

Article

# A Deep-Learning Approach for Diagnosis of Metastatic Breast Cancer in Bones from Whole-Body Scans

Nikolaos Papandrianos <sup>1,\*</sup>, Elpiniki Papageorgiou <sup>2</sup>, Athanasios Anagnostis <sup>3,4</sup> and Anna Feleki <sup>5</sup>

<sup>1</sup> Former Nursing Department, University of Thessaly, 35100 Lamia, Greece

<sup>2</sup> Faculty of Technology, University of Thessaly, Geopolis Campus, Larissa—Trikala Ring Road, 41500 Larissa, Greece; elpinikipapageorgiou@uth.gr

<sup>3</sup> Institute for Bio-economy and Agri-technology (iBO), Center for Research and Technology—Hellas (CERTH), 57001 Thessaloniki, Greece; a.anagnostis@certh.gr

<sup>4</sup> Department of Computer Science, University of Thessaly, 35131 Lamia, Greece

<sup>5</sup> Faculty of Technology, Geopolis Campus, University of Thessaly, 41500 Larissa, Greece; felekiann@gmail.com

\* Correspondence: npapandrianos@uth.gr

Received: 9 December 2019; Accepted: 22 January 2020; Published: 3 February 2020



**Abstract:** (1) Background: Bone metastasis is one of the most frequent diseases in breast, lung and prostate cancer; bone scintigraphy is the primary imaging method of screening that offers the highest sensitivity (95%) regarding metastases. To address the considerable problem of bone metastasis diagnosis, focused on breast cancer patients, artificial intelligence methods devoted to deep-learning algorithms for medical image analysis are investigated in this research work; (2) Methods: Deep learning is a powerful algorithm for automatic classification and diagnosis of medical images whereas its implementation is achieved by the use of convolutional neural networks (CNNs). The purpose of this study is to build a robust CNN model that will be able to classify images of whole-body scans in patients suffering from breast cancer, depending on whether or not they are infected by metastasis of breast cancer; (3) Results: A robust CNN architecture is selected based on CNN exploration performance for bone metastasis diagnosis using whole-body scan images, achieving a high classification accuracy of 92.50%. The best-performing CNN method is compared with other popular and well-known CNN architectures for medical imaging like ResNet50, VGG16, MobileNet, and DenseNet, reported in the literature, providing superior classification accuracy; and (4) Conclusions: Prediction results show the efficacy of the proposed deep learning approach in bone metastasis diagnosis for breast cancer patients in nuclear medicine.

**Keywords:** bone metastasis; breast cancer; whole body; scintigraphy; deep learning; image classification; convolutional neural networks

## 1. Introduction

Bone metastasis is cancer that initially started in another part of the body and spread to the bone. Various names can be found in the literature, such as secondary bone cancer or metastatic bone disease [1,2]. Bone metastasis commonly occurs in the spine, thigh and pelvis but far from that, it can also be seen in any bone. Sometimes, bone metastasis appears years after cancer treatment, or may be the first indication of cancer.

Metastatic cancer in bones is a life-threatening disease. Proper diagnosis and treatment of metastatic cancer in bones can save people's lives [3]. Identification of normal, benign, and malignant tissues seems to be a step of high importance before treatment of bone metastasis tumors. The most

common tumors, which frequently metastasize to bone, are those of the breast, lung and prostate, whereas the skeleton area especially in patients with advanced disease, accumulates most of the tumor burden [4]. In particular, breast cancer and prostate cancer are the most common types of cancer likely to spread to bone [3]. Previous studies clearly report that 60%–75% of metastatic breast cancers are initially diagnosed with bone metastasis [5,6].

Bone metastasis in women who suffer from breast cancer is clinically important [7], and has a significant impact on the quality of their life [8]. Rapid diagnosis of bone metastases can be achieved using modern imaging techniques such as scintigraphy, positron emission tomography (PET) and whole-body magnetic resonance imaging (MRI) [9–11]. Among them, bone scintigraphy offers the highest sensitivity (95%) regarding metastases, making it the primary imaging method of screening, [12–14]. In particular, whole-body bone scintigraphy emerges as the most common diagnostic procedure in nuclear medicine for clinical routine investigation [14–16] and can reveal the sites that bone metastases are distributed in all types of cancer, possibly predicting the outcome [17,18]. It actually presents high sensitivity whereas the changes of bone metabolism are identified by doctors and physicians earlier than the changes in bone structure detected by skeletal radiograms [2,7]. Infections, degenerative changes, benign or malignant diseases and other clinical entities are among indications in bone scintigraphy [16]. This method, however, shows low specificity, as it cannot distinguish between the causes of bone turnovers and those of metastatic origin (leukaemia, healing fracture, etc.), which is a significant drawback of this method.

At present, bone scintigraphy which uses single-photon emission computed tomography (SPECT) imaging, remains the most used and well-known imaging method for clinical diagnosis of metastatic cancer, both men and women [12,13,18]. It turned out to be an effective tool for early diagnosis of bone metastasis in breast and prostate cancer patients [15,17]. In particular, with the help of nuclear medical imaging of the targeted area of the body, which is used for the identification of benign and malignant conditions, the doctor and the physician are able to diagnose bone metastasis [10,19].

To address the considerable problem of bone metastasis diagnosis, machine learning (ML) methods focused on deep-learning algorithms for medical image analysis, have been recently investigated by many researchers. The advancements in the field of machine learning have led to more intelligent and robust computer-aided diagnosis (CAD) systems, as the learning ability of ML methods have been constantly improving [20,21]. Also, the groundbreaking performance of ML in CAD systems is based on deep learning (DL). DL has outstanding capabilities which make a very significant impact on improving the diagnostics potential of CAD systems in medical imaging [22–24].

A recent survey reveals the penetration of deep-learning techniques entirely into the field of medical image analysis in CAD; classification, detection, segmentation, registration, retrieval, image generation and enhancement, whereas the successful application of deep learning to medical imaging tasks is thoroughly examined [22–26]. Specifically, deep learning in medical imaging was achieved using convolutional neural networks (CNNs) [27–30]. CNNs appertain to deep, feed-forward artificial neural networks, able to extract the feature of an image and use this feature to classify the image. The aim of CNNs in all application fields is to convert the input into a feature vector by means of a matrix and to match it with trained feature vectors [31].

Some major and popular CNN models, introduced in published research articles in the context of image processing and deep learning, are: the AlexNet (2012) network [32], which has a very similar architecture to LeNet, by Yann LeCun et al., but is deeper, with more filters per layer and includes stacked convolutional layers, the VGGNet16 (2014), that consists of 16 convolutional layers and its very uniform architecture makes it very appealing [33], the ResNet model, that introduces the residual learning building block for extremely deep convolutional networks [34], the DenseNet (2017) [35] that offers the main advantage of alleviating the gradient vanishment problem with the direct connection of all the layers and, finally, the Deeplab, which deals with the introduction of convolutional layers atrous for semantic image segmentation in deep convolutional neural networks [36]. All the important

and most interesting applications of deep learning in medical image analysis and segmentation are gathered in recent review studies [22–24,29].

### 1.1. Related Research in Breast Cancer Diagnosis Using Convolutional Neural Networks (CNNs)

Regarding the medical image domain which deals with breast cancer diagnosis, previous studies examined the inclusion of machine learning approaches into breast cancer computer-aided diagnosis and image classification [29,37–41]. The state of the art on machine learning techniques for breast cancer computer aided diagnosis offers a wide range of analysis regarding the current status of CAD systems, when image modalities used and machine learning-based classifiers are taken into consideration [38,39,42,43]. Recent review studies in [44–47] presented a thorough survey on both traditional ML and DL literature with particular application in breast cancer diagnosis. Also, the bibliographic review in study [47] provided insightful characteristics of some well-known DL networks in breast cancer diagnosis, both in primary and metastatic breast cancer.

Investigating the area of computer-aided biomedical images applied in breast cancer diagnosis, there are some recent research studies which show the advantageous features of deep convolutional nets [48–50]. In particular, in [51] a CNN was applied for breast cancer histopathology images classification, in two different classes, benign and malignant. Suzuki et al. (2016) implemented CNNs for mass detection in mammographic images in computer-aided diagnosis [52]. Spanhol et al. (2016) applied CNNs for the classification of breast cancer from histopathological images [53], whereas Wichakam and Vateekul, (2016) combined deep convolutional networks and SVMs for mass detection on digital mammograms [54]. Swiderski et al. investigated the application of deep learning and non-negative matrix factorization for breast cancer recognition in mammograms [55]. In the same year, Kallenberg and his colleagues explored the capabilities of unsupervised deep learning to breast density segmentation and mammographic risk scoring [56]. Dealing with breast cancer image detection, CAD systems are fundamentally used in screening mammography for detection of mass [41,48,57,58]. In addition, CNNs have mostly been applied in mammography images and computed tomography (CT) images for classification and segmentation case studies [38,39,46–49]. Furthermore, deep CNNs and transfer-learning models have been applied for the prediction of lymph node metastasis in patients with primary breast cancer on the basis of US images [59], as well as on the basis of histopathological images [60].

It is evident that much research has already been done for the detection and diagnosis of primary breast cancer in the past few years. Some representative CNN applications for breast cancer diagnosis are summarized in Table 1.

**Table 1.** Summary of representative convolutional neural networks (CNNs) for breast cancer (BC) diagnosis.

| Type of BC                | Application Type                  | Modality                 | Reference                  | Data Available |
|---------------------------|-----------------------------------|--------------------------|----------------------------|----------------|
| Primary                   | classification                    | Histopathological images | Kumar K. et al. [51]       | NO             |
| Primary                   | detection                         | Mammograms               | Swiderski, B. et al. [55]  | YES            |
| Primary                   | classification                    | Histopathological        | Spanhol, F.A. et al. [53]  | NO             |
| Primary                   | mass detection                    | Mammograms               | Wichakam, I. et al. [54]   | NO             |
| Primary                   | mass detection                    | Mammograms               | Suzuki, S. et al. [52]     | NO             |
| Primary                   | segmentation                      | Mammograms               | Kallenberg, M. et al. [56] | NO             |
| Primary                   | mass detection                    | Mammograms               | Fenton J.J. et al. [58]    | NO             |
| Primary                   | mass detection and classification | Mammograms               | Cheng H. et al. [41]       | NO             |
| Primary                   | classification                    | Mammograms               | Chougrad, H et al. [57]    | NO             |
| Metastatic in lymph nodes | prediction                        | Ultrasound images        | Zhou Li-Q. et al. [59]     | NO             |
| Metastatic in lymph nodes | detection and classification      | Histopathological        | Steiner D.F. et al. [60]   | NO             |

However, our work is devoted to the investigation of deep-learning algorithms on bone metastasis diagnosis in breast cancer patients using bone scintigraphy. In nuclear medicine, bone scintigraphy and PET/CT are the two most used methods for bone metastasis detection in breast cancer [61–64]. Concerning the involvement and applicability of CNNs in diagnosis of metastatic breast cancer in bones from whole-body scan images, no previous works in CNNs exploration are reported. There are only a few research studies applying neural networks to deal with breast cancer bone metastasis classification in nuclear medicine. In what follows, the related papers of metastatic breast cancer diagnosis in bones using neural network methods are discussed.

### *1.2. Literature Review in Bone Metastasis Diagnosis from Breast Cancer Patients in Nuclear Medicine*

Focusing on bone scintigraphy, which is considered as the well-known imaging method of screening, concerning the diagnosis of bone metastasis, only two previous studies were reported in the literature that deal with metastatic breast cancer classification. The first study was devoted to the development of a computer-aided diagnosis system for bone scintigraphy scans (CADBOSS) [65]. CADBOSS is able to detect metastases in an accurate way, that combines an active contour segmentation algorithm for hotspots detection, an advanced method of image gridding to extract certain characteristics of metastatic regions and finally, an artificial neural network classifier for identifying possible metastases [65]. Another similar research study was conducted during a PhD thesis in [66] which deals with the investigation of CAD system to improve the diagnostic accuracy of planar whole-body bone scan interpretations. During this PhD thesis, four sub studies were accomplished sequentially. The first sub study investigated an automated method that determines the existence of metastases after bone scanning, using neural networks and image-processing techniques. The second study dealt with the discrepancies between observers as well as the performance, in terms of interpretations solely on bone scan images, regarding the existence of bone metastatic disease. The third study resulted in the development of a second CAD system based on improved image processing and artificial neural network to classify bone scans in two categories, considering a larger database of whole-body bone scans. Further investigation took place on how the proposed CAD system can benefit the physicians by reducing inter-observer variation. The final outcome of that thesis was the development of a totally automated computer-assisted diagnosis system that can identify metastases after examining bone scans, applying multi-layer perceptron artificial neural network techniques, and a small database of whole-body bone scans (135 patients). The highest sensitivity achieved from all the studies and accomplished during this thesis was approximately 89% [66]. None of these research studies provide publicly available data.

As there are no other research studies for breast cancer bone metastasis classification in nuclear medicine (and no research paper has been published, either in journals or conferences, which concerns bone scintigraphy analysis and classification, using deep learning), we briefly report some recent studies in nuclear medicine for bone metastasis diagnosis in men suffering from prostate cancer. Bone scintigraphy is a well-recognized image modality for metastatic prostate cancer diagnosis in bones.

### *1.3. Literature Review in Bone Metastasis Classification Using CNNs in Nuclear Medicine*

As far as prostate cancer bone metastasis classification is concerned, most of the studies applying CNNs are devoted to PET/CT imaging and SPECT scans [67–76]. These studies have been gathered in Table 2.

**Table 2.** Summary of CNNs applied in whole-body scans for bone metastasis classification in nuclear medicine imaging.

| Application Type  | Modality          | Reference | Data Available |
|---|-------------------|-----------|----------------|
| Metastatic prostate cancer classification                         | SPECT             | [70]      | NO             |
| Metastatic prostate cancer classification                         | NaF PET/CT images | [71]      | NO             |
| Malignancy detection  | FDG PET-CT        | [72]      | NO             |
| Metastatic prostate cancer classification                         | FDG PET           | [73]      | NO             |
| Patient's sex prediction  | FDG PET-CT        | [74]      | NO             |
| predicts whether physician's further diagnosis is required or not | FDG PET-CT        | [75]      | NO             |
| Segmentation  | SPECT             | [76]      | NO             |

In [70], a master thesis, was the first study devoted to CNN application by correctly classifying metastasis in bone scintigraphy images for prostate cancer. That thesis focused mainly on classification problems without considering any identification and segmentation tasks. The used dataset was provided by Exini Diagnostics AB, in the form of image patches of hotspots already found. Due to time frame restrictions, only the hotspots found in the spine had been used to train the CNN, as those were considered the easiest to classify. Bone scan index (BSI) was calculated for whole-body bone scans, by segmenting the entire skeleton from the background in both the anterior and posterior views. A shape model based on a mean shape of several normal whole-body scans was fitted to the skeleton, using an image analysis algorithm called Morphon registration. The outcomes of this master thesis [70] showed that the calculated accuracy of the validation set was 0.875, whereas the calculated accuracy of the testing set was 0.89.

In PET/CT imaging, a deep convolutional neural network has been implemented to distinguish benign from malignant lesions in NaF PET/CT images of patients with metastatic prostate cancer [71]. Also, a CNN-based system was examined to detect malignant findings in FDG (fluorodeoxyglucose)PET-CT examinations in a retrospective study [72]. Various CNN models with equivalent configurations, such as ResNet, DenseNet and Visual Geometry Group (VGG), were used to classify whole-body FDG PET images in three categories of benign, malignant and equivocal patient cases. Location of malignant uptake was classified at the same time into, (A) head-and-neck, (B) chest, (C) abdomen, (D) pelvic region. Different predictions were derived for each location [73].

In another research work [74], a simple CNN model with an equivalent configuration to VGG, that predicts patient sex from FDG PET-CT images was proposed. Kawauchi and his colleagues in [75] proposed an alternative CNN-based diagnosis system for whole-body FDG PET-CT that predicts whether physician's further diagnosis is required or not. A thorough analysis of the results showed an accuracy of  $93.2\% \pm 3.9\%$  regarding images of patients with malignant uptake while the respective accuracy for images of equivocal were  $87.8\% \pm 5.3\%$ . Furthermore, the task of segmentation in skeletal scintigraphy images with deep learning models has been discussed in the study of [76], in which the authors examined different approaches to convert convolutional neural networks, designed for classification tasks, into powerful pixel-wise predictors. However, this research domain of bone scintigraphy segmentation has not been established well yet with the inclusion of advanced deep neural networks.

Reviewing the relevant literature for diagnosis of bone metastasis using whole body images, the authors notice that all previous works focus mainly on PET imaging and prostate metastatic bone cancer, while no research paper seems to have been published, dealing with bone scintigraphy analysis and classification, in patients with breast cancer, using CNNs.

#### 1.4. Motivation and Aim of This Research Study

Although bone scintigraphy is extremely important for diagnosis of metastatic cancer in bones, in both men and women, there is currently no research paper regarding the classification of breast cancer metastasis in bones from whole-body images (bone scintigraphy) that uses efficient and robust CNNs.

Nowadays, the main challenge in bone scintigraphy, which is considered one of the most sensitive methods for imaging in nuclear medicine, regards building an algorithm that automatically identifies whether a patient is suffering from bone metastasis or not, only by looking at the whole-body scans. The algorithm has to be extremely accurate since people's lives could be at stake. Machine-learning approaches, with a focus on deep learning algorithms, have particularly shown a promising applicability in medical image analysis in the area of nuclear medicine. However, classification accuracy of bone scintigraphy, analyzed by machine-learning approaches, involving CNN systems, has not been well established yet.

The current work introduces an efficient CNN method for the diagnosis of bone metastasis disease, by identifying whether a patient is suffering from bone metastasis or not, based on whole-body images. After a thorough CNN exploration, taking into consideration both network architecture and hyperparameter configuration, the researchers propose a robust CNN to classify images in two classes, benign and malignant. A retrospective study involving 408 women patients suffering from breast cancer, was conducted to examine the accuracy of bone scintigraphy analyzed by CNNs. For validation purposes of the proposed classification methodology, a meticulous comparison analysis was performed between the proposed CNN method and a number of well-known and popular CNN architectures. In this work, popular and advanced CNN structures such as ResNet50, VGG16, and DenseNet that were previously proposed for medical image classification [22–24,30], are applied in the dataset, after their necessary parameterization and configuration. The experimental results reveal that the proposed method achieves superior performance as far as bone scintigraphy is concerned, outperforming other well-known methods of CNNs.

The novel contribution of this research work lies in the creation of a robust CNN model (being the most efficient deep learning method in medical image analysis) for bone metastasis diagnosis from breast cancer patients using whole-body scans. Following the literature review and the state-of-the-art deep learning methodology (as presented in Sections 1.1 and 1.2), it is obvious that there is no previous work on metastatic breast cancer identification in bones from whole-body scans using deep learning and CNNs.

The innovations and contributions of this paper are summarized as follows:

- The creation and demonstration of a customized, robust CNN-based classification tool for identification of metastatic breast cancer in bones from whole-body scans.
- The meticulous exploration of CNN hyper-parameter selection to define the best architecture and select the RGB mode selection that could lead to enhanced classification performance.
- The comparative experimental analysis performed for utilizing popular image classification CNN architectures, like ResNet50, VGG16, GoogleNET, etc.

The proposed methodology apparently improves the diagnostic effect of the deep-learning method. Moreover, after a thorough assessment of the results produced, the proposed approach is proven to be effective for classification of whole-body images in nuclear medicine.

This paper is structured in the following fashion: Section 2 presents the material and methods related to this research study. The proposed network based on CNN for bone metastasis diagnosis in breast cancer using whole body images is described in Section 3. The exploration analysis of CNNs in both RGB and grayscale modes is illustrated in Section 4, considering different parameters and configurations, thus providing the best CNN model for this case study. Section 4 also gathers all the performed experiments with the corresponding results, whereas Section 5 provides a thorough discussion about the analysis of results. The main outcomes of the paper are concluded in Section 6, which outlines the future steps, as well.

## 2. Materials and Methods

### 2.1. Breast Cancer Patient Images

This study was approved by the Director of the Diagnostic Medical Center “Diagnostico Iatriki A.E.” and the requirement to obtain informed consent was also waived by the Director of the Diagnostic Center, due to the study’s retrospective nature. All procedures in this work were in accordance with the Declaration of Helsinki. A retrospective review of 422 consecutive, previously selected whole-body scintigraphy images from 382 different patients (women) who visited the Nuclear Medicine Department of Diagnostic Medical Center “Diagnostico Iatriki A.E.” in Larissa, Greece, between June 2013 and December 2017, was performed. The selection criterion were breast cancer patients who had undergone whole-body scintigraphy, because of suspected bone metastatic disease.

Due to the fact that there are various remains and non-osseous uptake presented in images (i.e., urine contamination and medical accessories), as well as the frequent visible site of radiopharmaceutical injection [77], a preprocessing approach was accomplished to remove these artifacts from the original images. This preprocessing method was accomplished by a nuclear medicine physician, before the use of the dataset in the proposed classification approach.

Next, the selected images were diagnosed accordingly by a Nuclear Medicine Specialist with 15 years’ experience in bone scan interpretation. In total, 408 bone scan images were selected by the nuclear medicine specialist and classified accordingly into two categories (1) malignant and (2) benign; 221 bone scans concern malignant images whereas 187 bone scans concern benign images without bone metastasis (see Figure 1).

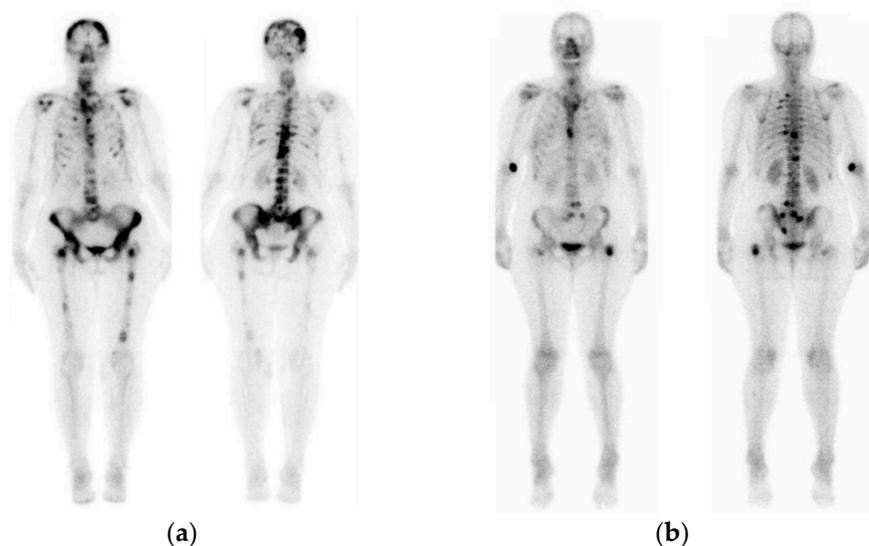
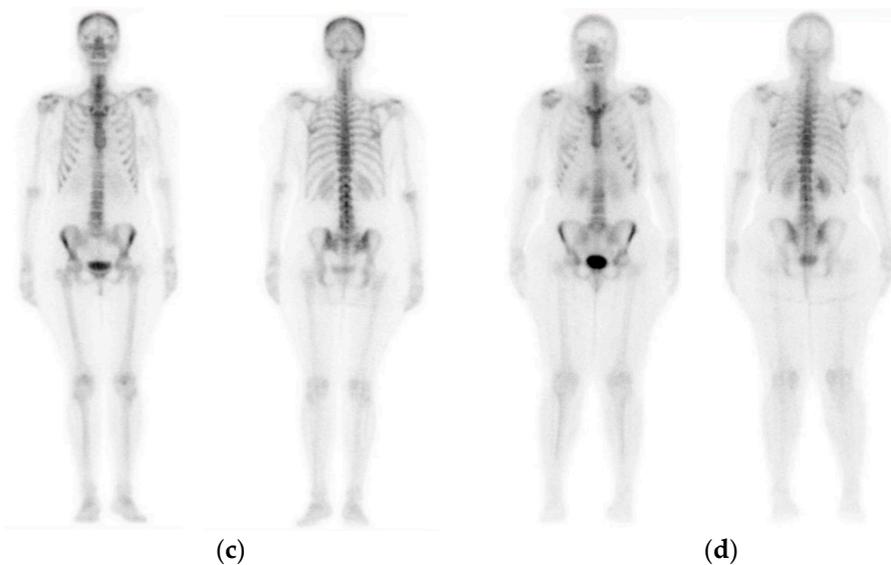


Figure 1. Cont.



**Figure 1.** Image breast cancer bone metastasis samples in the dataset (Label: (a,b) malignant and (c,d) benign).

The equipment used for scanning patients is a Siemens gamma camera Symbia S series SPECT System (by dedicated workstation and software Syngo VE32B) with two heads with low-energy high-resolution (LEHR) collimators, where the speed of scanning is 12 cm/min with no pixel zooming. Regarding the procedure followed for bone scintigraphy, two types of radionuclide were used, the  $^{99m}\text{Tc}$ -HDP (TechnoScan<sup>®</sup>) and the  $^{99}\text{Tc}$ -MDP (PoltechMDP 5mg). Approximately 3 h after an intravenous injection of radiopharmaceutical agent, a whole-body scintigraphy was acquired. The dosage of radiopharmaceutical agent ranged between 600 and 740 MBq depending on the body type of the patient, whereas the common intravenous injection is 670 MBq.

After bone scintigraphy, 586 planar bone scan images from patients with known P-Ca were reviewed, retrospectively. Anterior as well as posterior views of  $1024 \times 256$ -pixel resolution were digitally recorded using the whole-body field. The number of detected gamma decays in each spatial unit is represented by images of a 16-bit grayscale depth.

## 2.2. Main Aspects of Convolutional Neural Networks

One of the most powerful deep-learning techniques is the use of CNNs and they are common in analyzing and classifying images. Their architecture consists of the perceptron model, which means that it has fully connected layers, with each neuron connecting to every other neuron to the next layer. It is interesting to examine all the different types of layers that it contains.

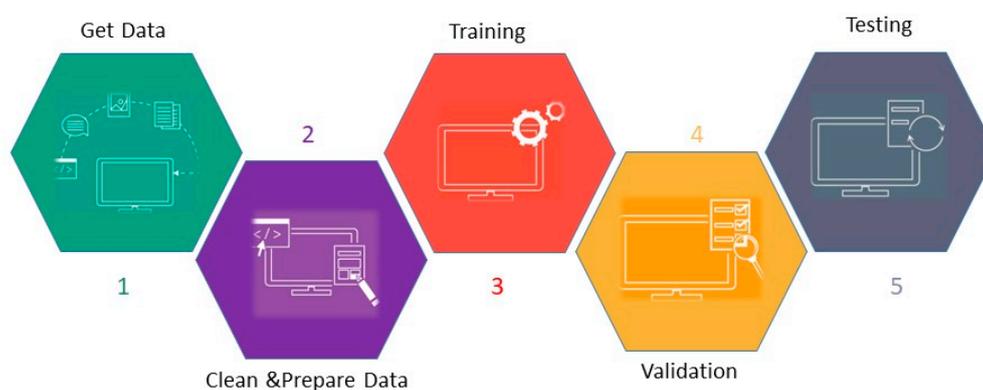
First, there is the convolutional layer, whose name is similar to the name of the neural network. This specific type of layer is fundamental for CNNs, because it creates activation maps. That means that it learns the patterns of the image, helping the algorithm to classify any new image based on these patterns [78,79].

Second comes the pooling layer, which is very important for down-sampling the image and removing any noise that may fuzzy the algorithm. It works on a specific threshold, where it keeps any pixel value higher than the threshold and discards any value lower than the threshold.

The last layer of a CNN is the fully connected layer, which “flattens” the output of previous node, which in definition, means that it converts it into a vector, so it will be ready for an output. Following this, our algorithm gives a label to each image. There are subcategories to the fully connected layer, as the first one does the vectorization and the last one gives the probabilities for each of the given class. [80–82].

### 2.3. Methodology

In medical analysis, the use of CNNs is common, so we can identify with an automated algorithm whether a patient has bone cancer metastasis, or not. The CNN method for bone metastasis classification includes five processing stages (see Figure 2): obtain data from whole-body scans, data pre-processing to normalize the collected image data, CNN training, CNN validation and finally, CNN testing followed by the evaluation of the classification results, as presented in Section 4.3. The five stages of the proposed methodology are thoroughly presented in the following.



**Figure 2.** Five stages of the proposed methodology.

#### Stage 1: Obtain Data

The whole-body scans given by the nuclear medicine physician were in RGB (red, green, blue) mode and were stored in the PC memory. Depending on their class, whether they were normal or malignant, we assigned to them the equivalent prefix. This technique helped us to identify them inside the deep-learning algorithm and to pass them the correct output. Specifically we gave 0 to malignant images as output and 1 to benign ones.

#### Stage 2: Clean and Prepare Data

##### Step 2.1: Data normalization

Data Normalization is used in every machine learning process. This procedure rescales the values of the data within the range of 0 and 1, while it is important to know or estimate accurately, the minimum and maximum observable values. The Min-Max normalization method has been selected as the most popular method, as regards the performance of the systems being examined. Moreover, reviewing the relevant literature, studies showed that the results' accuracy from this method is higher than of those being produced by other normalization methods [83]. The scale that all images are within zero and one also prevents having outliers which will confuse the algorithm.

##### Step 2.2: Data shuffle

From splitting the dataset into training and testing emerges the danger of choosing wrong samples for our algorithm, and so the solution to this problem is provided by the application of shuffling method. That gives a random order to the data.

##### Step 2.3: Data split

We split the available image data into three parts, which are training, validation and testing. Each partition has specific percentage, which had been taken from the whole dataset, meaning that 15% of the dataset was given to testing and 85% is for the other two parts, namely we split it 80% for training and 20% for validation. Therefore, three different datasets are created: Training, Validation and Testing datasets, respectively.

#### Stage 3: Training

##### Step 3.1: Data augmentation

The method of data augmentation is usually used when we have a small number of data and we do not want to create a model that does not have generalization capabilities. For instance, with

the techniques of rescale, rotation range, zoom range and flip, we create variant images and we avoid overfitting.

#### Step 3.2: Define CNN architecture

To define a proper architecture of a CNN for image classification, an exploration process is followed. In the experimentation and trial phase, the most important parameters that can lead to effective network architecture were explored; the number and the type of convolutional layers, the number of nodes, the number of pooling layers, the filters, the drop rate and the batch size, the number of dense nodes.

#### Step 3.3: Define the functions of the CNN architecture

After a search in the relevant literature, the main functions of the CNN architecture are the activation function which is the function that defines the output of the layer, and the loss function which is the function that is used for the optimization of the network's weights. After thorough experiments illustrated in previous works, ReLU (rectified linear unit) is the most used activation function for convolutional layers and dense layers. As regards the output layer, Sigmoid is one of the most commonly used functions [78].

#### Step 3.4: Train CNN

As the name indicates, training a CNN means training the model with images offering the opportunity to learn patterns, as the more differentiated the training is, the better the model becomes. The training process begins when the algorithm tries to create a function that describes the desired relation, based on the training data. Afterwards, it makes predictions based on this function (it is an error function) and moves to the validation step.

#### Stage 4: Validation

The validation dataset is used exclusively for the validation process. During validation, the weights are normalized, and the algorithm makes predictions on already known data. During the validation process, the main aim is to minimize the error which reveals the prediction efficiency of the proposed method.

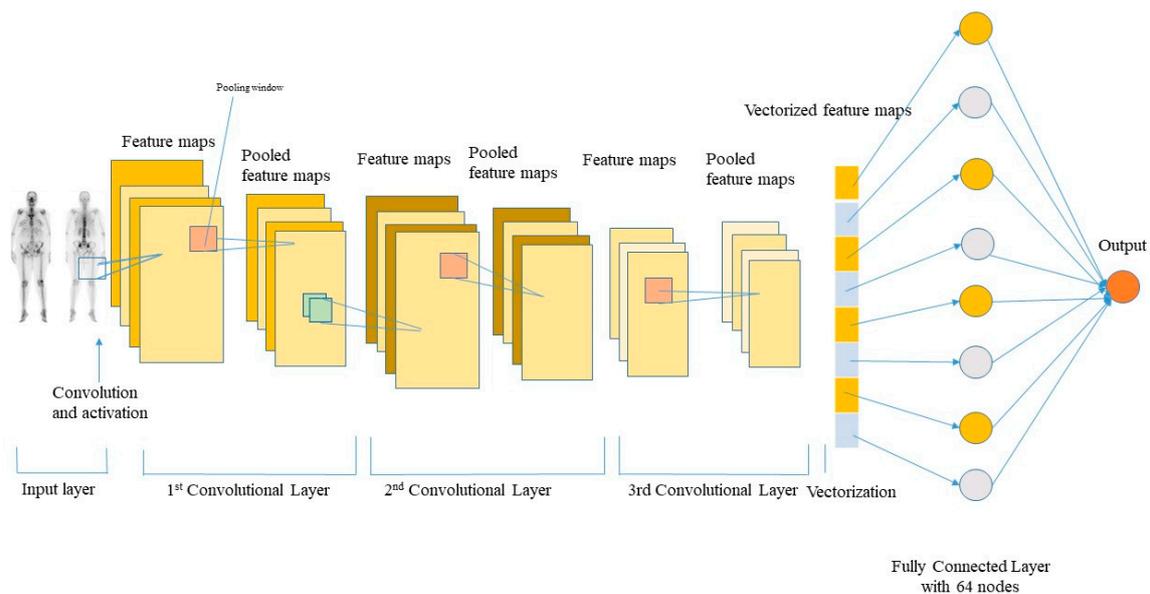
#### Stage 5: Testing/Evaluation

After the completion of the training and validation phases, the best learned model is deployed for the testing/evaluation process [84,85]. The evaluation process for the CNN model is accomplished using the testing data, as initially split, which is completely unknown to the model. During the testing phase, we check whether our classifier can make correct labeling to bot familiar images thus assessing the predictive ability of the current model. The classifier makes predictions on each image's class and finally compares the calculated predicted class to the true class. Next, for classifier evaluation, some well-known and popular performance metrics, such as the testing accuracy, precision, recall, sensitivity, specificity and f1-score of the model are computed, and accompanied by employing an error/confusion matrix for further model evaluation [84,86,87].

### 3. The Proposed CNN Architecture

The main goal of this research study is to create a deep neural network, and particularly, a CNN to precisely identify bone metastasis from whole-body scans, in women suffering from breast cancer. The developed CNN, which was applied in bone scintigraphy to automatically identify bone metastasis, will prove its ability to provide high accuracy with low computation time for whole-body image prediction. In the CNN exploration process, we had to experiment with different values for our parameters like pixels, epochs, drop rate, batch size, number of nodes and layers [79,80,88,89].

After a large number of experiments conducted with the aforementioned values, the best CNN architecture was derived. Actually, the best architecture consists of three convolutional layers followed by max pooling and dropout layer every time. Following the above layers, flatten comes next, along with the dropout layer. Then one dense layer is shown, and, in the end, the output layer takes over (see Figure 3). Table 1 summarizes all the information about the relevant architecture.



**Figure 3.** The CNN framework for bone scintigraphy classification.

The proposed CNN architecture has an image input size of  $256 \times 256$  pixels, and we pass the images in RGB mode. Three convolutional layers are present, where the max pooling and the dropout layers follow each one of them. It needs to be highlighted that all convolutional layers use ReLU as activation function and the values of filters are doubled for every convolutional layer. This means that the first convolutional layer has 8 filters, the second layer has 16 layers and so on, while all of them have a  $3 \times 3$  pixels kernel size. Furthermore, all pooling layers retain their max values and have  $2 \times 2$  pixels size. In addition, dropout layers have the value 0.7 as their drop rate parameter. Next, the flatten layer follows to convert the data into one dimension; hence, the data are ready for the fully connected layer [90].

In the next phase, the dropout layer takes over, having the same value for drop rate, and helps the model to avoid overfitting. Next, the fully connected layer takes control of the data, comprising 64 nodes and ReLU as activation function. Finally, yet essentially, there is the output layer which has one (1) node as it deals with binary classification and contains the Sigmoid as activation function.

In the following figure (Figure 4), the proposed CNN architecture is illustrated, which produces the best results distinguishing whether a person has bone metastasis or not. We concluded to this architecture after experimenting with different number of epochs, pixels, drop rates, number of nodes, number of layers, and so we infer that it performs in a perfect way, having reached a very high accuracy with low value of loss.

| Layers          | Output    | Description                       |
|-----------------|-----------|-----------------------------------|
| conv2d_1        | 254 x 254 | Kernel: 3 x 3, Filters : 8, ReLU  |
| max_pooling2d_1 | 127 x 127 | 2 x 2 max pooling                 |
| dropout_1       | 127 x 127 | Drop rate = 0.7                   |
| conv2d_2        | 125 x 125 | Kernel: 3 x 3, Filters : 16, ReLU |
| max_pooling2d_2 | 62 x 62   | 2 x 2 max pooling                 |
| dropout_2       | 62 x 62   | Drop rate = 0.7                   |
| conv2d_3        | 60 x 60   | Kernel: 3 x 3, Filters : 32, ReLU |
| max_pooling2d_3 | 30 x 30   | 2 x 2 max pooling                 |
| dropout_3       | 30 x 30   | Drop rate = 0.7                   |
| Flatten_1       | 28800     |                                   |
| dropout_4       | 28800     | Drop rate = 0.7                   |
| dense_1         | 64        | Nodes: 64, ReLU                   |
| dense_2         | 1         | Nodes: 1, Sigmoid                 |

Figure 4. The selected CNN architecture for diagnosis of metastatic breast cancer in bones.

## 4. Results

In this section, we have gathered the most representative and useful results of the conducted experiments and CNN exploration, after 10 executions. The final results are the averaged values of those produced from the experiments after each experiment was performed for 10 times.

This experiment was performed in Colaboratory, called Google Colab [91], which is a free Jupyter notebook environment in the cloud. The main reason for selecting this cloud environment of Google Colab is that supports free GPU. The frameworks Keras 2.0.2 and TensorFlow 2.0.0. were used, python language 3.7, CNN with structure (Convolution layer, 3; Maxpooling layer, 3) and Adam Optimizer.

Other CNN methods, including VGG16 [33], ResNet50 [34], DenseNet121 [35], and MobileNet [92], were also implemented so as a comparative analysis could be performed between them and the method proposed in this paper.

### 4.1. Classification Performance Evaluation

In the evaluation phase, we use specific metrics to see how our algorithm worked with the given images. The validation metrics that were used are accuracy, precision, recall, F1 score, sensitivity and specificity.

A predicted sample can be classified in one of the four states in binary classification problems: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). True positive means that the label of the image is benign, and it is classified correctly. False positive means that the label of the image is not benign, and it is classified as benign. True negative means that the label of the image is not benign, and it is classified as not benign. Similarly, false negative means that the label of the image is benign, and it is classified as malignant.

Mathematical formulations of the six performance metrics are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = \frac{2 * (Recall \times Precision)}{Recall + Precision} \quad (4)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (6)$$

Accuracy is the ratio between the correct labeled samples and all samples. Precision points out the ratio of corrected images of positive class to all images that were predicted for the same category. Recall deals with the division of corrected positive images to all corrected images of both categories.

F1-score is a harmonic mean between recall and precision. Sensitivity is the proportion of images which test positive for the malignancy (bone metastasis) among those which have the metastasis/malignancy. Finally, we have specificity, which is the balance between the corrected negative images and the sum of numerator and wrong labeled positive images.

#### 4.2. Confusion Matrix

It is common to use a confusion matrix in machine-learning problems because it is an efficient way to observe the performance of an algorithm, which in our case is image classification. It is a useful evaluation metric for calculating the errors made by the classifier and gives precisely the number of images that were classified at the wrong label. In this study the best confusion matrix of different CNN architectures is produced with the least classification errors.

#### 4.3. Presentation of Results

OpenCV 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. Python was used for coding, with the CNN being programmed with Keras (with Tensorflow [92]), the data normalization, data splitting, confusion matrices and classification reports with Sci-Kit Learn. The computations ranged between 2' to 4' per training for grayscale images and 2' to 5' per training (epoch) for RGB images (256 × 256 pixels), depending on the different input, with the grayscale images being the fastest and RGB images being the slowest to train.

To accomplish the aim of this study, we initially employed different CNN-based architectures and hyperparameter selection as defined in Section 3. After a thorough CNN exploration analysis, we concluded that the best CNN configuration was the following: a CNN with 3 convolutional layers, starting from 8 filters for the first layer, and for each convolutional layer that comes next, the number of filters is doubled (8->16->32), just like the following max-pooling layers. All of those filters have dimensions of 3 × 3. This set-up is shown in Table 1.

Following the steps described in Section 2.2, first we load the images and select the image type, RGB or grayscale. The images are in RGB mode by default and only after requesting are they transformed to grayscale mode. The second and important step is to normalize them, because we want to exclude any outliers and fuzzy data from our dataset. Furthermore, we should shuffle to randomize the order and then perform data augmentation to create variance to our data. Following that, we split our dataset into training, validation and testing and then we pass our images through the CNN. The proposed CNN network has been pre-trained on the ImageNet data set [93]. This data set has proven to be of high utility as it can provide another, more efficient method of weights initialization, in any image-related task.

We followed a two-class classification problem, for bone metastasis presence and absence in patients suffering from breast cancer, considering in total 408 samples of women patients.

Through our experimental analysis, working initially in RGB mode, we conducted experiments with various drop rates, epochs, number of dense nodes, pixel sizes and batch sizes. Different values for pixels were examined, such as 100 × 100, 200 × 200, 256 × 256, 300 × 300 and various values for batch sizes such as 8, 16 and 32 were investigated. In addition, variant drop rate values were studied, for example 0.2, 0.4, 0.7 and 0.9 and a divergent number of dense nodes like 16, 32 and 64 were explored. Epochs were explored from 100, 200, 300 and 500.

It is important to mention that we accomplished more runs with different epochs and pixels and we observed that for epochs = 200, for pixels (256,256), batch size 8, drop rate 0.7 and 64 nodes in dense layer, the CNN models worked better concerning the classification accuracy. In Appendix A, we provide some example simulations for different pixel sizes by also modifying the drop rate and batch size. Also, in Appendix A, some indicative results with different pixel sizes (100, 100), (200, 200) and (300, 300) are illustrated in Tables A1–A3 respectively.

After careful selection of number of epochs and pixels size (epochs = 200 and pixel = 256 × 256), we conducted further experiments with varying drop rates and batch sizes, to find out the best performance parameters.

The following Tables gather the performance analysis and results for the various CNN models with epochs = 200 and pixel = 256 × 256 (in Google Colab [91]).

After a thorough CNN exploration, the best CNN model has the following characteristics: batch-size = 8, dropout = 0.7, nodes (3 conv layers) = 8, 16, 32, Dense Nodes: 64, Epochs: 200 and Pixel = (256,256, 3). Figure 5 represents the accuracy and loss precision curves for the best CNN model.

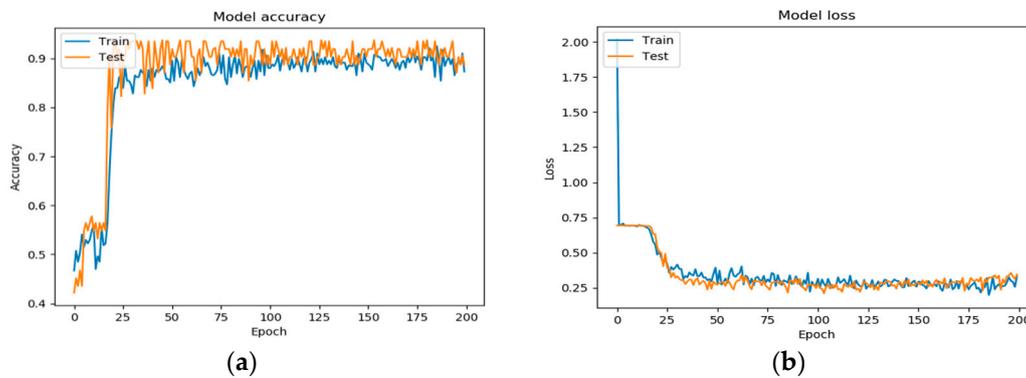


Figure 5. Precision curves for best CNN in RGB mode: (a) Accuracy and (b) Loss.

From the following Tables (see Tables 3–6), it is concluded that the best CNN model is the model with dropout = 0.7 and batch size = 8. The average time for 200 epochs is 788 s.

Table 3. CNN model 1 (epochs = 200, dropout = 0.2, pixel = 256 × 256) runs for different batch sizes.

|         | Batch Size = 8 |            |              |             | Batch Size = 16 |             |              |             | Batch Size = 32 |              |              |              |
|---------|----------------|------------|--------------|-------------|-----------------|-------------|--------------|-------------|-----------------|--------------|--------------|--------------|
|         | Acc.           | Loss       | Acc          | Loss        | Acc.            | Loss        | Acc          | Loss        | Acc.            | Loss         | Acc          | Loss         |
| Run 1   | 85.94          | 0.22       | 91.1         | 0.27        | 84.37           | 0.47        | 89.58        | 0.35        | 98.06           | 0.23         | 87.5         | 0.17         |
| Run 2   | 79.69          | 0.65       | 75           | 1.1         | 84.37           | 0.45        | 79.17        | 0.54        | 92.18           | 0.18         | 84.37        | 0.28         |
| Run 3   | 87.5           | 0.3        | 98.21        | 0.1         | 87.5            | 0.58        | 85.42        | 0.41        | 82.81           | 0.28         | 81.25        | 0.39         |
| Run 4   | 89.1           | 0.49       | 92.86        | 0.38        | 92.18           | 0.21        | 85.42        | 0.33        | 85.93           | 0.34         | 90.62        | 0.49         |
| Run 5   | 92.19          | 0.34       | 82.15        | 0.6         | 93.75           | 0.21        | 91.7         | 0.27        | 98.43           | 0.008        | 90.62        | 0.18         |
| Average | <b>86.89</b>   | <b>0.4</b> | <b>87.86</b> | <b>0.49</b> | <b>88.43</b>    | <b>0.38</b> | <b>86.26</b> | <b>0.38</b> | <b>91.48</b>    | <b>0.207</b> | <b>86.87</b> | <b>0.305</b> |

**Table 4.** CNN model 2 (epochs = 200, dropout = 0.7, pixel = 256 × 256) runs for different batch sizes.

|         | Batch Size = 8 |              |              |              | Batch Size = 16 |          |          |           | Batch Size = 32 |          |          |           |
|---------|----------------|--------------|--------------|--------------|-----------------|----------|----------|-----------|-----------------|----------|----------|-----------|
|         | Acc. Val       | Loss Val     | Acc Test     | Loss Test    | Acc. Val        | Loss Val | Acc Test | Loss Test | Acc. Val        | Loss Val | Acc Test | Loss Test |
| Run 1   | 93.75          | 0.21         | 89.28        | 0.37         | 93.75           | 0.15     | 83.33    | 0.37      | 89.06           | 0.42     | 93.75    | 0.207     |
| Run 2   | 95.31          | 0.35         | 94.64        | 0.26         | 90.62           | 0.29     | 89.58    | 0.23      | 95.31           | 0.16     | 87.5     | 0.33      |
| Run 3   | 92.18          | 0.23         | 92.86        | 0.18         | 95.31           | 0.16     | 93.75    | 0.22      | 90.62           | 0.22     | 93.75    | 0.18      |
| Run 4   | 93.75          | 0.24         | 92.85        | 0.19         | 89.1            | 0.29     | 89.58    | 0.22      | 95.31           | 0.15     | 87.5     | 0.314     |
| Run 5   | 87.5           | 0.29         | 92.85        | 0.24         | 90.62           | 0.23     | 87.51    | 0.26      | 92.18           | 0.24     | 93.75    | 0.239     |
| Run 6   | 89.06          | 0.3          | 89.28        | 0.22         | 89.1            | 0.29     | 97.92    | 0.14      | 82.81           | 0.35     | 90.62    | 0.227     |
| Run 7   | 90.62          | 0.287        | 94.64        | 0.21         | 84.38           | 0.32     | 95.83    | 0.2       | 98.43           | 0.16     | 93.75    | 0.21      |
| Run 8   | 92.18          | 0.22         | 87.5         | 0.25         | 92.18           | 0.23     | 93.75    | 0.19      | 85.93           | 0.28     | 87.5     | 0.232     |
| Run 9   | 84.37          | 0.33         | 96.42        | 0.16         | 85.7            | 0.32     | 95.83    | 0.24      | 90.62           | 0.21     | 93.75    | 0.19      |
| Run 10  | 93.75          | 0.176        | 94.64        | 0.154        | 89.06           | 0.37     | 87.5     | 0.26      | 85.93           | 0.39     | 87.5     | 0.32      |
| Average | <b>91.25</b>   | <b>0.263</b> | <b>92.49</b> | <b>0.224</b> | 89.98           | 0.27     | 91.46    | 0.23      | 90.62           | 0.26     | 90.94    | 0.25      |

**Table 5.** CNN model (epochs = 200, pixel = 256 × 256, conv (8->16->32), dense = 64) runs for different dropouts.

| Batch Size = 8 | Dropout = 0.2 |            |              |             | Dropout = 0.4 |             |              |            | Dropout = 0.9 |             |              |            |
|----------------|---------------|------------|--------------|-------------|---------------|-------------|--------------|------------|---------------|-------------|--------------|------------|
|                | Acc. Val      | Loss Val   | Acc Test     | Loss Test   | Acc. Val      | Loss Val    | Acc Test     | Loss Test  | Acc. Val      | Loss Val    | Acc Test     | Loss Test  |
| Run 1          | 85.94         | 0.22       | 91.1         | 0.27        | 93.75         | 0.22        | 93.75        | 0.23       | 95.31         | 0.31        | 85.71        | 0.43       |
| Run 2          | 79.69         | 0.65       | 75           | 1.1         | 87.5          | 0.42        | 85.71        | 0.32       | 82.81         | 0.37        | 82.14        | 0.38       |
| Run 3          | 87.5          | 0.3        | 98.21        | 0.1         | 84.37         | 0.34        | 89.28        | 0.23       | 79.69         | 0.47        | 73.21        | 0.48       |
| Run 4          | 89.1          | 0.49       | 92.86        | 0.38        | 85.94         | 0.29        | 89.28        | 0.23       | 87.5          | 0.36        | 80.36        | 0.43       |
| Run 5          | 92.19         | 0.34       | 82.15        | 0.6         | 85.94         | 0.23        | 83.93        | 0.39       | 84.37         | 0.36        | 89.28        | 0.27       |
| Average        | <b>86.89</b>  | <b>0.4</b> | <b>87.86</b> | <b>0.49</b> | <b>87.52</b>  | <b>0.32</b> | <b>87.68</b> | <b>0.3</b> | <b>85.94</b>  | <b>0.37</b> | <b>82.14</b> | <b>0.4</b> |

**Table 6.** CNN model (epochs = 200, pixel = 256 × 256, conv (8->16->32), dropout = 0.7) runs for different dense nodes.

| Batch size = 8 | Dense Nodes = 16 |             |              |             | Dense Nodes = 32 |             |              |             | Dense Nodes = 64 |             |              |             |
|----------------|------------------|-------------|--------------|-------------|------------------|-------------|--------------|-------------|------------------|-------------|--------------|-------------|
|                | Acc. Val         | Loss Val    | Acc Test     | Loss Test   | Acc. Val         | Loss Val    | Acc Test     | Loss Test   | Acc. Val         | Loss Val    | Acc Test     | Loss Test   |
| Run 1          | 95.31            | 0.19        | 91.1         | 0.18        | 85.94            | 0.39        | 91.1         | 0.23        | 93.75            | 0.21        | 89.28        | 0.37        |
| Run 2          | 95.31            | 0.23        | 91.1         | 0.22        | 92.19            | 0.18        | 85.71        | 0.25        | 95.31            | 0.35        | 94.64        | 0.26        |
| Run 3          | 85.94            | 0.29        | 87.5         | 0.32        | 92.19            | 0.25        | 87.50        | 0.27        | 92.18            | 0.237       | 92.86        | 0.18        |
| Run 4          | 95.31            | 0.17        | 91.1         | 0.23        | 90.62            | 0.23        | 91.10        | 0.27        | 93.75            | 0.238       | 92.86        | 0.19        |
| Run 5          | 87.5             | 0.27        | 92.86        | 0.16        | 90.62            | 0.22        | 94.64        | 0.21        | 87.5             | 0.29        | 92.85        | 0.24        |
| Average        | <b>91.87</b>     | <b>0.23</b> | <b>90.73</b> | <b>0.22</b> | <b>90.31</b>     | <b>0.25</b> | <b>90.01</b> | <b>0.25</b> | <b>92.50</b>     | <b>0.27</b> | <b>92.50</b> | <b>0.25</b> |

Table 7 depicts the performance metrics for both malignant (metastasis) and benign (no metastasis) patients for this network configuration. The best confusion matrix is illustrated in Table 8.

To further investigate the performance of the proposed CNN architecture, the grayscale mode was applied, for various CNN parameters with respect to drop-rate, batch size, dense nodes, pixel sizes and number of epochs. The same CNN exploration as in RGB mode was followed. After a thorough CNN exploration for grayscale mode, the best network parameters were selected: 3 convolutional layers (2, 4, 8), dense = 16, pixel size 150 × 150 and drop-rate = 0.2, epochs = 200, batch size = 8. The results for the best network performance for the bone metastasis classification problem, concerning the grayscale mode images, are gathered in Table 9. Table 10 depicts the average values of the five performance metrics and the best produced confusion matrix for grayscale mode. Figure 6 illustrates the plots of precision curves concerning accuracy and loss for the best CNN in grayscale mode. The best time was Best: 196 s, Average: 205 s, Steps: 28–30 step. Also, Figure 7 depicts the average prediction errors for the examined CNN architectures in RGB and Grayscale mode, showing the outperformance of the best CNN architecture with dropout=0.7.

**Table 7.** Performance metrics for the best CNN for 10 runs (batch size = 8, dropout = 0.7).

| Runs           |           | Precision | Recall | F1-Score | Sensitivity | Specificity |
|----------------|-----------|-----------|--------|----------|-------------|-------------|
| Run 1          | Malignant | 0.94      | 0.89   | 0.92     | 0.89        | 0.92        |
|                | Benign    | 0.85      | 0.92   | 0.88     | 0.85        | 0.94        |
| Run 2          | Malignant | 0.93      | 1.00   | 0.96     | 1.00        | 0.88        |
|                | Benign    | 1.00      | 0.88   | 0.93     | 0.88        | 1.00        |
| Run 3          | Malignant | 0.9       | 1.00   | 0.95     | 1.00        | 0.85        |
|                | Benign    | 1.00      | 0.85   | 0.92     | 0.85        | 1.00        |
| Run 4          | Malignant | 1.00      | 0.88   | 0.93     | 0.88        | 1.00        |
|                | Benign    | 0.88      | 1.00   | 0.94     | 1.00        | 0.88        |
| Run 5          | Malignant | 1.00      | 0.88   | 0.94     | 0.88        | 1.00        |
|                | Benign    | 0.88      | 1.00   | 0.94     | 1.00        | 0.88        |
| Run 6          | Malignant | 0.87      | 0.93   | 0.9      | 0.93        | 0.88        |
|                | Benign    | 0.94      | 0.88   | 0.91     | 0.88        | 0.93        |
| Run 7          | Malignant | 0.97      | 0.93   | 0.95     | 0.93        | 0.97        |
|                | Benign    | 0.94      | 0.97   | 0.95     | 0.97        | 0.93        |
| Run 8          | Malignant | 0.82      | 0.96   | 0.89     | 0.96        | 0.82        |
|                | Benign    | 0.97      | 0.82   | 0.89     | 0.82        | 0.96        |
| Run 9          | Malignant | 0.97      | 0.94   | 0.96     | 0.94        | 0.96        |
|                | Benign    | 0.93      | 0.96   | 0.95     | 0.96        | 0.94        |
| Run 10         | Malignant | 0.94      | 0.97   | 0.96     | 0.97        | 0.93        |
|                | Benign    | 0.96      | 0.93   | 0.94     | 0.93        | 0.97        |
| <b>Average</b> | Malignant | 0.934     | 0.938  | 0.936    | 0.94        | 0.92        |
|                | Benign    | 0.935     | 0.921  | 0.925    | 0.91        | 0.94        |

**Table 8.** Best confusion matrix for the proposed CNN.

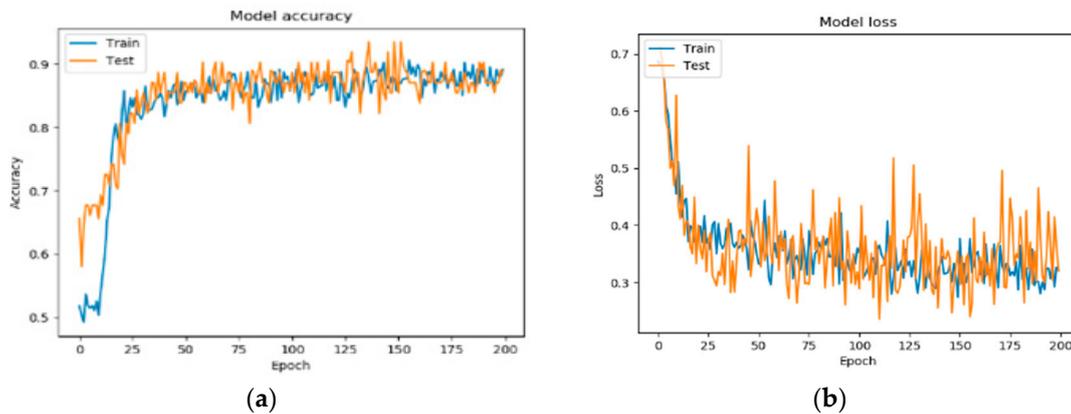
| Title 1   | Malignant | Benign |
|-----------|-----------|--------|
| Malignant | 31        | 1      |
| Benign    | 2         | 25     |

**Table 9.** CNN performance for grayscale images.

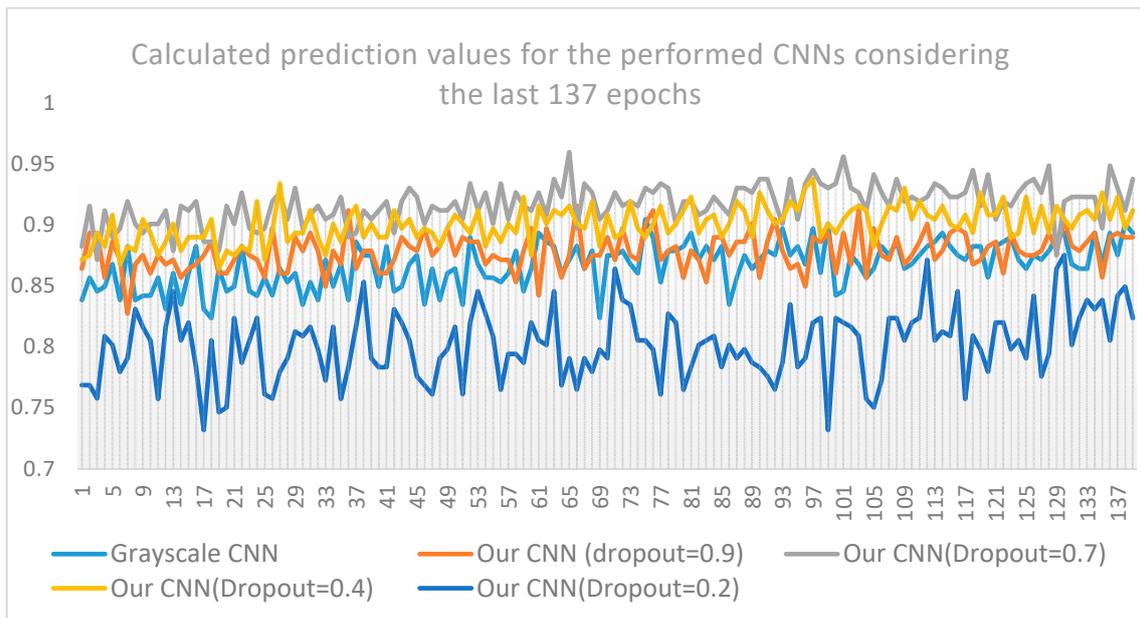
| Grayscale | CNN Architecture (2,4,8), Epochs = 200, Batch Size = 8, Dense = 16, Drop-Rate = 0.2, Pixel Size (150 × 150) |                   |                    |                |
|-----------|---|-------------------|--------------------|----------------|
| Runs      | Accuracy (Validation)   | Loss (Validation) | Accuracy (Testing) | Loss (Testing) |
| Run 1     | 85.94   | 0.31              | 85.71              | 0.42           |
| Run 2     | 92.19   | 0.28              | 80.36              | 0.45           |
| Run 3     | 78.12   | 0.52              | 75                 | 0.51           |
| Run 4     | 87.5  | 0.26              | 78.57              | 0.61           |
| Run 5     | 84.37   | 0.33              | 89.28              | 0.3            |
| Average   | 85.62   | 0.34              | 81.78              | 0.46           |

**Table 10.** Performance metrics and confusion matrix for grayscale images.

| Grayscale               | Precision        | Recall        | F1-score | Sensitivity | Specificity |
|-------------------------|------------------|---------------|----------|-------------|-------------|
| Malignant               | 0.82             | 0.97          | 0.89     | 0.82        | 0.96        |
| Benign                  | 0.96             | 0.76          | 0.85     | 0.96        | 0.82        |
| <b>Confusion matrix</b> | <b>Malignant</b> | <b>Benign</b> |          |             |             |
| Malignant               | 32               | 1             |          |             |             |
| Benign                  | 7                | 22            |          |             |             |



**Figure 6.** Precision curves for best CNN in grayscale mode: (a) accuracy and (b) loss.



**Figure 7.** Average prediction errors for the best CNN architecture.

**4.4. Comparison with Benchmark CNNs**

In what follows, an extensive comparative analysis between the state-of-the-art CNNs such as VGG16 [33], ResNet50 [34], Mobile Net [92] and Densenet [35] and our best model was performed. The following well-known CNN architectures were used: (i) ResNet50, is a 50-weight layer deep version of ResNet (Residual neural Network), with 152 layers, based on “network-in-network” micro-architectures [94]. ResNet has less parameters than the VGG network, demonstrating that extremely deep networks can be trained using standard SGD (Stochastic gradient descent) and a reasonable initialization function through the use of residual modules. (ii) VGG16 [33], is an extended

version of VGG as it contains 16 weight layers within the architecture. VGGs are usually constructed by using  $3 \times 3$  convolutional layers which are stacked on top of each other.

DenseNET is an extension of ResNet [95]. Based on recent research, it has been proven that using Neural Networks which have smaller connections between layers, mainly in the case where the dataset is relatively small, give better classification performance. Based on this result, DenseNet was created. It follows the feed-forward technique and each layer has as the input all previous output feature maps [35].

MobileNet convolutes each channel separately instead of combining and flattening them all, with the use of depthwise separable convolutions [92]. Its architecture combines convolutional layers, depthwise and pointwise layers to a total number of 30.

These popular CNNs were used for transfer learning and were used with weights that were trained on the ImageNet [93] dataset. Tables 11–13 gather the results of these benchmark CNN models compared with our best CNN configuration.

**Table 11.** Accuracy and loss of the proposed CNN model and benchmark CNNs.

| Average         | Best CNN | ResNet50 | VGG16 | MobileNet | DenseNet |
|-----------------|----------|----------|-------|-----------|----------|
| Acc. Validation | 92.50    | 91.87    | 82.49 | 88.13     | 93.75    |
| Loss Validation | 0.27     | 0.24     | 0.37  | 0.25      | 0.15     |
| Acc. Testing    | 92.50    | 91.87    | 83.75 | 85.36     | 95       |
| Loss Testing    | 0.25     | 0.19     | 0.34  | 0.35      | 0.176    |

**Table 12.** Malignant disease class performance comparison of different methods.

|             | Best CNN | ResNet50 | VGG16 | MobileNet | DenseNet |
|-------------|----------|----------|-------|-----------|----------|
| Accuracy    | 92.50    | 91.87    | 83.75 | 85.36     | 95       |
| Precision   | 0.93     | 0.84     | 0.85  | 0.78      | 0.92     |
| Recall      | 0.94     | 1        | 0.9   | 0.99      | 0.99     |
| F1-score    | 0.94     | 0.91     | 0.86  | 0.87      | 0.95     |
| Sensitivity | 0.94     | 1        | 0.89  | 0.99      | 0.99     |
| Specificity | 0.92     | 0.78     | 0.75  | 0.66      | 0.9      |

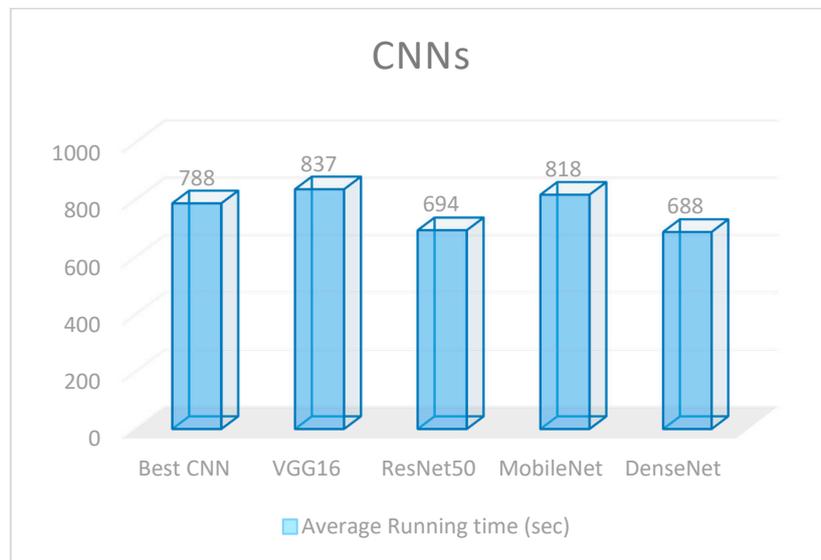
**Table 13.** Benign class performance comparison of different methods.

|             | Best CNN | ResNet50 | VGG16 | MobileNet | DenseNet |
|-------------|----------|----------|-------|-----------|----------|
| Accuracy    | 92.50    | 91.87    | 83.75 | 85.36     | 95       |
| Precision   | 0.94     | 1        | 0.87  | 0.99      | 0.98     |
| Recall      | 0.92     | 0.78     | 0.75  | 0.66      | 0.9      |
| F1-score    | 0.93     | 0.87     | 0.77  | 0.79      | 0.94     |
| Sensitivity | 0.91     | 0.78     | 0.75  | 0.66      | 0.91     |
| Specificity | 0.94     | 1        | 0.89  | 0.99      | 0.99     |

In this research work, after an extensive exploration with the provided architectures of popular CNNs, the following parameters for the well-known CNNs were defined.

- Best CNN: batch-size = 8, dropout = 0.7, nodes (3 conv layers) = 8,16,32, dense nodes: 64, epochs:200, pixel = (256,256,3)
- VGG16: pixel size (200 × 200), batch size = 32, dropout = 0.2, dense nodes 2 × 512, epochs = 200
- ResNet50: pixel size (224 × 224), batch size = 32, dropout = 0.5, dense nodes 32 × 32, epochs = 200
- MobileNet: pixel size (250 × 250), batch size = 8, dropout = 0.5, global average pooling, dense = 1500 × 1500, epochs = 200
- DenseNet: pixel size (200X200), batch size = 8, dropout = 0.9, global average pooling, dense nodes = (128 × 128), epochs = 200.

In Appendix B, Figures A1–A4 illustrate the average prediction accuracies with their respective loss curves for the optimal architectures of the four benchmark CNNs. The average running time for each CNN architecture for all models was calculated and Figure 8 depicts the values of running time for 200 epochs. It is obvious that the proposed CNN model has similar average run time to the other popular CNN models.



**Figure 8.** Average run time for the compared best CNN architectures.

## 5. Discussion

Through the trials and experiments performed, it emerges that the accuracy of the selected CNN architecture is higher in the case of RGB images. Also, the loss (validation and testing) is quite low, and the confusion matrix is better in terms of accuracy than the corresponding one in the case of grayscale images, showing the superiority of RGB images use for the classification task. The CNN in RGB process spent 4 min approximately for training each fold dataset and less than 30 s for prediction.

The best performing CNN algorithm was then compared with some state-of-the-art CNNs that are commonly used for image classification problems. To directly compare the classification performance of our model with other popular CNN architectures we used, besides the accuracy, the evaluation metrics of precision, recall, F1 score, sensitivity and specificity indicators, which are listed in the tables above. We highlight that our proposed algorithm performed better when compared to popular CNN algorithms for the specific problem. This shows that the improvements made to network architecture and the configuration of hyperparameter were rather effective. This study validates the premise that CNNs are algorithms that can offer high accuracy in medical image classification-based problems. This can have a direct application on medical imaging where the automatic identification of diseases is crucial for the patients.

The main outcomes of this study can be summarized as follows:

- The proposed CNN method exhibits outstanding performance when RGB analysis is performed, for the examined images of this case study. This can emanate from the fact that the results produced from the application of the CNN architecture are based on distinct features that appear specifically on the bone metastasis presence scans, compared to the healthy scans (no malignant spots).
- The proposed CNN architecture is superior compared to three out of four benchmarks and well-known CNN architectures (ResNet50, VGG16, MobileNet, and DenseNet) which have been previously and efficiently used in medical image-processing problems. As observed from the aforementioned tables and figures, the results deriving from the application of the CNN method,

are better in terms of classification accuracy, prediction, sensitivity, specificity and f1 score, than those coming from popular and well-known CNN approaches found in the relevant literature. Even if CNN has an inferior performance compared to DenseNet, the accuracy and loss values of the proposed CNN show a quite low variability whereas the corresponding DenseNet values present substantial variations among predicted outcomes (see Figure A3).

- The proposed bone scan deep learning performs efficiently despite the fact that it was trained on a small number of images.
- Overall, the deployed process seemingly improves the diagnostic effect of the deep-learning method, making it more efficient compared to other benchmark CNN architectures for image analysis. For the purpose of medical image analysis and classification, the CNN approach can be used effectively in the classification of whole-body images in nuclear medicine, outweighing the popular CNN architectures for medical image analysis.

To sum up, this is the first research study concerning the application of CNNs efficiently in bone metastasis diagnosis from breast cancer patients in bone scintigraphy. Following the proposed methodology, a robust CNN algorithm is created for bone metastasis diagnosis using whole-body scans from breast cancer patients. This algorithm can be implemented in any system without specific requirements, providing the nuclear medicine physician with an efficient and robust decision support tool to automatically identify whether a patient is suffering from breast cancer or not, by looking at whole-body scintigraphy images. Due to its high classification accuracy and performance, this algorithm has a significant positive impact in clinical diagnosis and decision making in nuclear medicine by providing diagnosis to metastatic breast cancer patients.

The proposed classification algorithm can be applied in any type of whole-body scans in bone scintigraphy and can be integrated into a CAD system of whole-body diagnosis or a clinical decision support system in nuclear medical imaging. Hence, it can be used by physicians in medical practice.

## 6. Conclusions

In this research study, convolutional neural network algorithms have been used to deal with the problem of bone metastasis diagnosis in breast cancer patients due to their significant applicability in image analysis. A total of 408 images were acquired with similar numbers of both benign and malignant cases (metastasis present). Two different types of processing images, RGB-mode and grayscale-mode, denoting their colored or monochromatic tone, were tested in order to evaluate their contribution to the increase of performance. Several CNN architectures were tested, and only one ended up with the best performance under all hyperparameter selection cases, with the results ranging from 85.35% to 93.75%.

In conclusion, these preliminary image data suggest that bone scintigraphy combined with the CNN method can have a considerable effect in the detection of bone metastasis. In particular, this approach allows an easier and more precise interpretation of the images that can have a positive impact on diagnosis accuracy as well as on decision making regarding the treatment that will be further administered.

Due to an insufficient amount of data in our study, a more detailed classification and diagnosis of breast cancer diseases needs to be carried out, and that is going to be our next priority in our future research. Since deep neural networks are based on complex, inter-connected hierarchical representations of the training data, which are used to make predictions, the task of interpreting these predictions becomes quite demanding. Moreover, due to the resemblance between deep neural networks and “blackbox” models, interpretability needs to be enhanced, while the research community should further investigate how to measure sensitivity and visualize features in deep learning.

In future work, the authors intend to show how the proposed algorithm could be integrated into a CAD system of whole-body diagnosis in nuclear medicine and also provide evidence that the clinical decision support system could be generalized across continents.

**Author Contributions:** Conceptualization, N.P. and E.P.; methodology, N.P., E.P. and A.A.; software, A.F. and A.A.; validation, N.P., E.P. and A.A.; formal analysis, N.P.; investigation, N.P.; resources, N.P.; data curation, N.P.; writing—original draft preparation, N.P.; writing—review and editing, N.P., and E.P.; visualization, A.F. and A.A.; supervision, N.P. and E.P.; project administration, E.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** CNNs (epochs = 200, dropout = 0.7, conv (8->16->32), dense = 64) runs for different pixel sizes.

| Batch size<br>= 8 | CNN (100 × 100 × 3) |             |              |             | CNN (200 × 200 × 3) |             |              |             | CNN (300 × 300 × 3) |             |              |             |
|-------------------|---------------------|-------------|--------------|-------------|---------------------|-------------|--------------|-------------|---------------------|-------------|--------------|-------------|
|                   | Acc. Val            | Loss Val    | Acc Test     | Loss Test   | Acc. Val            | Loss Val    | Acc Test     | Loss Test   | Acc. Val            | Loss Val    | Acc Test     | Loss Test   |
| Run 1             | 82.81               | 0.76        | 73.21        | 1.04        | 85.94               | 0.17        | 92.86        | 0.17        | 87.5                | 0.296       | 96.428       | 0.157       |
| Run 2             | 82.81               | 0.76        | 87.5         | 0.3         | 92.19               | 0.2         | 91.1         | 0.29        | 89.06               | 0.247       | 94.64        | 0.167       |
| Run 3             | 81.25               | 0.39        | 80.36        | 0.64        | 81.25               | 0.28        | 87.5         | 0.24        | 87.5                | 0.27        | 89.28        | 0.206       |
| Run 4             | 84.37               | 0.37        | 87.5         | 0.45        | 84.37               | 0.25        | 83.93        | 0.27        | 92.187              | 0.21        | 89.28        | 0.255       |
| Run 5             | 81.25               | 0.39        | 76.78        | 0.54        | 92.19               | 0.2         | 92.86        | 0.15        | 87.5                | 0.276       | 89.28        | 0.235       |
| Average           | <b>82.5</b>         | <b>0.53</b> | <b>81.07</b> | <b>0.59</b> | <b>87.19</b>        | <b>0.22</b> | <b>89.65</b> | <b>0.22</b> | <b>88.75</b>        | <b>0.26</b> | <b>91.78</b> | <b>0.20</b> |

**Table A2.** CNN Model (epochs = 200, pixel = 200 × 200, conv (8->16->32), dense = 64) runs for different dropouts.

| Batch size<br>= 8 | Dropout = 0.2 |          |          |           | Dropout = 0.4 |          |          |           | Dropout = 0.9 |          |          |           |
|-------------------|---------------|----------|----------|-----------|---------------|----------|----------|-----------|---------------|----------|----------|-----------|
|                   | Acc. Val      | Loss Val | Acc Test | Loss Test | Acc. Val      | Loss Val | Acc Test | Loss Test | Acc. Val      | Loss Val | Acc Test | Loss Test |
| Run 1             | 93.75         | 0.22     | 91.1     | 0.21      | 87.5          | 0.44     | 92.86    | 0.18      | 71.87         | 0.55     | 71.43    | 0.57      |
| Run 2             | 84.37         | 0.45     | 83.93    | 0.29      | 89.1          | 0.47     | 85.71    | 0.34      | 81.25         | 0.41     | 87.5     | 0.37      |
| Run 3             | 83.37         | 0.38     | 85.67    | 0.28      | 87.5          | 0.21     | 89.28    | 0.17      | 92.19         | 0.36     | 85.71    | 0.41      |
| Run 4             | 95.31         | 0.7      | 89.29    | 0.17      | 87.5          | 0.25     | 85.71    | 0.4       | 79.69         | 0.43     | 87.5     | 0.46      |
| Run 5             | 90.62         | 0.21     | 92.86    | 0.23      | 82.1          | 0.36     | 87.25    | 0.43      | 84.37         | 0.4      | 78.57    | 0.43      |
| Average           | 89.48         | 0.39     | 85.57    | 0.24      | 86.74         | 0.34     | 88.16    | 0.3       | 81.87         | 0.43     | 82.14    | 0.45      |

**Table A3.** VGG16 (epochs = 200, dropout = 0.2, pixel = 256 × 256) runs for different batch sizes.

| VGG16 (Batch Size=32) |          |          |          |           |
|-----------------------|----------|----------|----------|-----------|
|                       | Acc. Val | Loss Val | Acc Test | Loss Test |
| Run 1                 | 84.37    | 0.3      | 81.25    | 0.36      |
| Run 2                 | 73.44    | 0.53     | 78.12    | 0.26      |
| Run 3                 | 75       | 0.46     | 84.37    | 0.38      |
| Run 4                 | 89.06    | 0.31     | 90.62    | 0.22      |
| Run 5                 | 90.62    | 0.26     | 84.37    | 0.51      |
| Average               | 82.49    | 0.37     | 83.74    | 0.34      |

Appendix B

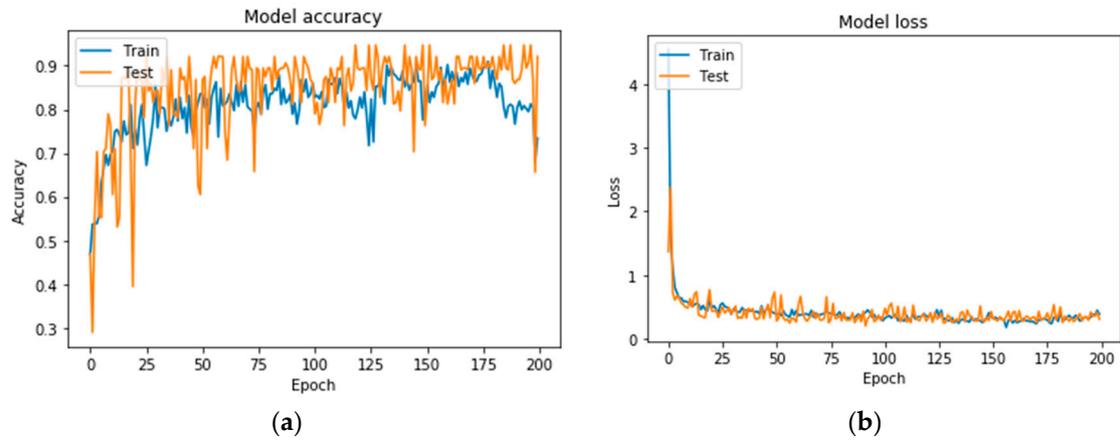


Figure A1. Precision curves for VGG16: (a) accuracy and (b) loss.

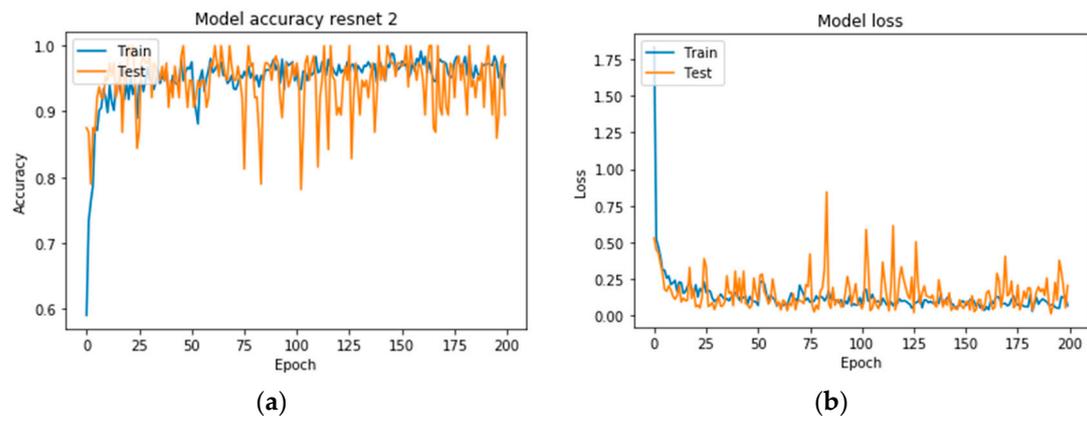


Figure A2. Precision curves for Resnet50 (a) accuracy and (b) loss.

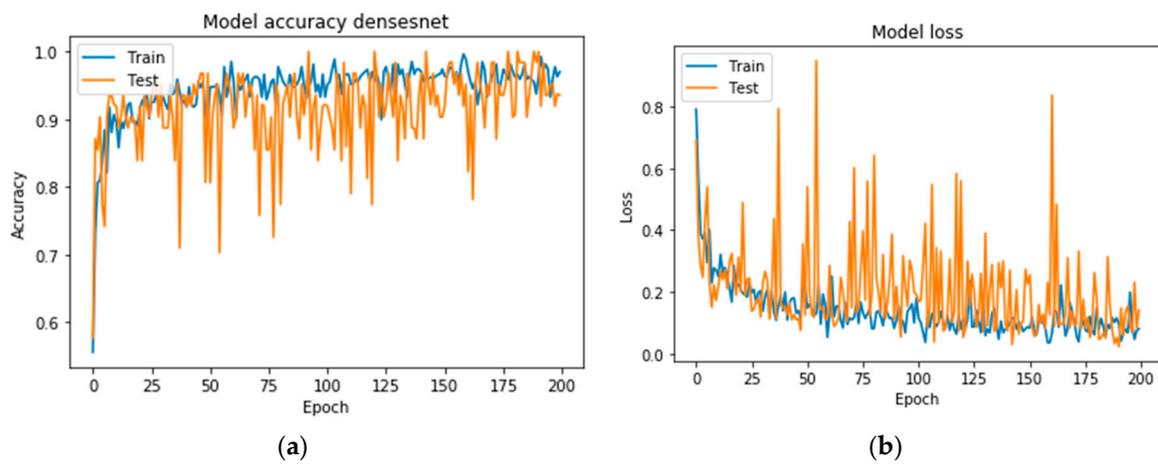
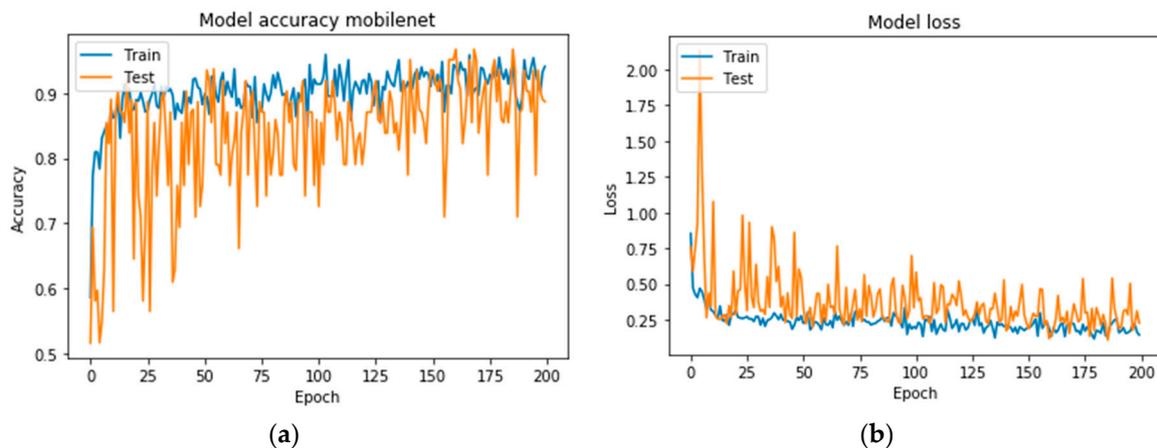


Figure A3. Precision curves for DenseNet (a) accuracy and (b) loss.



**Figure A4.** Precision curves for MobileNet (a) accuracy and (b) loss.

## References

- Roodman, G.D. Mechanisms of bone metastasis. *N. Engl. J. Med.* **2004**, *350*, 1655–1664. [[CrossRef](#)] [[PubMed](#)]
- Coleman, R.E. Metastatic bone disease: Clinical features, pathophysiology and treatment strategies. *Cancer Treat. Rev.* **2001**, *27*, 165–176. [[CrossRef](#)] [[PubMed](#)]
- Macedo, F.; Ladeira, K.; Pinho, F.; Saraiva, N.; Bonito, N.; Pinto, L.; Goncalves, F. Bone Metastases: An Overview. *Oncol. Rev.* **2017**, *11*, 321. [[CrossRef](#)] [[PubMed](#)]
- Battafarano, G.; Rossi, M.; Marampon, F.; Del Fattore, A. Management of bone metastases in cancer: A review. *Crit. Rev. Oncol. Hematol. Int. J. Mol. Sci.* **2005**. [[CrossRef](#)]
- Manders, K.; van de Poll-Franse, L.V.; Creemers, G.J.; Vreugdenhil, G.; van der Sangen, M.J.; Nieuwenhuijzen, G.A.; Roumen, R.M.; Voogd, A.C. Clinical management of women with metastatic breast cancer: A descriptive study according to age group. *BMC Cancer* **2006**, *6*, 179. [[CrossRef](#)] [[PubMed](#)]
- Yazdani, A.; Dorri, S.; Atashi, A.; Shirafkan, H.; Zabolinezhad, H. Bone Metastasis Prognostic Factors in Breast Cancer. *Breast Cancer* **2019**, *13*. [[CrossRef](#)]
- Coleman, R.E. Clinical features of metastatic bone disease and risk of skeletal morbidity. *Clin. Cancer Res.* **2006**, *12*, 6243s–6249s. [[CrossRef](#)]
- Muhammad, A.; Mehwish, I.; Muhammad, D.; Khan, A.U. Awareness and current knowledge of breast cancer. *Biol. Res.* **2017**, *50*, 33. [[CrossRef](#)]
- Talbot, J.N.; Paycha, F.; Balogova, S. Diagnosis of bone metastasis: Recent comparative studies of imaging modalities. *Q. J. Nucl. Med. Mol. Imaging* **2011**, *55*, 374–410.
- O’Sullivan, G.J.; Carty, F.L.; Cronin, C.G. Imaging of bone metastasis: An update. *World J. Radiol.* **2015**, *7*, 202–211. [[CrossRef](#)]
- Chang, C.Y.; Gill, C.M.; Simeone, J.F.; Taneja, A.K.; Huang, A.J.; Torriani, M.; Bredella, M.A. 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. [[CrossRef](#)] [[PubMed](#)]
- Savelli, G.; Maffioli, L.; Maccauro, M.; De Deckere, E.; Bombardieri, E. Bone scintigraphy and the added value of SPECT (single photon emission tomography) in detecting skeletal lesions. *Q. J. Nucl. Med.* **2001**, *45*, 27–37. [[PubMed](#)]
- Ghosh, P. The role of SPECT/CT in skeletal malignancies. *Semin. Musculoskelet. Radiol.* **2014**, *18*, 175–193. [[CrossRef](#)]
- Rieden, K. Conventional imaging and computerized tomography in diagnosis of skeletal metastases. *Radiologe* **1995**, *35*, 15–20.
- Hamaoka, T.; Madewell, J.E.; Podoloff, D.A.; Hortobagyi, G.N.; Ueno, N.T. Review—Bone imaging in metastatic breast cancer. *J. Clin. Oncol.* **2004**, *22*, 2942–2953. [[CrossRef](#)]
- Wyngaert, T.; Strobel, K.; Kampen, W.U.; Kuwert, T.; Bruggen, W.; Mohan, H.K.; Gnanasegaran, G.; Delgado-Bolton, R.; Weber, W.A.; Beheshti, M.; et al. The EANM practice guidelines for bone scintigraphy. *Eur. J. Nucl. Med. Mol. Imaging* **2016**, *43*, 1723–1738. [[CrossRef](#)]

17. Even-Sapir, E.; Metser, U.; Mishani, E.; Lievshitz, G.; Lerman, H.; Leibovitch, I. The detection of bone metastases in patients with high-risk prostate cancer: 99mTc-MDP Planar bone scintigraphy, single- and multi-field-of-view SPECT, 18F-fluoride PET, and 18F-fluoride PET/CT. *J. Nucl. Med.* **2006**, *47*, 287–297.
18. Hahn, S.; Heusner, T.; Kümmel, S.; Königer, A.; Nagarajah, J.; Müller, S.; Boy, C.; Forsting, M.; Bockisch, A.; Antoch, G. Comparison of FDG-PET/CT and bone scintigraphy for detection of bone metastases in breast cancer. *Acta Radiol.* **2011**, *52*, 1009–1014. [[CrossRef](#)]
19. Drakopoulos, M.; Vejdani, K.; Osman, M. Comparison of diagnostic certainty and accuracy of 18F-NaF PET/CT and planar 99mTc-MDP bone scan in patients with prostate cancer. *J. Nucl. Med.* **2014**, *55*, 1669.
20. Nemoto, M.; Masutani, Y.; Nomura, Y.; Hanaoka, S.; Miki, S.; Yoshikawa, T.; Hayashi, N.; Ootomo, K. Machine Learning for Computer-aided Diagnosis. *Igaku Butsuri* **2016**, *36*, 29–34. [[CrossRef](#)]
21. Suzuki, K. *Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis: 9781466600591: Medicine & Healthcare Books*; University of Chicago: Chicago, IL, USA, 2012.
22. Litjens, G.; Kooi, T.; Bejnordi, B.; Setio, A.; Ciompi, F.; Ghafoorian, M.; Laak, J.; Ginneken, B.; Sánchez, C. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)]
23. Biswas, M.; Kuppili, V.; Saba, L.; Edla, D.R.; Suri, H.S.; Cuadrado-Godia, E.; Laird, J.R.; Marinho, R.T.; Sanches, J.M.; Nicolaides, A.; et al. State-of-the-art review on deep learning in medical imaging. *Front. Biosci.* **2019**, *24*, 392–426.
24. Shen, D.; Wu, G.; Suk, H.I. Deep Learning in Medical Image Analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [[CrossRef](#)] [[PubMed](#)]
25. Sahiner, B.; Pezeshk, A.; Hadjiiski, L.M.; Wang, X.; Drukker, K.; Cha, K.H.; Summers, R.M.; Giger, M.L. Deep learning in medical imaging and radiation therapy. *Med. Phys.* **2019**, *46*. [[CrossRef](#)] [[PubMed](#)]
26. Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik* **2019**, *29*, 102–127. [[CrossRef](#)] [[PubMed](#)]
27. Yang, L.; Xie, X.; Li, P.; Zhang, D.; Zhang, L. Part-based convolutional neural network for visual recognition. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017.
28. Komeda, Y.; Handa, H.; Watanabe, T.; Nomura, T.; Kitahashi, M.; Sakurai, T.; Okamoto, A.; Minami, T.; Kono, M.; Arizumi, T.; et al. Computer-Aided Diagnosis Based on Convolutional Neural Network System for Colorectal Polyp Classification: Preliminary Experience. *Oncology* **2017**, *93* (Suppl. S1), 30–34. [[CrossRef](#)]
29. Anwar, S.M.; Majid, M.; Qayyum, A.; Awais, M.; Alnowami, M.; Khan, M.K. Medical Image Analysis using Convolutional Neural Networks: A Review. *J. Med. Syst.* **2018**, *42*, 226. [[CrossRef](#)]
30. Gao, J.; Jiang, Q.; Zhou, B.; Chen, D. Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview. *Math. Biosci. Eng.* **2019**, *16*, 6536–6561. [[CrossRef](#)]
31. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2014; pp. 818–833.
32. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. *ImageNet Classification with Deep Convolutional Neural Networks*; Advances in Neural Information Processing Systems 25; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Dutchess County, NY, USA, 2012; pp. 1097–1105.
33. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
35. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, Hawaii, 21–26 July 2016.
36. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv* **2016**, arXiv:1606.00915. [[CrossRef](#)]
37. Nahid, A.; Kong, Y. Involvement of Machine Learning for Breast Cancer Image Classification: A Survey. *Comput. Math. Methods Med.* **2017**, *2017*, 29. [[CrossRef](#)]
38. Cheng, H.; Shi, X.; Min, R.; Hu, L.; Cai, X.; Du, H. Approaches for automated detection and classification of masses in mammograms. *Pattern Recognit.* **2006**, *39*, 646–668. [[CrossRef](#)]

39. Ponraj, D.N.; Jenifer, M.E.; Poongodi, D.P.; Manoharan, J.S. A survey on the preprocessing techniques of mammogram for the detection of breast cancer. *J. Emerg. Trends Comput. Inf. Sci.* **2011**, *2*, 656–664.
40. Jiang, Y.; Chen, L.; Zhang, H.; Xiao, X. Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module. *PLoS ONE.* **2019**, *14*, e0214587. [[CrossRef](#)]
41. Sert, E.; Ertekin, S.; Halici, U. Ensemble of convolutional neural networks for classification of breast microcalcification from mammograms. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2017**, *2017*, 689–692.
42. Rangayyan, R.M.; Ayres, F.J.; Desautels, J.L. A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. *J. Frankl. Inst.* **2007**, *344*, 312–348. [[CrossRef](#)]
43. Magna, G.; Casti, P.; Jayaraman, S.V.; Salmeri, M.; Mencattini, A.; Martinelli, E.; Natale, C.D. Identification of mammography anomalies for breast cancer detection by an ensemble of classification models based on artificial immune system. *Knowl. Based Syst.* **2016**, *101*, 60–70. [[CrossRef](#)]
44. Yassin, N.I.R.; Omran, S.; El Houby, E.M.F.; Allam, H. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Comput. Methods Programs Biomed.* **2018**, *156*, 25–45. [[CrossRef](#)]
45. Gardezi, S.J.S.; Elazab, A.; Lei, B.; Wang, T. Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review. *J. Med. Internet Res.* **2019**, *21*, e14464. [[CrossRef](#)]
46. Munir, K.; Elahi, H.; Ayub, A.; Frezza, F.; Rizzi, A. Cancer Diagnosis Using Deep Learning: A Bibliographic Review. *Cancers* **2019**, *11*, 1235. [[CrossRef](#)]
47. Chougrad, H.; Zouaki, H.; Alheyane, O. Deep Convolutional Neural Networks for breast cancer screening. *Comput. Methods Programs Biomed.* **2018**, *157*, 19–30. [[CrossRef](#)]
48. Abdelhafiz, D.; Yang, C.; Ammar, R.; Nabavi, S. Deep convolutional neural networks for mammography: Advances, challenges and applications. *BMC Bioinform.* **2019**, *20* (Suppl. S11), 481. [[CrossRef](#)]
49. CNNs Applied in Breast Cancer Classification. Available online: <https://towardsdatascience.com/convolutional-neural-network-for-breast-cancer-classification-52f1213dcc9> (accessed on 10 November 2019).
50. Kumar, K.; Chandra Sekhara Rao, A. Breast cancer classification of image using convolutional neural network. In Proceedings of the 2018 4th International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, India, 15–17 March 2018; Available online: <https://ieeexplore.ieee.org/abstract/document/8389034> (accessed on 6 January 2020).
51. Suzuki, S.; Zhang, X.; Homma, N.; Ichiji, K.; Sugita, N.; Kawasumi, Y.; Ishibashi, T.; Yoshizawa, M. Mass detection using deep convolutional neural networks for mammographic computer-aided diagnosis. In Proceedings of the 55th Annual Conference of the Society of Instruments and Control Engineers of Japan (SICE), Tsukuba, Japan, 20–23 September 2016; pp. 1382–1386.
52. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. Breast cancer histopathological image classification using convolutional neural networks. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 2560–2567.
53. Wichakam, I.; Vateekul, P. Combining deep convolutional networks and SVMs for mass detection on digital mammograms. In Proceedings of the 8th International Conference on Knowledge and Smart Technology (KST), Bangkok, Thailand, 3–6 February 2016; pp. 239–244.
54. Swiderski, B.; Kurek, J.; Osowski, S.; Kruk, M.; Barhoumi, W. Deep learning and non-negative matrix factorization in recognition of mammograms. In Proceedings of the Eighth International Conference on Graphic and Image Processing, International Society of Optics and Photonics, Tokyo, Japan, 8 February 2017; Volume 10225, p. 102250B.
55. Kallenberg, M.; Petersen, K.; Nielsen, M.; Ng, A.Y.; Diao, P.; Igel, C.; Vachon, C.M.; Holland, K.; Winkel, R.R.; Karssemeijer, N.; et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE Trans. Med. Imaging* **2016**, *35*, 1322–1331. [[CrossRef](#)]
56. Giger, M.L.; Vybomy, C.L.; Huo, Z.; Kupinski, M.A. Computer-aided diagnosis in mammography. In *Handbook of Medical Imaging*, 2nd ed.; Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review; SPIE Digital Library: Cardiff, Wales, 2000; pp. 915–1004.
57. Fenton, J.J.; Taplin, S.H.; Carney, P.A.; Abraham, L.; Sickles, E.A.; D’Orsi, C.; Berns, E.A.; Cutter, G.; Hendrick, R.E.; Barlow, W.E.; et al. Influence of computer-aided detection on performance of screening mammography. *N. Engl. J. Med.* **2017**, *356*, 1399–1409. [[CrossRef](#)]

58. Zhou, L.Q.; Wu, X.L.; Huang, S.Y.; Wu, G.G.; Ye, H.R.; Wei, Q.; Bao, L.Y.; Deng, Y.B.; Li, X.R.; Cui, X.W.; et al. Lymph Node Metastasis Prediction from Primary Breast Cancer US Images Using Deep Learning. *Radiology* **2020**, *294*, 19–28. [CrossRef]
59. Steiner, D.F.; MacDonald, R.; Liu, Y.; Truszkowski, P.; Hipp, J.D.; Gammage, C.; Thng, F.; Peng, L.; Stumpe, M.C. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. *Am. J. Surg. Pathol.* **2018**, *42*, 1636–1646. [CrossRef]
60. Takayoshi, U.; Sachiko, Y.; Seigo, Y.; Takeshi, A.; Naoki, M.; Masahiro, E.; Hiroyoshi, F.; Yoshihiro, U.; An Junichiro, W. Comparison of FDG PET and SPECT for Detection of Bone Metastases in Breast Cancer. *Breast Imaging Am. J. Roentgenol. Diagn. Adv. Search* **2005**, *184*, 1266–1273.
61. Soyeon, P.; Yoon, J.K.; Lee, S.J.; Kang, S.Y.; Yim, H.; An, Y.S. Prognostic utility of FDG PET/CT and bone scintigraphy in breast cancer patients with bone-only metastasis. *Medicine* **2017**, *96*, e8985.
62. Łukaszewski, B.; Nazar, J.; Goch, M.; Łukaszewska, M.; Stępiński, A.; Jurczyk, M.U. Diagnostic methods for detection of bone metastases. *Contemp. Oncol.* **2017**, *21*, 98–103. [CrossRef]
63. Aslantas, A.; Dandil, E.; Sağlam, S.; Çakiroğlu, M. CADBOSS: A computer-aided diagnosis system for whole-body bone scintigraphy scans. *J. Can. Res. Ther.* **2016**, *12*, 787–792. [CrossRef]
64. Sadik, M. Computer-Assisted Diagnosis for the Interpretation of Bone Scintigraphy: A New Approach to Improve Diagnostic Accuracy. Ph.D. Thesis, University of Gothenburg, Gothenburg, Sweden, June 2019.
65. Fogelman, I.; Cook, G.; Israel, O.; Van der Wall, H. Positron emission tomography and bone metastases. *Semin. Nucl. Med.* **2005**, *35*, 135–142. [CrossRef]
66. Pianou, N.K.; Stavrou, P.Z.; Vlontzou, E.; Rondogianni, P.; Exarhos, D.N.; Datseris, I.E. More advantages in detecting bone and soft tissue metastases from prostate cancer using 18F-PSMA PET/CT. *Hell. J. Nucl. Med.* **2019**, *22*, 6–9. [CrossRef]
67. Newberg, A. Bone Scans. In *Radiology Secrets Plus*, 3rd ed.; Elsevier: Cambridge, MA, USA, 2011; pp. 376–381. ISBN 978-0-323-06794-2. [CrossRef]
68. Dang, J. Classification in Bone Scintigraphy Images Using Convolutional Neural Networks. Master's Thesis, Lund University, Lund, Sweden, 2016.
69. Bradshaw, T.; Perk, T.; Chen, S.; Hyung-Jun, I.; Cho, S.; Perlman, S.; Jeraj, R. Deep learning for classification of benign and malignant bone lesions in [F-18]NaF PET/CT images. *J. Nucl. Med.* **2018**, *59* (Suppl. S1), 327.
70. Furuya, S.; Kawauchi, K.; Hirata, K.; Manabe, O.; Watanabe, S.; Kobayashi, K.; Ichikawa, S.; Katoh, C.; Shiga, T.J. A convolutional neural network-based system to detect malignant findings in FDG PET-CT examinations. *Nucl. Med.* **2019**, *60*, 1210.
71. Furuya, S.; Kawauchi, K.; Hirata, K.; Manabe, O.; Watanabe, S.; Kobayashi, K.; Ichikawa, S.; Katoh, S.; Shiga, T.J. Can CNN detect the location of malignant uptake on FDG PET-CT? *Nucl. Med.* **2019**, *60* (Suppl. S1), 285.
72. Kawauchi, K.; Hirata, K.; Katoh, C.; Ichikawa, S.; Manabe, O.; Kobayashi, K.; Watanabe, S.; Furuya, S.; Shiga, T. A convolutional neural network based system to prevent patient misidentification in FDG-PET examinations. *Sci. Rep.* **2019**, *9*, 7192. [CrossRef]
73. Kawauchi, K.; Hirata, K.; Ichikawa, S.; Manabe, O.; Kobayashi, K.; Watanabe, S.; Haseyama, M.; Ogawa, T.; Togo, R.; Shiga, T.; et al. Strategy to develop convolutional neural network-based classifier for diagnosis of whole-body FDG PET images. *Nucl. Med.* **2018**, *59* (Suppl. S1), 326.
74. Gjertsson, K. Segmentation in Skeletal Scintigraphy Images Using CNNs. Master's Thesis, Lund University, Lund, Sweden, 2017.
75. Weiner, M.G.; Jenicke, L.; Miller, V.; Bohuslavizki, H.K. Artifacts and nonosseous uptake in bone scintigraphy. Imaging reports of 20 cases. *Radiol. Oncol.* **2001**, *35*, 185–191.
76. O'Shea, K.T.; Nash, R. An Introduction to Convolutional Neural Networks. 2015. Available online: <https://arxiv.org/abs/1511.08458> (accessed on 12 November 2019).
77. Albelwi, S.; Mahmood, A. A Framework for Designing the Architectures of Deep Convolutional Neural Networks. *Entropy* **2017**, *19*, 242. [CrossRef]
78. Springenberg, J.T.; Dosovitskiy, A.; Brox, T.; Riedmiller, M. Striving for simplicity: The all convolutional net, Proceedings of ICLR-2015. *arXiv* **2014**, arXiv:1412.6806.
79. Yamashita, R.; Nishio, M.; Kinh, R.; Togashi, G. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef]

80. Fully Connected Layers in Convolutional Neural Networks: The Complete Guide. Available online: <https://missinglink.ai/guides/convolutional-neural-networks/fully-connected-layers-convolutional-neural-networks-complete-guide/> (accessed on 5 January 2020).
81. Jayalakshmi, T.; Santhakumaran, A. Statistical normalization and back propagation for classification. *Int. J. Comput. Theory Eng.* **2011**, *3*, 89–93. [CrossRef]
82. Bishop, C.M.; Hart, P.E.; Stork, D.G. *Pattern Recognition and Machine Learning*; Springer: Berlin, Germany, 2006.
83. Moustakidis, S.; Christodoulo, E.; Papageorgiou, E.; Kokkotis, C.; Papandrianos, N.; Tsaopoulos, D. Application of machine intelligence for osteoarthritis classification: A classical implementation and a quantum perspective. *Quantum Mach. Intell.* **2019**. [CrossRef]
84. Theodoridis, S.; Koutroumbas, K.; Stork, D.G. *Pattern Recognition*; Academic Press: Cambridge, MA, USA, 2009.
85. Labatut, V.; Cherifi, H. Accuracy measures for the comparison of classifiers. In Proceedings of the 5th International Conference on Information Technology, Amman, Jordan, 11–13 May 2011.
86. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
87. Lofte, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
88. Flatten Layer What Does. Available online: <https://www.google.com/search?q=flatten+layer+what+does&aq=flatten+layer+what+does&aqs=chrome.69i57j0l3.5821j0j7&sourceid=chrome&ie=UTF-8> (accessed on 3 January 2020).
89. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
90. Chollet, F. “Keras.” GitHub Repository. 2015. Available online: <https://github.com/fchollet/keras> (accessed on 9 December 2019).
91. Google Colab, Colaboratory Cloud Environment Supported by Google. Available online: <https://colab.research.google.com/> (accessed on 25 January 2020).
92. Jia, D.; Wei, D.; Richard, S.; Li-Jia, L.; Kai, L.; Li, F.-F. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009.
93. Lin, M.; Chen, Q.; Yan, S. Network in Network. Proceedings in ICLR 2013. *arXiv* **2013**, arXiv:1312.4400.
94. Densely Connected Convolutional Networks. Available online: <https://arthurdouillard.com/post/densenet/> (accessed on 3 January 2020).
95. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, J.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis. (IJCV)* **2015**, *115*, 211–252. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).