	89,28	0,33	94,53	0,167	90,17	0,253	90,62	0,25	92,71	0,21
Run 6	87,5	0,36	89,29	0,27	86,71	0,328	89,28	0,393	87,50	0,44	93,75	0,27
Run 7	84,37	0,499	91,07	0,34	90,62	0,22	91,07	0,19	92,97	0,24	92,71	0,19
Run 8	90,62	0,24	91,96	0,35	95,31	0,23	93,75	0,139	89,06	0,41	87,50	0,30
Run 9	92,19	0,27	88,39	0,4	92,96	0,217	92,85	0,16	85,94	0,37	89,58	0,36
Run 10	89,84	0,25	88,39	0,65	85,93	0,45	89,28	0,22	87,50	0,46	82,29	0,39
Average	87,26	0,325	89,46	0,36	**91,56**	**0,23**	**91,42**	**0,22**	89,38	0,33	89,99	0,27

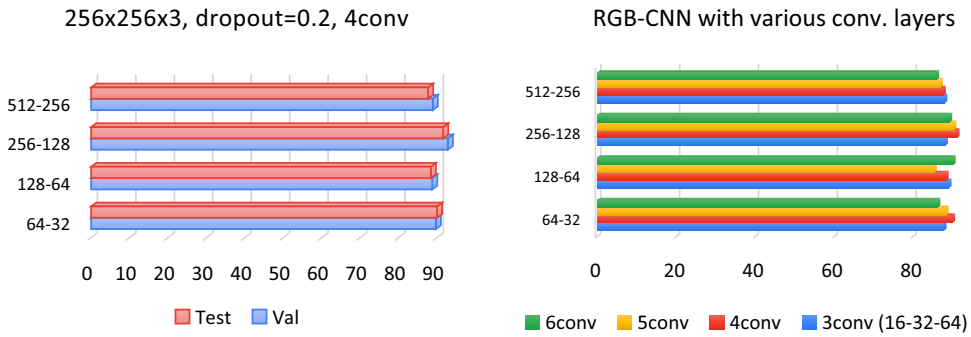

Fig. 4 **a** Validation and testing accuracies for various dense node configurations compared to the suggested RGB-CNN and **b** testing accuracies for various dense node configurations and convolutional layers

Table 2 Various pixel sizes for the best CNN architecture

Pixel sizes	256 × 256 × 3				300 × 300 × 3				350 × 350 × 3			
	AV	LV	AT	LT	AV	LV	AT	LT	AV	LV	AT	LT
Run 1	91,41	0,23	91,07	0,2	95,31	0,15	89,28	0,23	85,16	0,4	91,07	0,18
Run 2	94,53	0,15	91,96	0,25	93,75	0,19	89,28	0,26	84,37	0,41	90,18	0,29
Run 3	92,19	0,12	93,75	0,15	92,97	0,29	85,71	0,36	88,28	0,28	87,5	0,34
Run 4	91,41	0,22	91,07	0,24	91,41	0,18	95,53	0,16	92,97	0,33	84,82	0,36
Run 5	94,53	0,17	90,17	0,25	95,31	0,2	93,75	0,18	92,19	0,2	92,86	0,2
AVE	92,81	0,18	91,60	0,22	93,75	0,20	90,71	0,24	88,59	0,32	89,29	0,27

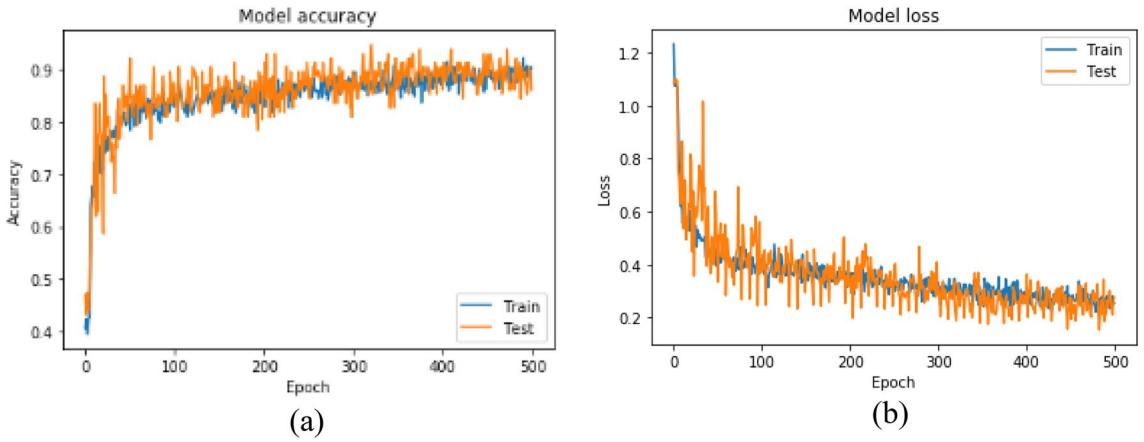

Fig. 5 Precision curves for best CNN in RGB mode: **a** accuracy and **b** loss

between state-of-the-art CNNs, and our best model, was performed. Well-known CNN architectures, such as (i) ResNet50 [15], (ii) VGG16 [12, 13], (iii) Inception V3 [33], and (iv) Xception [34], were used and are described as follows:

(i) ResNet50 is a 50-weight layer deep version of ResNet (Residual neural Network), with 152 layers, based on "network-in-network" micro-architectures [15]. ResNet50 has fewer parameters than the VGG network, demonstrating the fact that extremely deep networks can be trained using standard SGD (and a reasonable initialization function) with residual modules.

(ii)VGG16 [13] is an extended version of Visual Geometry Group (VGG), as it contains 16 weight layers within the architecture. VGGs are usually constructed using 3×3 convolutional layers, which are stacked on top of each other.

(iii) Inception V3 [33], Inception's third installment, includes new factorization ideas. It is a 48-layer deep network which incorporates RMSProp optimizer and computes 1×1, 3×3, and 5×5 convolutions within the same module of the network. Its original architecture is GoogleNet. Szegedy et al. proposed an updated version of the Inception V3 architecture, which was included in Keras inception module [33], capable of further boosting the classification accuracy of ImageNet.

(iv) Xception, which is an extension of the Inception's architecture, replaces the standard Inception's modules with depth-wise separable convolutions [34].

After the extensive exploration of all the provided architectures of popular CNNs, the authors defined the respective parameters, as follows:

- Best RGB CNN: pixel size ($256\times256\times3$), batch size = 16, dropout = 0.2, conv 16–32–64–128, dense nodes 256×128, epochs = 500 (average run time = 3348 s),
- VGG16: pixel size ($250\times250\times3$), batch size = 32, dropout = 0.2, Flatten, dense nodes 512×512, epochs = 200 (average run time = 2786 s),
- ResNet50: pixel size ($250\times250\times3$), batch size = 64, dropout = 0.2, global average pooling, dense nodes 1024×1024, epochs = 200 (average run time = 1853 s),
- Inception V3: pixel size ($250\times250\times3$), batch size = 32, dropout = 0.7, global average pooling 2D, dense nodes 1500×1500, epochs = 200 (average run time = 2634 s),
- Xception: pixel size ($300\times300\times3$), batch size = 16, dropout = 0.2, flatten, dense nodes 512×512, epochs = 200 (average run time = 3083 s).

Figure 6a illustrates the classification accuracy for validation and testing for all the examined CNNs, whereas Fig. 6b depicts the corresponding loss for validation and testing.

It is important to highlight that the authors have conducted an exploratory analysis for all benchmark CNNs for different dropouts, paying special attention to avoid overfitting [26] by reducing the number of weights. Thus, not only did they select the best performing model for each CNN architecture in terms of accuracy, but at the same time, they determined the best CNN that avoids overfitting in 10 runs.

Discussion of results and conclusions

The problem of diagnosis of bone metastasis in prostate cancer patients has been tackled with the use of CNN algorithms, which are considered powerful methods for detecting complex visual patterns in the field of medical image analysis. In the current work, a total of 778 images were acquired, including almost the same numbers of healthy, degenerative and malignant cases from metastatic prostate cancer patients. Several CNN architectures were tested, leading to the one that performed optimally under all hyperparameter

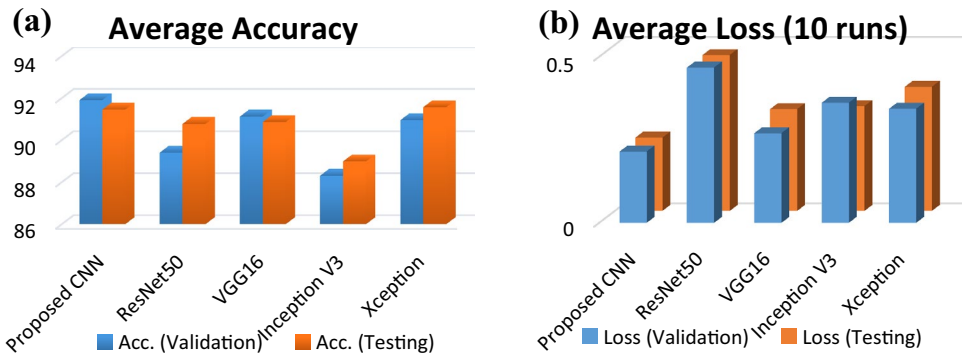

Fig. 6 Overview of **a** classification accuracy and **b** loss (validation and testing) for all the examined CNNs

selection cases and regularization methods, with classification accuracies ranging from 89.78% up to 93.06%. The authors selected the RGB-CNN model with 4 convolutional layers, batch size = 16, dropout = 0.2 and 256-128 dense nodes as the simplest and optimally performed CNN, with respect to testing accuracy and loss.

To further assess the produced results, the classification performance of the proposed model is compared (even indirectly) with that of other CNNs, which were reported in the literature and have been already applied in the specific problem.

Previous works that are highly related to this research study, are those carried out in [18] and [25], regarding bone metastasis classification. Both works employ CNNs for metastatic/non-metastatic hotspot classification, providing classification accuracies up to 89% (see Table 1). According to the related work in bone metastasis diagnosis in prostate cancer patients from whole-body scintigraphy images, there is inadequate formation regarding the classification accuracy of BS when CNN systems are considered.

To highlight the classification performance of the proposed CNN method, the authors performed a straightforward comparison of the accuracy and several other evaluation metrics like precision, recall, sensitivity, specificity, and F1-score of our model with those of other state-of-the-art CNNs, commonly used for image classification problems. From the results listed in Fig. 6, it arises that the proposed algorithm exhibits better or similar performance to other well-known CNN architectures, applied in the specific problem. This study validates the premise that CNNs are algorithms that can offer high accuracy in medical image classification-based problems and can be directly applied on medical imaging where the automatic identification of diseases is crucial for the patients.

The main advantages of the proposed RGB-CNN architecture are described as follows:

(1) RGB-CNN can efficiently train a model on a small dataset: regardless of the relevant literature, in which CNNs can work properly and effectively only when large datasets of medical images are available, in this nuclear medical image analysis problem with a small dataset, the proposed RGB-CNN model is proven to be able to be trained efficiently, too.

(2) RGB-CNN reduces architecture complexity: it comes with a simpler architecture, exhibiting better performance than other well-known CNN architectures in the field of medical image analysis, and also operates with less running time, due to its reduced architectural complexity.

Overall, for the purpose of medical image analysis and classification, the CNN methodology is proven to be sufficiently powerful in the nuclear medicine domain and particularly for BS, outweighing the popular CNN architectures for image analysis, like VGG16, ResNet50, InceptionV3, and Xception. The proposed classification algorithm can be applied in any type of whole-body scans, concerning BS and can be integrated into a CAD system of whole-body diagnosis or a clinical decision support system in nuclear medical imaging.

Some potential limitations of this research study include: (i) the use of a small dataset, since most of the notable accomplishments of deep learning are typically based on very large amounts of data, and (ii) no explanation on the decisions is provided. These limitations will be addressed in our future work.

Future work is oriented in two different directions: (i) to perform a further investigation of the proposed architecture using more images from prostate cancer patients, as well as patients suffering from other types of metastatic cancer, like breast, kidney, lung or thyroid cancer. This will demonstrate the generalization capabilities of the proposed method, and (ii) to offer interpretability and transparency in the classification process, while simultaneously increasing the accuracy of the model, and investigating new efficient AI architectures.

Funding This research received no external funding.

Data availability The datasets analyzed during the current study are available from the corresponding author and nuclear medicine physician on reasonable request.

Compliance with ethical standards

Conflicts of interest The authors declare that there are no conflicts of interest regarding the publication of this paper.

Ethical approval This research work does not report human experimentation; not involve human participants following an experimentation in subjects. All procedures in this study were in accordance with the Declaration of Helsinki.

Informed consent This study was approved by the Board Committee Director of the Diagnostic Medical Center "Diagnostico-Iatriki A.E." Dr. Vasilios Parafestas and the requirement to obtain informed consent was waived by the Director of the Diagnostic Center due to its retrospective nature.

References

1. Coleman RE. Metastatic bone disease: clinical features, pathophysiology and treatment strategies. Cancer Treat Rev. 2001;27:165–76.
2. Lukaszewski B, Nazar J, Goch M, Lukaszewska M, Stępiński A, Jurczyk MU. Diagnostic methods for detection of bone metastases. 2017;21(2):98.
3. Chang CY, Gill CM, Joseph Simeone F, Taneja AK, Huang AJ, Torriani M, et al. Comparison of the diagnostic accuracy of 99 m-Tc-MDP bone scintigraphy and 18 F-FDG PET/CT for the detection of skeletal metastases. Acta Radiol. 2016;57:58–65.
4. Doi K. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. Comput Med Imaging Graph. 2007;31(4–5):198–211.
5. Rieden K. Conventional imaging and computerized tomography in diagnosis of skeletal metastases. Radiologe. 1995;35:15–20.
6. Van Den Wyngaert T, Strobel K, Kampen WU, Kuwert T, Van Der Bruggen W, Mohan HK, et al. The EANM practice guidelines for bone scintigraphy On behalf of the EANM Bone & Joint Committee and the Oncology Committee. European Journal of Nuclear Medicine and Molecular Imaging. 2016;.
7. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88.
8. Biswas M, Kuppili V, Saba L, Edla DR, Suri HS, Cuadrado-Godia E, et al. State-of-the-art review on deep learning in medical imaging. Front Biosci. 2019;24:392–426.
9. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19:221–48.
10. Qian N. On the momentum term in gradient descent learning algorithms. Neural Netw. 1999;12(1):145–51.
11. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks BT—Computer Vision–ECCV 2014. European Conference on Computer Vision (ECCV). Cham: Springer; 2014. p. 818–33.
12. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd International conference on learning representations, ICLR 2015-conference track proceedings. arXiv; 2014. p. 1409–556.
13. Simonyan K, Zisserman A. VGG-16. arXiv preprint. 2014;.
14. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Proceedings of the IEEE computer society conference on computer vision and pattern recognition. 2015. p. 1–9.
15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE computer society conference on computer vision and pattern recognition. 2016. p. 770–8.
16. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proceedings - 30th IEEE conference on computer vision and pattern recognition, CVPR 2017. 2017.
17. Horikoshi H, Kikuchi A, Onoguchi M, Sjöstrand K, Edenbrandt L. Computer-aided diagnosis system for bone scintigrams from Japanese patients: importance of training database. Ann Nucl Med. 2012;26:622–6.
18. Dang J. Classification in Bone Scintigraphy Images Using Convolutional Neural Networks. 2016.
19. Erdi YE, Humm JL, Imbriaco M, Yeung H, Larson SM. Quantitative bone metastases analysis based on image segmentation. J Nucl Med. 1997;38:1401–6.
20. Šajn L, Kononenko I, Milčinski M. Computerized segmentation and diagnostics of whole-body bone scintigrams. Comput Med Imaging Graph. 2007;31:531–41.
21. Elfarra FG, Calin MA, Parasca SV. Computer-aided detection of bone metastasis in bone scintigraphy images using parallelepiped classification method. Ann Nucl Med. 2019;33:866–74.
22. Bradshaw T, Perk T, Chen S, Hyung-Jun I, Cho S, Perlman S, et al. Deep learning for classification of benign and malignant bone lesions in [F-18]NaF PET/CT images. J Nucl Med. 2018;59:327.
23. Furuya S, Kawauchi K, Hirata K, Manabe W, Watanabe S, Kobayashi K, et al. A convolutional neural network-based system to detect malignant findings in FDG PET-CT examinations. J Nucl Med. 2019;60:1210.
24. Furuya S, Kawauchi K, Hirata K, Manabe W, Watanabe S, Kobayashi K, et al. Can CNN detect the location of malignant uptake on FDG PET-CT? J Nucl Med. 2019;60:285.
25. Belcher L. Convolutional neural networks for classification of prostate cancer metastases using bone scan images. 2017.
26. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. 2018;9:611–29.
27. O'Shea KT, Nash R. An Introduction to Convolutional Neural Networks. arXiv preprint [Internet]. 2015; Available from: https://arxiv.org/abs/1511.08458.
28. Komeda Y, Handa H, Watanabe T, Nomura T, Kitahashi M, Sakurai T, et al. Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: preliminary experience. Oncology. 2017;93:30–4.
29. Ma L, Ma C, Liu Y, Wang X. Thyroid diagnosis from SPECT images using convolutional neural network with optimization. Comput Intell Neurosci. 2019;2019:6212759.
30. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 2014.
31. Colaboratory cloud environment supported by Google [Internet]. Google Colab. Available from: https://colab.research.google.com/.
32. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vision. 2015;115(3):211–52.
33. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE computer society conference on computer vision and pattern recognition. 2016.
34. Chollet F. Xception: Deep learning with depthwise separable convolutions. 2016.