Bulletin of Electrical Engineering and Informatics
Vol. 13, No. 3, June 2024, pp. 1802~1813
ISSN: 2302-9285, DOI: 10.11591/eei.v13i3.6861

A deep learning-based system for accurate diagnosis of pelvic bone tumors

Mona Shouman[1], Kamel Hussein Rahouma[1], Hesham Fathy Aly Hamed[2,3]
[1]Department of Computer and System Engineering, Faculty of Engineering, Minia University, Minia, Egypt
[2]Department of Electrical and Communication Engineering, Faculty of Engineering, Minia University, Minia, Egypt
[3]Dean of College of Artificial Intelligence- Russian University, Cairo, Egypt

Article Info

Article history:

Received Jun 2, 2023
Revised Oct 2, 2023
Accepted Nov 14, 2023

Keywords:

Classification
Computed tomography
Deep learning
DenseNet
Pelvic
Pelvic bone tumors

ABSTRACT

Bone image analysis and categorizing bone cancers have both seen advancements thanks to deep learning (DL), more notably convolution neural networks (CNN). This study suggests a brand-new CNN-based methodology for categorizing pelvic bone tumors specifically. This work aims to create a pelvic bone computed tomography (CT) image categorization system based on deep learning. The proposed technique uses a convolutional neural network (CNN) architecture to automatically extract information from the CT images and classify them into distinct categories of tumors. A total of 178 3D CT pictures was discovered and added retroactively. DenseNet created the image-based model with Adam optimizer and cross entropy loss. The suggested system's accuracy is assessed using a variety of performance indicators, including sensitivity, specificity, and F1-score. As demonstrated by the experiment findings, the suggested deep learning based classification system has a high degree of accuracy (94%), making it useful for the diagnosis and treatment of pelvic bone tumors. Our promising results might hasten the use of DL-assisted CT diagnosis for pelvic bone tumors in the future.

Corresponding Author:

Mona Shouman
Department of Computer and System Engineering, Faculty of Engineering, Minia University
Minia, Egypt
E-mail: mona.shouman1985@gmail.com

1. INTRODUCTION

Primary malignancies of the bone and joints are the third most prevalent cause of mortality for cancer patients under age 20 [1]. Different therapies are required depending on whether a bone tumor is classified as benign, intermediate, or malignant by the world health organization (WHO) [2]. Pelvic bone tumors come in a variety of types, and they are frequently treated differently. Understanding the form of a tumor is crucial for developing the appropriate treatment strategy. Bone tumors are generally benign and do not tend to spread. Although they can appear in any bone, the largest ones are often where they are located. These comprise the humerus (upper arm bone), tibia (shinbone), femur (thighbone), and pelvis. Certain kinds are more prevalent in particular regions, including the spine or the area around the greatest bones' development plates. Benign bone tumors are a broad category that includes many different forms of tumors. The most typical ones include osteochondromas, chondroblastomas, giant cell tumors, periosteal chondromas, chondroblastomas, osteoid osteomas, chondroblastomas, endochondromas, and chondromyxoid fibromas. Benign bone tumors include a number of diseases, including fibrous dysplasia, unicameral bone cysts, and cysts. Even though they aren't actually tumors, they are frequently treated similarly. Compared to benign tumors, malignant (cancerous) tumors are more harmful and uncommon. When a tumor is classified as malignant, it indicates that there is a moderate to high chance that it may spread from its original site. The lymphatic or blood vessels are the routes via which the cancer cells spread. Most frequently, malignant bone cancers spread to neighboring bones or the lungs. Bone cancerous tumors can develop at nearly any age. Two of the most prevalent malignant bone tumors, osteosarcoma, and ewing's sarcoma, often affect individuals 30 years of age or younger. On the other hand, malignant tumors called chondrosarcomas, which develop as cartilage-like tissue, typically appear beyond the age of thirty. Chondrosarcoma, chordoma, ewing's sarcoma, neuroblastoma, and osteosarcoma are examples of malignant bone tumors. Since radiographs usually evaluate the lesion's location, internal matrix, boundaries, and concurrent periosteal reaction, they are the best first-line imaging modality for evaluating bone lesions [3]. A differential diagnosis of bone lesions may frequently be made using these lesion characteristics and the patient's age [3], [4]. The issue is that radiographs have their limits. The radiographic diagnosis may be more difficult because of superimpositions, poorly seen partial cortical loss, and difficulties interpreting flat and short bones and soft tissues [5]. Radiologists have expressed interest in using contemporary computer-aided diagnostic (CAD) tools to save time and effort [6]. We use a set of CT scans because of their excellent spatial and contrast resolution. Using computed tomography (CT), it is possible to examine minute bone characteristics. Furthermore, these features enable accurate classification of minor lesions and assessment of the substitution of soft tissue for fatty marrow in metastatic lesions. In general, CT is more readily available and less expensive to study. The primary drawback of CT is that it exposes patients to more radiation than a bone scan would [7].

Many research studies have used the DL method in recent years for medicinal purposes [8], [9]. One advantage of such a method is an automatically applied feature learning model that can handle large data sets. Convolutional neural networks (CNN) are one DL approach that is thought to have the optimal architecture in many applications for classifying images. High performance in the classification, segmentation, and detection tasks of medical pictures is suggested by a CNN approach with many architectures, especially in a medical application [10]. An essential part of CNNs' conceptual design is down sampling space, weight sharing, and local perception area [11]. A lot of studies have been published recently that use different medical imaging to detect and classify bone tumors. Several DL algorithms are used in these experiments. In 2018, 3D CT scans of the spine were performed on extremely sick patients to address the segmentation and classification of difficult-to-define lytic and sclerotic metastatic lesions. Here, a CNN provided independent feature extraction [7]. Additionally, in 2018, methods for automated abdominal anatomy segmentation utilizing CT images were developed, along with methods for diagnosis, treatment planning, and therapy delivery [12]. The segmentation approaches included statistical models and multi-atlas label fusion (MALF) [13]. A new CNN architecture was used to classify three types of brain cancers. It was compared to pre-existing, pre-trained networks and found to be simpler after being tested on T1-weighted contrast-enhancedMRI. In 2020, two 10-fold cross-validation techniques, two datasets, and four methodologies were coupled to assess the network's performance [14]. Furthermore, the 2021 classification of pathologists' histological bone tumor aggressiveness will be contrasted with that of VGG- and deep learning (DL). As part of whole slide imaging (WSI), 427 pathological slides of bone tumors were created. Pathologists annotated the WSI tumor area. Four pathologists with varying levels of expertise were contrasted with the most effective models [15]. Additionally, studies were done in 2021 to develop a deep learning system that uses patient demographics and regular magnetic resonance imaging (MRI) to discern between benign and malignant bone lesions. Using T1- and T2-weighted pre-operative MRI, 1,060 histologically verified bone lesions were identified and included retrospectively [16]. Furthermore, fusion models utilizing deep learning and machine learning were developed in 2022 to categorize bone cancers as benign, malignant, or intermediate utilizing the lesion's conventional radiography as well as possibly pertinent clinical information [17]. A proposal was made in 2021 to create an intelligent clinical decision support system that would use pictures from the risk of malignancy index (RMI) to diagnose and classify brain cancers. After pre-training on an ImageNet dataset that had been appropriately matched to MRI images of brain tumors extracted from the brain tumor segmentation (BraTS) 19 database, seven CNN architectures were chosen [18]. Deep convolutional neural networks (DCNNs) were used in 2021 for deep ensemble learning, which classified skin lesions using CNNs. Using an ensemble learning method increased the F1-score performance, sensitivity, specificity, accuracy, and precision of three DCNN architectures: DenseNet 201, Inception V3, and Inception ResNet V2 [11]. Furthermore, a number of deep learning models were developed in 2021 to automatically classify brain cancers. To classify brain cancers, four deep learning models AlexNet, VGG16, GoogleNet, and RestNet50 were used [19]. In 2022, a deep neural network architecture based on the transfer learning theory was created in order to identify and categorize pictures of the histology of breast cancer. The researchers used three pre-trained CNN architectures (ResNet50, inception-v3, and VGG-16) inside the proposed framework to extract features from images and distinguish between benign and malignant tumor cells in the histopathology images of breast cancer. After that, they loaded their gathered characteristics into a fully connected (FC) layer by concatenating them [20]. A total of 3586 images of brain tumors both benign and malignant were utilized in 2022. CNN classified phase was built on top of ResNet50 architecture. To prevent over-fitting, a global average pooling (GAP) layer was included in the design for the ResNet50 output. Finally, a sigmoid layer was used to obtain the classification [6]. Additionally, in 2022, the study presented new mathematical modeling that uses effective processing to evaluate and identify the features of artificial pulse audio signals. During the training phase, which is mainly made possible by deep neural network-based learning, the numerous functional blocks of the numerical modeling are further connected with the recurrent structure of long-short-term memory (R-LSTM) feedback connections (FCs) (DNNL) [21]. Research that was recently carried out in 2023 focuses on the employment of an interval selection technique in conjunction with a heterogeneous client balancing strategy to classify EEG data using the ResNet50 deep architecture [22].

Because bone tumors can occur in a number of ways and are not common, few radiologists have the skills necessary to make an accurate diagnosis. Additionally, early tumor detection can greatly aid doctors in early management and determining the necessary treatment strategy. Thus, the goal of this work is to develop a deep learning-based system for classifying CT images of pelvic bones. The suggested technique automatically extracts information from 3D CT scans and classifies the images into several tumor types using a CNN architecture. A number of people's pelvic CT scans were used to train and evaluate the CNN model.

2. METHOD

This study proposes a novel 14-category classification scheme for pelvic bone tumors. The recommended deep learning method for pelvic bone malignancy detection is shown in Figure 1. The analysis starts with data pre-processing techniques. Pre-processing is done on images before they are added to the network. CNN classification is the next level in our pipeline. A thorough description of the data sets, CNN network hyper-parameter values, optimization method, training computations, and performance calculations can be found in the next subsection.

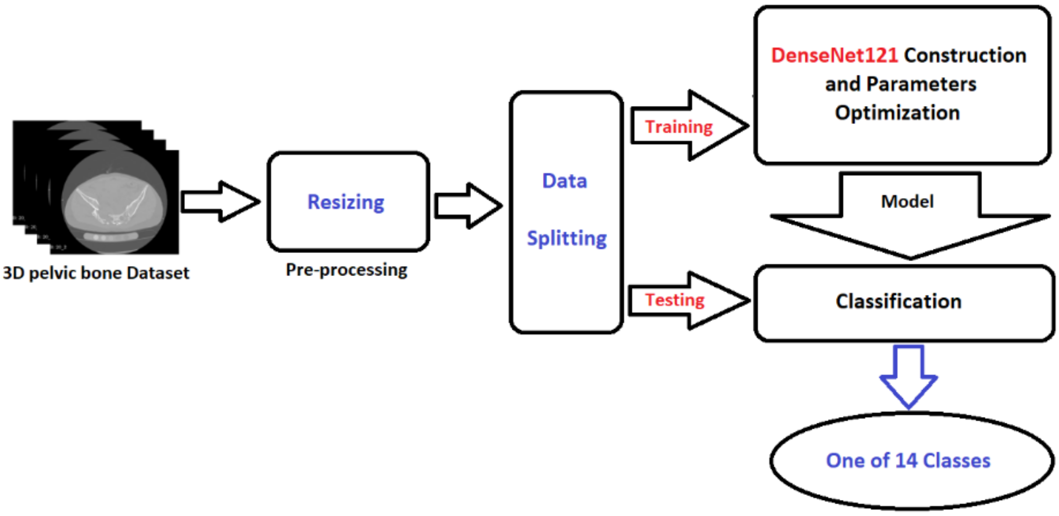

Figure 1. Block schematic illustrating the suggested approach

2.1. Dataset

We used an online dataset found at https://zenodo.org/record/4588403#.YEyLq 0zaCo on the zenodo website was established in the year 2021. This is a pelvic CT dataset that includes 75 CT scans with metal artifacts and 184 CT volumes (more than 320 K CT slices) from different manufacturers and fields [23] and the practice of multi-bone labeling is well known. Seven sources, including two clinics and five existing CT databases, provided the images [24]–[27]. These seven sub-datasets, which were independently gathered from multiple locations and sources, each contain unique characteristics that are frequently observed in clinics. In situations that are really poor quality or lack a pelvic region, the unconnected parts outside the pelvis are deleted. All images are in neuroimaging informatics technology initiative (NIfTI) format to simplify data processing using Python. Examples of the dataset with varied properties are shown in Figure 2. Table 1 presents a detailed summary of this dataset.

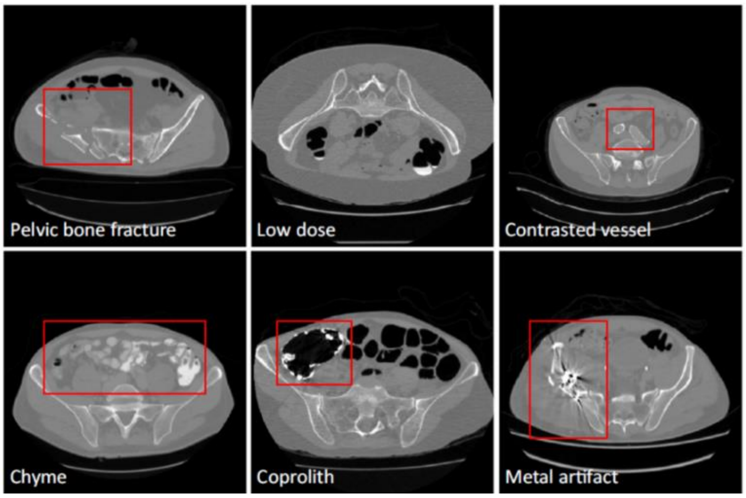

Figure 2. Images from the pelvic CT dataset examples in various contexts [23]

Table 1. Summary of the pelvic CT dataset

No.	Dataset name	Number of 3D volumes	Mean spacing (mm)	Mean size	Year
1	ABDOMEN	35	(0.76,0.76,3.80)	(512,512,73)	2015
2	COLONOG	731	(0.75,0.75,0.81)	(512,512,323)	2008
3	MSD_T10	155	(0.77,0.77,4.55)	(512,512,63)	2019
4	KITS19	44	(0.82,0.82,1.25)	(512,512,240)	2019
5	CERVIX	41	(1.02,1.02,2.50)	(512,512,102)	2015
6	CLINIC	103	(0.85,0.85,0.80)	(512,512,345)	2020
7	CLINIC-metal	75	(0.83,0.83,0.80)	(512,512,334)	2020
	Total pelvic CT dataset	1,184	(0.78,0.78,1.46)	(512,512,273)	8-2-2021
	Our dataset (CTs)	178	(0.85,0.85,0.80)	(512,512,345)	2022

The dataset has seven sub-datasets. These two sub-datasets, CLINIC and CLINIC-meta, were acquired from an orthopedic hospital and are pertinent to pelvic fractures. While CLINIC-metal is acquired from photos taken after surgery that do, CLINIC is gathered from images taken before surgery that do not. This group of data was generated by the KITS19 kidney and renal tumor segmentation challenge [24]. The CERVIX and ABDOMEN sub-datasets were produced as a result of the workshop and challenge for multi-atlas labeling outside of the cranial vault. These multi-organ segmentation datasets are all original, one-of-a-kind datasets for various body parts. Colon tumors are segmented in the ninth sub-dataset of the medical segmentation decathlon, or MSD T10 [25], [26]. COLONOG: the CT COLONOGRAPHY dataset's sub-dataset, focuses on CT colonography research [27]. We only focus on the 178 CTs that contain tumors in the pelvic bones. This dataset contains both benign and malignant pelvic bone cancers.

Our dataset is 178 3D CT images which is CLINIC and CLINIC-meta sub-datasets only because those only has the CT image with the tumors. All rest sub-datasets are labeling of the pelvic bones.

2.2. Data pre-processing

To reduce the total processing time for training and testing, a pre-processing step is first performed on the original images. Pre-processing is the term for a few basic abstraction-level operations on images. It explains every change done to the raw data before the DL model is fed into it. It is applied to improve the image and remove any prior incorrect information. Initially, every CT scan was adjusted to have a mean spacing of (0.85, 0.85, 0.80) mm with a mean size of (512, 512, 345) mm. Since the accuracy was greater while evaluating the performance with only 178 images, there is no requirement for data augmentation. Next, divide the data into two categories: training (80%) and testing (20%). Then, in order to speed up processing, resize the images one more time to the mean size of (50, 50, 50) mm.

Pelvic bone tumors have 14 classes depending on their types. The tumor is classified into its type which is very important to early detect and diagnose the tumor. The many forms of benign and malignant pelvic bone tumors are listed in Table 2. The data collection includes tumor kinds that are both benign and malignant. The specific image dataset numbers are displayed in Table 3.

Table 2. Different kinds of pelvic bone tumors, benign and malignant [28]-[34]

	Benign		Malignant
a.	Osteochondroma	a.	Osteosarcoma
b.	“Eosinophilic granuloma” (EG)	b.	Chondrosarcoma
c.	Periosteal chondroma or juxtacortical chondroma	c.	Ewing’s sarcoma (ESFTs) “ewing sarcoma family of tumors”
d.	“Chondromyxoid fibroma” (CMF)	d.	Chordoma
e.	Desmoplastic fibroma or collagenous fibroma	e.	Lymphoma
f.	Benign fibrous histiocytoma	f.	Metastatic bone carcinoma
		g.	Angiosarcoma
		h.	Hemangiopericytoma

Table 3. Details of the dataset of images

Number	Tumour	Images number
1	Osteosarcoma	53
2	Chondrosarcoma	33
3	Ewing’s sarcoma (ESFTs)	18
4	Chordoma	11
5	Lymphoma	10
6	Metastatic bone carcinoma	3
7	Angiosarcoma	2
8	Hemangiopericytoma	2
9	Osteochondroma	16
10	“EG”	12
11	Periosteal chondroma or juxtacortical chondroma	11
12	“CMF”	3
13	Desmoplastic fibroma or collagenous fibroma	2
14	Benign fibrous histiocytoma	2
	Total	178

2.3. DenseNet121 architecture

The training model is built using the dense convolutional network (DenseNet) [35]. Figure 3 shows the construction of DenseNet. Each layer uses concatenated feature maps from all previous layers as inputs rather than average feature maps. DenseNets need fewer parameters than a similar regular CNN, allowing for feature reuse because redundant feature mappings are eliminated. Consequently, the feature maps of all layers preceding the lth layer serve as its input, x_0,...,x_{l-1}:

$$x_l = H_l([x_0, x_1, \dots \dots \dots .., x_{l-1}]) \tag{1}$$

where [x_0,...,x_{l-1}] represents the concatenation of feature maps, or the output from all layers before l (0, l-1). A rectified linear unit (ReLU) [36], a batch normalization (BN) [37] and a 3×3×3 convolution (conv) are the three operations that make up the composite function H_l (). To make implementation simpler, H_l's many inputs were combined into a single tensor.

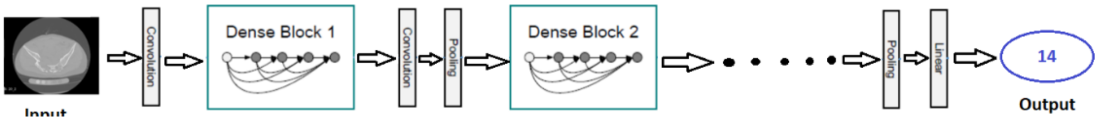

Figure 3. Block schematic of the suggested DenseNet architecture

The feature map gets larger after going through each thick layer because each layer adds 'k' features on top of the global state, or features that were already there. The rate of network expansion, denoted by parameter 'k,' dictates the amount of data that gets added to every network tier. If every function H_l produces k feature maps (where K_0 is the number of channels in the input layer), then the lth layer comprises input feature maps.

$$k_l = k_0 + k * (l - 1) \tag{2}$$

Even though each layer only produces k output feature maps, a substantial amount of input may be required, particularly for levels that come after. To increase processing speed and efficiency, a (1×1×1) convolution layer may be added as a bottleneck layer before each (3×3×3) convolution [38], [39].

DenseNet-121 has 6 AvgPools and 13 convolutions illustrated in Table 4. Each layer in the DenseNet121 design receives all of the outputs from the layers that came before it, creating an extremely dense network architecture that enables deep supervision. Each convolution block in our suggested model employs a skip relation to get around the problem of vanishing gradients. Lastly, the output layer of the original ResNet50 has been replaced with GAP. A (M×M×N) feature map, in which (M×M) represents the picture size and N the number of filters, is transformed into a (1×N) feature map via the GAP layer. Convolutional blocks are used for extracting features. The following are additional advantages for the GAP layer: At this layer, over-fitting is prevented because parameter modification is not required. It functions as a flattened layer, reducing a multidimensional input vector of extracted characteristics to a single dimension. Additionally, it requires less time.

Table 4. DenseNet121 configuration

Layers	Output Size	DenseNet-121
Convolution	50*50*50	3*3*1 conv, stride 3
Pooling	50*50*50	3*3*3 max pool, stride 3
Dense block (1)	50*50*50	$\begin{bmatrix} \mathbf{1*1} & \boldsymbol{conv} \\ \mathbf{3*3} & \boldsymbol{conv} \end{bmatrix} * \mathbf{6}$
Transition layer (1)	50*50*50	1*1*1 conv
	25*25*25	2*2*2 average pool, stride 3
Dense block (2)	25*25*25	$\begin{bmatrix} \mathbf{1*1} & \boldsymbol{conv} \\ \mathbf{3*3} & \boldsymbol{conv} \end{bmatrix} * \mathbf{12}$
Transition layer (2)	25*25*25	1*1*1 conv
	12*12*12	2*2*2 average pool, stride 3
Dense block (3)	25*25*25	$\begin{bmatrix} \mathbf{1*1} & \boldsymbol{conv} \\ \mathbf{3*3} & \boldsymbol{conv} \end{bmatrix} * \mathbf{24}$
Transition layer (3)	25*25*25	1*1*1 conv
	6*6*6	2*2*2 average pool, stride 3
Dense block (4)	6*6*6	$\begin{bmatrix} \mathbf{1*1} & \boldsymbol{conv} \\ \mathbf{3*3} & \boldsymbol{conv} \end{bmatrix} * \mathbf{16}$
Transition layer (4)	6*6*6	1*1*1 conv
	3*3*3	2*2*2 average pool, stride 3
Classification layer	1*1*1	3*3*3 global average pool
		1000D fully-connected, softmax

The 'vanishing gradient' problem manifests when the CNN's layer count rises or as the layers get deeper. It follows that when the distance between the input and output layers lengthens, some information may 'vanish' or 'get lost,' which has an impact on how well the network can train. DenseNets addresses this problem by changing the conventional CNN architecture and minimizing the link between layers. Each layer has an identical feature map size and is physically connected to the others. To keep the feed-forward nature intact, each layer broadcasts its own feature maps to all upper levels and accepts additional inputs from all previous layers. This dense connection design has the somewhat surprising benefit of requiring fewer parameters than conventional convolutional networks since duplicate feature maps do not need to be relearned. Every feature mapping in the network is used to inform the final classifier's decision. Because DenseNet layers are very tiny (e.g., 12 filters per layer), only a small fraction of feature maps contribute to the network's "collective knowledge" while the remaining feature maps remain unchanged. Among the many advantages of DenseNets are the reduction of vanishing gradient problems, improved feature propagation, ease of feature reuse, and a significant decrease in the number of parameters [35].

2.4. Training

We have used Python (v3.7), which used DenseNet121 and the help of MONAI [40]. Models were trained in PyTorch (v1.6) and Python (v3.7) using an AMD Ryzen 7 5800H graphics processing unit (GPU) running at 3.2 GHz. In the model, the numerical index labeling method is used. Assign each tumor class to a label index. The study has 14 labels. The networks are trained by defined transforms like resizing the image and enabling randrotate90 to detect any rotation of the image. The batch size is 60 for 100 epochs, respectively. Because of GPU limitations, the starting learning rate is set to 0.1 and divided by 10 at epoch 90. memory constraints. Furthermore, the DenseNet was created with spatial dimensions 3, input channels 1, and out channels=14 and enabled the cross-entropy loss [41] and Adam optimizer [42]. Finally, evaluate the performance metric.

2.4.1. Cross entropy loss

Let S stand for the sample space and L for the finite number of labels in problems requiring multi-class classification, where L=$\{l_1; l_2;:: ; l_m\}$; m > 2. The mapping relationship between sample x and label set L is many-to-one as each sample may only have one label, but several samples may share the same label. A multi-class classification challenge involves determining a sample's class based on many classes. The connection weight between the i^{th} and j^{th} neurons is shown by the symbol w_{ij}. Following is a display of the softmax.

$$y_i = \frac{\exp(z_i)}{\sum_{j=1}^{m} \exp(z_j)} \tag{3}$$

where $i \in \{1,2, \dots \dots \dots, m\}$, and the neural output $z_j = \sum_j w_{ij} x_{ij}$.
The cross-entropy function is.

$$f^t(W) = -\sum_{i=1}^{m} \widetilde{y_i} \log\left(softmax(w_i x)\right) = -\sum_{i=1}^{m} \widetilde{y_i} \log(y_i) \tag{4}$$

where m denotes the total number of classes, y_i denotes the i^{th} prediction class of MPCE, and $\widetilde{y_i}$ denotes the i^{th} true class of training samples.

2.4.2. Adam optimizer

Adam optimization is a stochastic gradient descent method that relies on an adaptive estimate of first- and second-order moments. Adam is an optimization technique that, in contrast to the traditional stochastic gradient descent method, may be used to continuously update network weights based on training data. In their 2015 ICLR paper, Diederik Kingma from OpenAI and Jimmy Ba from the University of Toronto made Adam's initial presentation. It is suitable for issues requiring large amounts of data and/or parameters because of its computational efficiency, low memory needs, invariance to diagonal rescaling of the gradients, and low memory requirements. The next algorithm shows the Adam optimization method:

Algorithm: *Adam*, our proposed algorithm for stochastic optimization.
g_t^2 indicates the elementwise square $g_t \odot g_t$.
Good default setting for the tested machine learning problems is $\alpha = 0.001$.
$\beta_1 = 0.9, \beta_2 = 0.99$ and $\epsilon = 10^{-8}$. All operations on vectors are element-wise.
With β_1^t and β_2^t we denote β_1 and β_2 to the power *t*.

```
Require: $\alpha$: Stepsize
Require: $\beta_1, \beta_2 \in [0,1)$: Exponential decay rates for the moment estimates
Require: $f(\theta)$: Stochastic objective function with parameters $\theta$
Require: $\theta_0$: Initial parameter vector
  $m_0 \leftarrow 0$ (Initialize $1^{st}$ moment vector)
  $v_0 \leftarrow 0$ (Initialize $2^{nd}$ moment vector)
  $t \leftarrow 0$ (Initialize timestep)
  while $\theta_t$ not converged do
     $t \leftarrow t + 1$
     $g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$ (Get gradients w.r.t stochastic objective at timestep t)
     $m_t \leftarrow \beta_1 . m_{t-1} + (1 - \beta_1) . g_t$ (Update biased first moment estimate)
     $v_t \leftarrow \beta_2 . v_{t-1} + (1 - \beta_2) . g_t^2$ (Update biased second raw moment estimate)
     $\widehat{m_t} \leftarrow m_t / (1 - \beta_1^t)$ (Compute bias-corrected first moment estimate)
     $\widehat{v_t} \leftarrow v_t / (1 - \beta_2^t)$ (Compute bias-corrected second raw moment estimate)
     $\theta_t \leftarrow \theta_{t-1} - \alpha . \widehat{m_t} / (\sqrt{\widehat{v_t}} + \epsilon)$ (Update parameters)
  end while
  return $\theta_t$ (Resulting parameters)
```

3. RESULTS AND DISCUSSION

Because of having multi-class classification (14 pelvic bone tumors), using multi-label confusion matrix (MLCM) [43] to calculate the accuracy of the classification technique. By measuring the classification overlap, the confusion matrix is an effective tool for performance evaluation. A two-dimensional matrix called a confusion matrix has rows that show the real labels and columns that show the classifier's anticipated labels. It is easy and simple to create a confusion matrix for a multi-class classifier.

$$M(r, c) = \sum_{i=1}^{m} \left(I(y_i = r) I(h(x_i) = c) \right), \quad All\ r, c\ \epsilon \{0, \dots, q-1\} \tag{5}$$

where *M* is the confusion matrix, r and c are the row and column of the confusion matrix, m is the number of instances in the test data set, *I*( ) is the indicator function, x_i is the i-th input to classifier *h*( ), y_i is the true label assigned to input x_i, and q is the number of classes. The next algorithm shows the method for creating the confusion matrix for a multi-class classifier.

```
Algorithm multi-class confusion matrix
  for each input instance do
    $r = assigned\ label$
    $c = predicted\ lable$
    $M(r,c) += 1$
  end for
```

Through the use of a confusion matrix, the distribution of guesses across all classes is displayed in an understandable and concise manner. After normalizing the confusion matrix row by row, the percentage of FN for every class that corresponds to each row is obtained. The cells along the main diagonal of the normalized confusion matrix show the recall for the relevant class. In the event of imbalanced data sets, the confusion matrix with the real number of counting the true and false prediction will provide information on the size of each class, while the normalized matrix is useful for figuring out the percentage of the true and false prediction. The ability to see and examine the distribution and overlap of accurately predicted labels over other labels in a single view is the primary benefit of having a confusion matrix. In order to determine accuracy, recall, and F1-score, the most important and commonly used metrics in classifier evaluation, it is also utilized to compute TP, TN, FP, and FN. The resulting multi-class confusion matrix is displayed in Table 5. The accuracy, recall, and F1-score values for every class are shown in Table 6. These may be computed in the following way for every class:

$$MPrecision = \frac{TP}{TP+FP} \tag{6}$$

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

$$F1 - score = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

Table 5. Resulted confusion matrix

		Predicted classes													
	Classes	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
True classes	C0	**11**	0	0	0	0	0	0	0	0	0	0	0	0	0
	C1	0	**9**	0	0	0	0	0	0	**1**	0	0	0	0	0
	C2	0	0	**3**	0	0	0	0	0	0	0	0	0	0	**1**
	C3	0	0	0	**53**	0	**1**	0	0	0	0	0	0	0	0
	C4	0	**1**	0	0	**10**	0	0	0	0	0	0	0	0	0
	C5	0	0	0	0	0	**3**	0	0	**1**	0	0	0	0	0
	C6	0	0	0	0	**1**	0	**33**	0	0	0	**1**	0	0	0
	C7	0	0	0	0	0	0	0	**2**	0	0	0	0	0	0
	C8	0	0	0	**1**	0	0	0	0	**14**	0	0	0	0	0
	C9	1	0	0	0	0	0	0	0	0	**17**	0	0	0	0
	C10	0	0	0	0	0	0	0	0	0	0	**10**	0	0	0
	C11	**1**	0	0	0	0	0	0	**1**	0	0	0	**3**	0	0
	C12	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0
	C13	0	0	0	0	0	0	**1**	0	0	**1**	0	0	0	**6**

Table 6. The results of each measurement for each class

Class	Precision	Recall	F1-score
C0	0.99	0.917	0.995
C1	0.9	0.9	0.989
C2	0.75	0.99	0.99
C3	0.98	0.98	0.99
C4	0.99	0.91	0.995
C5	0.75	0.99	0.995
C6	0.97	0.97	0.995
C7	0.99	0.667	0.99
C8	0.99	0.99	0.984
C9	0.99	0.944	0.99
C10	0.99	0.99	0.957
C11	0.99	0.99	0.989
C12	0.99	0.99	0.989
C13	0.99	0.99	0.984

Despite its name, the multilabel_confusion_matrix function in the sklearn Python package [44] is a one-versus-rest confusion matrix (much like the binary confusion matrix). This means that TP, TN, FP, and FN are determined for each class using the multilabel_confusion_matrix function in sklearn.

One of the measures most commonly employed in multi-class classification is accuracy, which is computed directly from the confusion matrix.

$$Total\ Accuracy = \frac{Total\ TP}{Total\ TP + Total\ FP} \tag{9}$$

The accuracy formula considers the total number of true positive and true negative components in the numerator as well as the total number of entries in the confusion matrix in the denominator. The items on the confusion matrix's major diagonal that the model properly recognized are known as the true positives and true negatives, while everything outside of the major diagonal that the model incorrectly categorized is contained in the denominator. Table 5 of the confusion matrix displays the 94% accuracy of the model. While it enhances some classes' performance and is less sensitive to imbalanced data, relying only on it might be deceptive [45]. As a result, we also used f1 score, accuracy, and recall metrics. Because of this, we ought to have a reliable indicator of model generalization across the dataset of imbalanced CT images. Since later metrics are affected by class inequality, they are used to show the model's overall performance independent of the number of each particular class.

The confusion matrix shows the network’s predictions for 178 images in each category. If all values on the diagonal were 178, this would indicate that each test image was correctly classified. Clearly, for our network, this is not the case. The values outside the diagonal give a sense of which category is getting misclassified. If there are more images, the accuracy will be 100%.

Positive predictive value, or precision, is a metric that expresses the percentage of positive class predictions that really fall into the positive class. For most of the classes’ precision is higher than 0.9 which is illustrated in Table 6 except for class number two, which means the classifier is correct by more than 90% to indicate the correct type of tumor. The number of accurate class predictions made from all of the positive examples in the dataset is measured by recall, sometimes referred to as sensitivity. In our method, the recall value for most of the classes is more than 90% except for class number seven, which means the model is correct by more than 90% shown in Table 6. Sometimes the two metrics are combined into the F1-score (or f-measure), which offers a single measurement for a system. It provides a single score that resolves the memory and accuracy problems into a single figure. For every class, we attain an F1-score of greater than 90%.

Because of their shorter connections, DenseNet may be more accurate in part because each layer receives more supervision from the loss function. One may say that DenseNets performs a form of "deep supervision." Deep-supervised networks function by promoting the intermediate layers that accumulate discriminative features [46], which incorporate classifiers coupled to every hidden layer, have previously proved the benefits of deep supervision. By using a maximum of two or three transition layers, a single classifier placed on top of the network provides direct supervision to all levels. In an implied approach, DenseNets carry out a similar deep supervisory function in this way. Since the same loss function is used for all layers in DenseNets, the gradient and loss function are easier to understand.

Even though the single DenseNet121 with a fine-tuning model has fewer parameters than many CNN networks, it still produces extremely good results. DenseNet 121's performance in this experiment demonstrates that it can be trained on many datasets display the classification results using the ensemble learning approach, with an average accuracy of 94%. In addition, the use of DenseNet overcomes the vanishing gradient problem. Unfortunately, this study has a limitation such as a lack of pelvic bone datasets but with these few medical images, the model obtains very good results.

In addition, a review of previous studies comparing different approaches to classifying different types of bone lesions was done to assess the approach proposed in this publication. Table 7 reports and compares the new methodology approach's outcomes with the previous method on a range of bone lesions using a variety of medical images. The suggested ensemble model generates more accurate results than prior research, as demonstrated in Table 7. Because the ensemble model has a more complex architecture, the performance outcomes are evaluated during the training, validation, and testing phase to ensure higher performance. Table 7 compares our research to various techniques that have already been published. The table shows for MRI the maximum accuracy was 86% at year 2021 [17]. But when using WSI the accuracy was 85.96% [15]. However, our method using 3D CT images gives higher result (94%).

Table 7. Accuracy of published methods

Method	Year	Medical image type	Accuracy (%)
Our method	2023	3D CT	94
Liu *et al.* [17]	2022	MRI	86
Eweje *et al.* [16]	2021	MRI	76
Tao *et al.* [15]	2021	WSI	85.96
Chmelik *et al.* [7]	2018	3D CT	80

4. CONCLUSION

The large intra-class variations and inter-class similarities in tumor size, location, and appearance make pelvic bone tumor diagnosis extremely challenging. In summary, this study demonstrates that 94% of the model's ultimate accuracy is achieved after training. We can attain a classification performance that is significantly greater than the results of prior studies based on the suggested ensemble model. The experimental findings show that the suggested frameworks produce effective results. Addition, using of DenseNet solves the vanishing gradient problem. Deep learning is more effective than young radiologists in classifying pelvic bone tumors from CT scans with similar accuracy as subspecialists. Our findings are encouraging and might hasten the use of DL-assisted pelvic bone tumor diagnosis in the future.

REFERENCES

[1] E. Ward, C. DeSantis, A. Robbins, B. Kohler, and A. Jemal, "Childhood and adolescent cancer statistics, 2014," *CA: A Cancer Journal for Clinicians*, vol. 64, no. 2, pp. 83–103, Mar. 2014, doi: 10.3322/caac.21219.

[2] H. A. Loomans-Kropp and A. Umar, "Increasing Incidence of Colorectal Cancer in Young Adults," *Journal of Cancer Epidemiology*, vol. 2019, pp. 1–9, Nov. 2019, doi: 10.1155/2019/9841295.

[3] I. Zambo and K. Veselý, "[WHO classification of tumours of soft tissue and bone 2013: the main changes compared to the 3rd edition].," *Ceskoslovenska patologie*, vol. 50, no. 2, pp. 64–70, 2014.

[4] P. Picci *et al.*, Eds., Diagnosis of Musculoskeletal Tumors and Tumor-like Conditions. *Springer International Publishing*, 2020. doi: 10.1007/978-3-030-29676-6.

[5] P. Y. Huang *et al.*, "Osteomyelitis of the femur mimicking bone tumors: A review of 10 cases," *World Journal of Surgical Oncology*, vol. 11, no. 1, p. 283, Dec. 2013, doi: 10.1186/1477-7819-11-283.

[6] M. Ahmed, F. Khalifa, H. E. D. Moustafa, G. A. Saleh, and E. AbdElhalim, "A deep learning based system for accurate diagnosis of brain tumors using T1-w MRI," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 2, pp. 1192–1202, Nov. 2022, doi: 10.11591/ijeecs.v28.i2.pp1192-1202.

[7] J. Chmelik *et al.*, "Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data," *Medical Image Analysis*, vol. 49, pp. 76–88, Oct. 2018, doi: 10.1016/j.media.2018.07.008.

[8] S. Nurmaini, A. Darmawahyuni, A. N. S. Mukti, M. N. Rachmatullah, F. Firdaus, and B. Tutuko, "Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification," *Electronics (Switzerland)*, vol. 9, no. 1, p. 135, Jan. 2020, doi: 10.3390/electronics9010135.

[9] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of Deep Learning and Reinforcement Learning to Biological Data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2063–2079, Jun. 2018, doi: 10.1109/TNNLS.2018.2790388.

[10] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018, doi: 10.1007/s13244-018-0639-9.

[11] R. A. Pratiwi, S. Nurmaini, D. P. Rini, M. N. Rachmatullah, and A. Darmawahyuni, "Deep ensemble learning for skin lesions classification with convolutional neural network," *IAES International Journal of Artificial Intelligence*, vol. 10, no. 3, pp. 563–570, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp563-570.

[12] E. Gibson *et al.*, "Automatic Multi-Organ Segmentation on Abdominal CT with Dense V-Networks," *IEEE Transactions on Medical Imaging*, vol. 37, no. 8, pp. 1822–1834, Aug. 2018, doi: 10.1109/TMI.2018.2806309.

[13] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 611–623, Mar. 2013, doi: 10.1109/TPAMI.2012.143.

[14] M. M. Badža and M. C. Barjaktarović, "Classification of brain tumors from mri images using a convolutional neural network," *Applied Sciences (Switzerland)*, vol. 10, no. 6, p. 1999, Mar. 2020, doi: 10.3390/app10061999.

[15] Y. Tao *et al.*, "Qualitative Histopathological Classification of Primary Bone Tumors Using Deep Learning: A Pilot Study," *Frontiers in Oncology*, vol. 11, Oct. 2021, doi: 10.3389/fonc.2021.735739.

[16] F. R. Eweje *et al.*, "Deep Learning for Classification of Bone Lesions on Routine MRI," *EBioMedicine*, vol. 68, p. 103402, Jun. 2021, doi: 10.1016/j.ebiom.2021.103402.

[17] R. Liu *et al.*, "A deep learning–machine learning fusion approach for the classification of benign, malignant, and intermediate bone tumors," *European Radiology*, vol. 32, no. 2, pp. 1371–1383, 2022, doi: 10.1007/s00330-021-08195-z.

[18] H. El Hamdaoui *et al.*, "High precision brain tumor classification model based on deep transfer learning and stacking concepts," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 1, pp. 167–177, Oct. 2021, doi: 10.11591/ijeecs.v24.i1.pp167-177.

[19] A. A. Abbood, Q. M. Shallal, and M. A. Fadhel, "Automated brain tumor classification using various deep learning models: A comparative study," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 1, pp. 252–259, Apr. 2021, doi: 10.11591/ijeecs.v22.i1.pp252-259.

[20] A. M. Hassan, M. B. El-Mashade, and A. Aboshosha, "Deep learning for cancer tumor classification using transfer learning and feature concatenation," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 6, pp. 6736–6743, Dec. 2022, doi: 10.11591/ijece.v12i6.pp6736-6743.

[21] M. N. Meqdad, S. O. Husain, A. M. Jawad, S. Kadry, and A. R. Khekan, "Classification of electroencephalography using cooperative learning based on participating client balancing," *International Journal of Electrical and Computer Engineering*, vol.

13, no. 4, pp. 4692–4699, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4692-4699.

[22] M. V. V. P. Kantipudi and S. Kumar, “A computationally efficient learning model to classify audio signal attributes,” *International Journal of Electrical and Computer Engineering*, vol. 12, no. 5, pp. 4926–4934, Oct. 2022, doi: 10.11591/ijece.v12i5.pp4926-4934.

[23] P. Liu *et al.*, “Deep learning to segment pelvic bones: large-scale CT datasets and baseline models,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 5, pp. 749–756, May 2021, doi: 10.1007/s11548-021-02363-8.

[24] B. Landman, Z. Xu, J. E. Igelsias, M. Styner, T. R. Langerak, and A. Klein, “Multi-atlas labeling beyond the cranial vault-workshop and challenge,” *In Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge*, vol. 5, p. 12, 2017, doi: 10.7303/syn3193805.

[25] N. J. Sathianathen *et al.*, “Automatic Segmentation of Kidneys and Kidney Tumors: The KiTS19 International Challenge,” *Frontiers in Digital Health*, vol. 3, Jan. 2022, doi: 10.3389/fdgth.2021.797607.

[26] C. D. Johnson *et al*, "Accuracy of CT colonography for detection of large adenomas and cancers," *New England Journal of Medicine,*, vol. 359, no. 12, 2008, doi: 10.1056/NEJMoa0800996

[27] A. L. Simpson *et al.*, “A large annotated medical image dataset for the development and evaluation of segmentation algorithms,” *arXiv*, 2019, doi: 10.48550/arXiv.1902.09063.

[28] E. Michel, “Fundamentals of Skeletal Radiology. 2nd ed,” *Radiology*, vol. 198, no. 3, pp. 850–850, 1996, doi: 10.1148/radiology.198.3.850.

[29] T. T. Miller, “Bone tumors and tumorlike conditions: Analysis with conventional radiography,” *Radiology*, vol. 246, no. 3, pp. 662–674, Mar. 2008, doi: 10.1148/radiol.2463061038.

[30] R. Chugh, H. Tawbi, D. R. Lucas, J. S. Biermann, S. M. Schuetze, and L. H. Baker, “Chordoma: The Nonsarcoma Primary Bone Tumor,” *The Oncologist*, vol. 12, no. 11, pp. 1344–1350, Nov. 2007, doi: 10.1634/theoncologist.12-11-1344.

[31] L. G. Kindblom, “Bone tumors: epidemiology, classification, pathology,” *Imaging of bone tumors and tumor-like lesions: techniques and applications*, pp. 1–15, 2009.

[32] J. L. Bloem and I. I. Reidsma, “Bone and soft tissue tumors of hip and pelvis,” *European Journal of Radiology*, vol. 81, no. 12, pp. 3793–3801, Dec. 2012, doi: 10.1016/j.ejrad.2011.03.101.

[33] G. Girish, K. Finlay, D. Fessell, D. Pai, Q. Dong, and D. Jamadar, “Imaging review of skeletal tumors of the pelvis malignant tumors and tumor mimics,” *The Scientific World Journal*, vol. 2012, pp. 1–12, 2012, doi: 10.1100/2012/240281.

[34] M. Sundaram and D. Vanel, “Tumors and tumor-like lesions of bone,” *Musculoskeletal Diseases: Diagnostic Imaging and Interventional Techniques*, pp. 62–67, 2005, doi: 10.1007/88-470-0339-3_10.

[35] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.

[36] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” *Journal of Machine Learning Research*, vol. 15, pp. 315–323, 2011.

[37] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *32nd International Conference on Machine Learning, ICML 2015*, vol. 1, pp. 448–456, 2015.

[38] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[39] M. Shafiq and Z. Gu, “Deep Residual Learning for Image Recognition: A Survey,” *Applied Sciences (Switzerland)*, vol. 12, no. 18, p. 8972, Sep. 2022, doi: 10.3390/app12188972.

[40] MONAI, “Medical Open Network for Artificial Intelligence”, [Online]. Available: https://monai.io/

[41] Y. Zhou, X. Wang, M. Zhang, J. Zhu, R. Zheng, and Q. Wu, “MPCE: A Maximum Probability Based Cross Entropy Loss Function for Neural Network Classification,” *IEEE Access*, vol. 7, pp. 146331–146341, 2019, doi: 10.1109/ACCESS.2019.2946264.

[42] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *arXiv*, 2015, doi: 10.48550/arXiv.1412.6980

[43] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-Label Confusion Matrix,” *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.

[44] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

[45] T. Tazin *et al.*, “A Robust and Novel Approach for Brain Tumor Classification Using Convolutional Neural Network,” *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–11, Dec. 2021, doi: 10.1155/2021/2392395.

[46] Y. K. Park and R. K. Kalil, “Tumors and tumor-like lesions of bone for surgical pathologists, orthopedic surgeons and radiologists,” *Tumors and Tumor-Like Lesions of Bone: For Surgical Pathologists, Orthopedic Surgeons and Radiologists*, pp. 1–994, 2015, doi: 10.1007/978-1-4471-6578-1.

BIOGRAPHIES OF AUTHORS

Mona Shouman earned her B.Sc. and M.Sc. in Communications and Electronic Engineering from the Faculty of Engineering, Cairo University in June 2007 and October 2011 respectively. She was an assistant lecturer at HBRC from 2008 to 2010. From 2011 to 2012, she was an assistant lecturer at Ain-Shams University. She was an assistant lecturer at MTI University from 2013 to 2018. In addition, worked at PHI as an assistant lecturer from 2019 to 2021. Her research interests are artificial intelligence, satellites, automatic control, and acoustics. She can be contacted at email: mona.shouman1985@gmail.com.

Kamel Hussein Rahouma earned his B.Sc. and M.Sc. in Communications and Electronic Engineering from the Faculty of Engineering, Cairo University in June 1984 and March 1988 respectively. He earned his first doctoral degree in Communications and Electronics Engineering jointly from the University of Kent at Canterbury (England) and Minia University (Egypt) in 1996. In May 2001, he earned his second doctoral degree in Computer Science from the University of Salzburg, Austria. Professor Rahouma has multidisciplinary research interests. This includes (but is not limited to): cryptography and information security, artificial intelligence applications, biomedical applications, bioinformatics, geolocation systems, lifi communication systems, smart card systems, embedded systems, driverless vehicles, design and implementation of satellite systems, and educational applications. He is currently supervising and guiding more than 40 postgraduate students who are working for their Master and Doctoral degrees. He has taught and is teaching various undergraduate courses in Egypt and Saudi Arabia. This may include (but not limited to): analog and digital communications, electronics, electronic circuits, digital design, electric circuits, engineering mathematics, probability and statistics, control systems, artificial intelligence, image processing, embedded systems, computer programming (FORTRAN, java, C, C++, MATLAB, basic, visual basic, ….), cryptography and network security, operating systems, information systems, optimization, computer networks, operational research, distributed, and parallel computer systems. He is also guiding many bachelors graduation projects in different fields of computer science, communications and electronics engineering, and biomedical engineering. He has a strong ambition to work with teams in different areas of science. He is dreaming of a clean healthy society free of diseases. He can be contacted at email: kamel_rahouma@yahoo.com.

Hesham Fathy Aly Hamed was born in Giza, Egypt, in 1966. He received the B.Sc. degree in Electrical Engineering, the M.Sc. and Ph.D. degrees in Electronics and Communications Engineering from Minia University, EL-Minia, Egypt, in 1989, 1993, and 1997 respectively. He was the dean of Faculty of Engineering, Minia University. He was a Visiting Researcher at Ohio University, Athens, Ohio. From 1989 to 1993 he worked as a Teacher Assistant in the Electrical Engineering Department, Minia University. From 1993 to 1995 he was a visiting scholar at Cairo University, Cairo, Egypt. From 1995 to 1997 he was a visiting scholar at Texas A&M University, College Station, Texas (with the group of VLSIS). From 1997 to 2003 he was an Assistant Professor in the Electrical Engineering Department, Minia University. From 2003 to 2005 he was Associate Professor in the same University. He is currently the dean of Faculty of Artificial Intelligence, Russian University, Cairo. He has published more than 200 papers. His research interests include analog and mixed-mode circuit design, low voltage low power analog circuits, current mode circuits, nano-scale circuits design, FPGA, and applications of artificial intelligence. He can be contacted at email: hfah66@yahoo.com.