Applied Computer Systems
ISSN 2255-8691 (online)
ISSN 2255-8683 (print)
June 2023, vol. 28, no. 1, pp. 66–77
https://doi.org/10.2478/acss-2023-0007
https://content.sciendo.com

Recognition and 3D Visualization of Human Body Parts and Bone Areas Using CT Images

Hai Thanh Nguyen[1*], My N. Nguyen[1,2], Bang Anh Nguyen[1], Linh Chi Nguyen[1], Linh Duong Phung[3]
[1]*College of Information and Communication Technology, Can Tho University, Can Tho, Vietnam*
[2]*Kyoto Institute of Technology, Kyoto, Japan*
[3]*College of Information Science and Engineering, Ritsumeikan University, Kyoto, Japan*

***Abstract* – The advent of medical imaging significantly assisted in disease diagnosis and treatment. This study introduces to a framework for detecting several human body parts in Computerised Tomography (CT) images formatted in DICOM files. In addition, the method can highlight the bone areas inside CT images and transform 2D slices into a visual 3D model to illustrate the structure of human body parts. Firstly, we leveraged shallow convolutional Neural Networks to classify body parts and detect bone areas in each part. Then, Grad-CAM was applied to highlight the bone areas. Finally, Insight and Visualization libraries were utilized to visualize slides in 3D of a body part. As a result, the classifiers achieved 98 % in F1-score in the classification of human body parts on a CT image dataset, including 1234 slides capturing body parts from a woman for the training phase and 1245 images from a male for testing. In addition, distinguishing between bone and non-bone images can reach 97 % in F1-score on the dataset generated by setting a threshold value to reveal bone areas in CT images. Moreover, the Grad-CAM-based approach can provide clear, accurate visualizations with segmented bones in the image. Also, we successfully converted 2D slice images of a body part into a lively 3D model that provided a more intuitive view from any angle. The proposed approach is expected to provide an interesting visual tool for supporting doctors in medical image-based disease diagnosis.**

***Key Terms* – 3D visualization, bone highlighting, convolutional neural networks, CT image classification, DICOM Grad-CAM, human body parts.**

I. Introduction

Medical image processing has recently attracted many scientists with extensive research and applications based on different technologies. This field success helps doctors give fast and precise diagnoses, improving treatment efficiency. With medical imaging, the physical measurements are acquired and reflected as images, providing an in-depth insight into the human body's structure and function. Some contemporary ones hear a lot about acquisition techniques, such as X-rays, Computed Tomography Scans (CT Scan), Magnetic Resonance Imaging (MRI), and Ultrasound. Furthermore, X-ray produces images of the inside of the body, especially the painless bones, and takes a few minutes. It uses X-radiation with wavelengths ranging from 10 picometers to 10 nanometers. Depending on how dense a body part is, X-rays will result in different intensities on the detector after passing through it, creating different brightness levels on the image. Dense parts of the body, like the bones, hamper X-rays as they pass through, resulting in bright areas on the image, while other softer parts that mostly contain air, like the lungs, will be represented as dark areas. This method is mainly used for detecting problems in the bones, teeth, lungs, and heart. CT Scan takes an X-ray image of a body part several times from different angles to create a series of slices. Then a computer combines these slices to create a more detailed image of that part. This method gives doctors more information about the body part structure as they can view it from any perspective. CT Scan does not cause any pain for the patient and takes about 15 to 30 min. It can be used for several purposes, such as detecting tumours, bone fractures, heart disease, and internal injuries. Unlike the above techniques, MRI uses powerful magnets and radio waves to take two-dimensional images of the inside of the body instead of X-rays. A computer then combines those images to produce an intuitive 3D image of the internal body parts. MRI is a valuable technique for observing many body parts, including the brain, bone, heart, liver, etc. Ultrasound is a technique that produces internal organ images by emitting high-frequency sound waves toward the body. Based on the echoes, a computer represents them as an image on the monitor corresponding to what is happening inside the patient's body. Ultrasound is primarily used in antenatal examination. It can also be used to examine the condition of other organs, including the heart, liver, kidneys, bladder, uterus, etc.

Digital Imaging and Communications in Medicine (DICOM) are universally adopted for the file format. DICOM is an international medical image standard encompassing more than just an image, such as patient information, acquisition setting, and communication protocol. DICOM standardizes the management of digital medical images and related data. In addition, it enables communication among medical imaging devices from different vendors, allowing them to exchange medical image data with any device that supports the DICOM format. When working with medical images, doctors only view them from a 2D perspective. However, this manner needs to be more intuitive, and the natural shape of the target object needs

* Corresponding author's e-mail: nthai.cit@ctu.edu.vn

©2023 Hai Thanh Nguyen, My N. Nguyen, Bang Anh Nguyen, Linh Chi Nguyen, Linh Duong Phung.
This is an open access article licensed under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0).

to be appropriately reflected. Many platforms that support the 3D rendering of medical images have come into place to tackle this hitch, such as AMI Medical Imaging, 3D Slicer, and ITK/VTK Viewer. Besides, machine learning algorithms have been widely adopted in medical image processing – these algorithms aid doctors in making quick and accurate treatment decisions for their patients. Some studies have solved the problem of medical image classification, including studies of [1], [2], and the detection of skin cancer with images by [3]. Furthermore, many deep learning-based approaches have become successful in clinical practice, including decision support systems (CDSS), diagnosis, and prediction of the invasion of lung carcinoma based on radio waves and clinical features by [4].

This study focuses on several main tasks: recognising parts of the human body, highlighting the bones from DICOM files, and visualizing them into 3D models. Firstly, we trained a CNN to classify DICOM images of human body parts. Then, after splitting the bone and non-bone pixels in each body part image, we trained another CNN to classify them. Next, Grad-CAM was used to localise and highlight the bones in the original images. Finally, using the Insight Toolkit (ITK) and the Visualization Toolkit (VTK) libraries by [5] is our method to transform those 2D images into lively 3D models. These directions have powerfully attracted numerous studies by a vast of scientists. The main contributions can be summarised as follows:

- We deploy shallow convolutional neural networks for classifying CT images of four body parts in DICOM format. We explore the potential of shallow architectures with one (CNN-1), two (CNN-2), or three (CNN-3) stacks of a convolutional layer, including 32 filters (64 filters for the second stack of CNN-2 and 128 filters for the third stack of CNN-3) of the size of 3×3, followed by a ReLU, and a max pooling layer of 2×2 (stride 2) in human body parts classifications and detecting bone areas in CT images. For four considered parts, including the head, hip, pelvis, and shoulder, the model trained on slices from a woman can reach an accuracy of 98 % for predicting body parts from slices of a man. In addition, we can see that images captured from a gender can be used for predicting another gender.
- We generate images that contain only bone areas by setting a threshold in CT images from the images capturing human body parts. From such images, including bone and non- bone areas, we fetched into a shallow convolutional neural architecture for the training phase for bone detection. Then, we highlight the bones in CT images using visual Grad-CAM explanations from the trained model.
- We build a web application allowing users to combine 2D CT slices of a body part to create an interactive 3D model viewed from any angle. Moreover, we can examine the texture inside the 3D model by viewing its slices along the different axes.

The remainder of this paper is organised as follows. Section II presents relevant studies. Section III describes the dataset used for our experiments, the data pre-processing steps for generating bone images for bone detection tasks, and architectures for classification tasks. The study outcomes are presented in Section IV. Finally, we discuss and summarise our work in Section V.

II. Related Work

Numerous approaches were presented to solve the problems related to classifying parts of the human body in images. The paper [6] introduced a novel approach to generating 'visual explanations' for decisions from deep networks called Grad-CAM (Gradient-weighted Class Activation Mapping). The technique supports predicting the concept through a generated map in which all the potential areas are highlighted, and it can be used with many CNN-based models. The authors used Grad-CAM and existing fine-grained visualizations to make a high-resolution class-discriminative visualization for image classification, captioning, and visual question answering. Grad-CAM is proven to evaluate deep networks successfully. The proposed approach is used to leverage the possibility of the fashion item localization in an image by drawing bounding boxes and labelling them. It can extract fashion item label-pixel data with a pixel-wise semantic segmentation function. The authors also introduced a relevant study running on video data and showed that the system flexibility, output quality, and reliability are better than other platforms. Furthermore, they have proven that the solution can give localization and segmentation services, and the core architecture is still extensible and straightforward. The study [7] described the importance and challenges of post-processing DICOM imaging data – an involved step of the process of fabricating a 3D-printed physical model that would allow the manufacturing of both patient-specific 3D-printed models and medical devices. This work requires different techniques from traditional image post-processing, and a compatible digital file format with 3D printers is needed. The paper [8] aimed at providing a new solution to the intensity inhomogeneities in image segmentation. Initially, each point is derived from a local intensity clustering property of the image intensities. Then, a local clustering criterion function for the image intensities in their neighbourhood is defined and combined with the neighbourhood centre to get a global criterion of image segmentation. The proposed method can segment the image and estimate the bias field simultaneously by minimizing energy defined concerning the level set functions representing a partition of the image domain and a bias field that accounts for the image intensity heterogeneity. The method is proven to give desired results for the experiments on synthetic and real images of various modalities and outperform other models. It has been used to apply the segmentation and bias correction of magnetic resonance (MR) images. The study [9] was based on the 3D human body measurement technology conducted on 211 female body data aged from 50 to 59 in central China and explained the specific characteristics of various categories of the body. The researchers [10] classified human body parts from work pieces. They compared the performance between a support vector machine (SVM) and a random forest classifier for classification tasks on body parts. The study [11] explored and performed the

three-dimensional morphological feature extraction to show the front and side of the human body. This task aimed at determining the control points and their specific values that would best reveal their body shape. The authors [12] evaluated the diagnostic performance of virtual single-frame 4D-DSA and compared to 2D-DSA images by checking out clinicians' ability to evaluate cerebrovascular pathology. Two neuro-interventional radiologists were utilised for rating single-frame images of four projections from 4D-DSA and their corresponding 2D-DSA images ($n = 15$). Diagnostic quality (0 = non-diagnostic, 1 = poor, 2 = acceptable, 3 = good) acted as the basis for grading images. Each case's dose area product (DAP) was recorded for all 2D-DSA and 4D-DSA acquisitions and the overall procedure. The mean diagnostic quality of all four 4D-DSA projections from both raters was 1.75, and that of 2D-DSA projections was 2.8. The diagnostic quality of 4D-DSA and 2D-DSA at all four projections ($p < 0.001$) was considerably different for the student's t-test. 4D-DSA acquisitions accounted for 30 % dose compared to the overall average aggregated dose per procedure. The study indicated that the difference in image quality between virtual single-frame 4D-DSA and their corresponding 2D-DSA images was essential for statistics, and 4D-DSA acquisitions had a needless radiation dose than conventional procedures with 2D-DSA acquisitions. Therefore, enhancing medical report confidentiality in DICOM images is vital in the medical profession. The authors [13] introduced a new blind watermarking method. Each bit was inserted to subtract adjacent DC coefficients of the FFT (fast Fourier transformation) using an essential kw. The proposed system generating a bit sequence in the critical kw generated eXOR with each watermark bit. The proposed method was robust to single, double, and triple attacks and outperformed other proposed watermarking schemes. Some interesting works on human body classification were introduced in other studies [14]–[18].

Techniques for 3D visualization from 2D images are also proposed in many studies. The research [19] presented the advantages of adopting DICOM for digital and computational pathology (reusing enterprise-wide infrastructure for storage, security, and business continuity) and its sound characteristics. To recognize intussusception in children, the authors [20] implemented a deep convolutional neural network (CNN) algorithm using labelled datasets of plain abdominal X-rays from affected children (1449 images of intussusception patients and 9935 images from patients without intussusception). Single Shot MultiBox Detector was used for detecting abdominal, and intussusception was classified using ResNet. The internal test values (training with two hospital data sets) and external test values (using the remaining data set) were pretty high. They proved that deep CNN and plain abdominal X-rays to detect intussusception could help screen children. The authors [21] presented the difficulties they faced while developing machine learning methods to aid in diagnosing and radiology reading of whole-body magnetic resonance imaging (MRI) scans in cancer staged patients (MALIBO study). The two vital characteristics of data for clinical trials incorporating machine learning were data quality and uniformity. Also, efficient ways for storing the machine learning outcomes and the clinical images for clinical purposes were indispensable for obtaining good clinical results in the MALIBO study. Lastly, the authors emphasised that we should consider the problems related to legal, ethical, and clinical acceptance when applying computer-assisted tools to clinical practice. The study [22] implemented a simple Diagonal-kernel with a diagonal kernel and an anti-diagonal kernel. CIFAR and ImageNet classification datasets evaluated the Diagonal kernel performance. The experimental outcomes on the CIFAR and ImageNet classification datasets show that the Diagonal-kernel can give high accuracy when reducing parameters and computational cost, have larger local receptive fields, and are more efficient. However, the fine-grained and imbalanced image dataset results indicate that Diagonal-kernel has a more considerable accuracy loss than coarse-grain images. Still, the loss is acceptable, and imbalanced data does not affect its performance. Because of the small number of weights for deep convolution, Diagonal-kernel is mostly for traditional convolution, not depth-wise convolution. A novel CNN containing a concatenation of multiple networks, namely C-Net, was proposed by [23] for automatically classifying cancer types. Unlike conventional deep learning models in biomedical image classification, no prior knowledge is employed. The model contains multiple CNNs, namely Outer, Middle, and Inner. Six networks of the first two parts of the architecture extract features. Then they are utilized for classifying the images in malignancy and benignancy at the Inner network. The experiments of this study run on BreakHis and Osteosarcoma public datasets. The experimental outcomes show that the proposed model outperforms all other models on the individual metrics for both datasets and gets no misclassification. Some other research on medical images visualizations are presented by [24]–[29] with methods to provide virtual reality applications, detection, and segmentation for pulmonary nodule 3D visualization diagnosis. The authors [30] conducted 3D visualization from a volume rendering technique. In contrast, researchers, including [31], attempted a multi-planar reformatting, surface rendering, and hardware-accelerated volume rendering for visualizing bone in 3D.

Although numerous studies on 3D visualization have been proposed, bone visualization of human body parts in 3D spaces using machine learning model explanations such as Grad-CAM is still limited. Therefore, our study has attempted shallow architectures to detect bones in human parts and visualize them in 3D space using Grad-CAM.

III. METHODS

The overall workflow of our proposed framework is illustrated in Fig. 1 with three major tasks as follows:

- *The body part image classification and bone detection tasks*: These tasks are formatted with the blue (for human body part classification) and orange (for bone detection) backgrounds in Fig. 1 and detailed in Fig. 4. The task reads DICOM files containing CT images and categorises them into folders of four corresponding body parts (head, hip, pelvis, and shoulder). Then, the data were fetched into

learning models (using the Convolutional Neural Networks as exhibited in Figs. 2 and 3 for human body parts, bone detection tasks, respectively) to perform the classification task of human body parts. The dataset for bone detection was generated by keeping pixels equal to or greater than 150 and labelled as bone, while the remaining was considered a non-bone area.

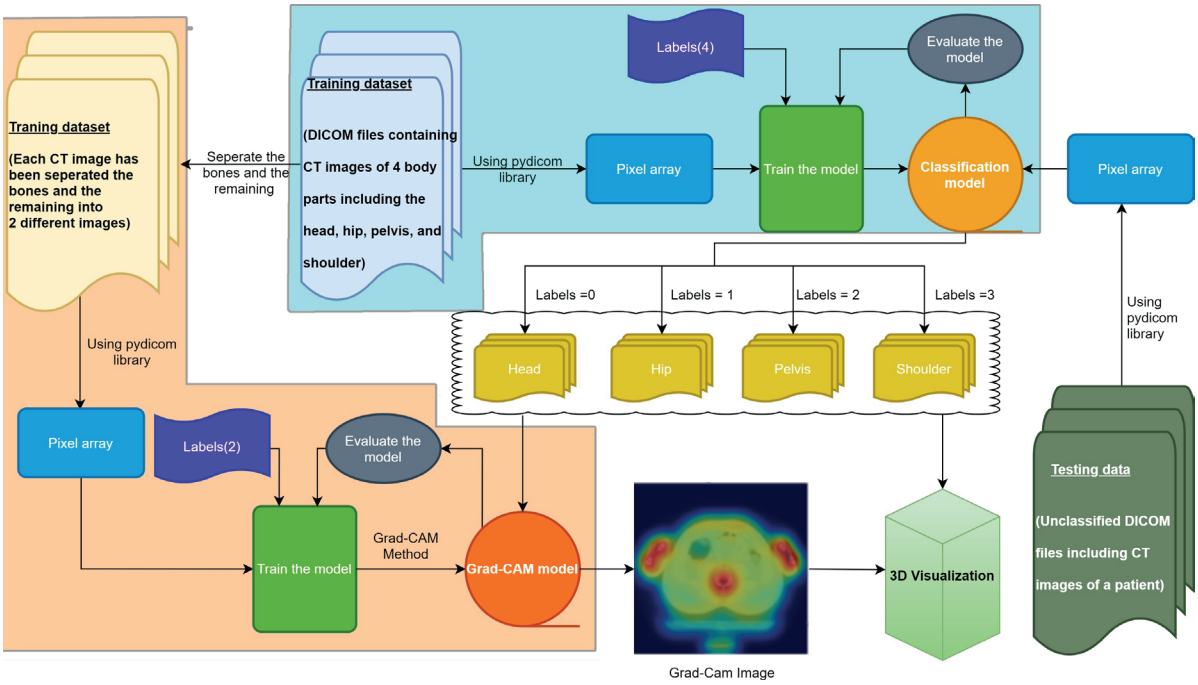


Fig. 1. The proposed workflow.

- *The bone highlighting task.* This task (filled with an orange round and the area marked "Grad-Cam image" in Fig. 1) looks at the output of the previous block. Then, it applies Grad-CAM to produce images in which the bones are highlighted with a general architecture illustrated in Fig. 5. In addition, this task aims at rendering a 3D model of the bones in the next block.
- *The 3D visualization task.* Deploying ITK and VTK library, this task takes CT images of a body part with the bones highlighted from the previous blocks and visualizes them into a 3D model. It provides a more intuitive way to easily view CT images to discover abnormalities.

Details of such tasks are presented in the following sections. First, section III-A describes the data used in the experiments. Then, shallow convolutional neural networks for body parts classification will be introduced in Section III-C. Next, the proposed techniques for highlighting the bones in CT images are presented in Section III-D. Finally, we illustrate the 3D visualization method in Section III-E.

A. Data Preparation

In this study, our experiments were conducted on the Visible Human Project Dataset collected by [32]. This dataset represents the average human body exhibited in CT images of the male and female bodies with the section interval of one millimeter and one-third of a millimeter, respectively. The data were stored in DICOM format and downloaded for general research use. Each includes DICOM files, and each contains a grey CT image of size 512×512, and patient data are categorized in folders by the body part. Several folders and files of each dataset used in the study are presented in Table I with 479, 800, 290, and 910 images of 512×512 for parts of head, hip, pelvis, and shoulder, respectively.

The dataset mentioned earlier consists of the Visible Female and the Visible Male CT datasets. Slices from the female patient were fetched into the training phase, while we used slices from the male patient for testing, as revealed in Table I.

TABLE I
THE SAMPLES OF THE CONSIDERED DATASETS

Dataset	Labels				Total	Usage
	Head	Hip	Pelvis	Shoulder		
Visible Female CT	234	400	150	450	1234	Training
Visible Male CT	245	400	140	460	1245	Testing
Total	479	800	290	910	2479	

B. Data Pre-processing Steps and Data Augmentation

After downloading, we extracted image data of all four body parts from DICOM files to four corresponding labels. Then, augmentation techniques, including rotation, shifting, shearing, zooming, and flipping, were leveraged to increase the training data to 7404 images used in the training phase.

We separated each 8-bit training image into a bone and a non-bone image. The bone images can be obtained by setting grey pixel values equal to or greater than 150 (as conducted by [33]) and assigning zero to other pixel values as non-bone images. Similarly, the pixels with less than 150 were used to generate a non-bone image. Those images were then augmented to increase the number of training images. There are two subsets

in the dataset that we used. One includes CT images taken from a male cadaver and the other from a female cadaver. In order to have an objective evaluation of the performance of the models, we decided to use one subset for training and the other for testing. The high testing accuracies reported in our study have proven the use of the two subsets.

C. Shallow Convolutional Neural Networks for CT Image Classification Tasks

We investigate the effects of the stack of a convolutional layer, a ReLU, and a max pooling layer on image classification performance. CNN-1 contains a convolutional layer consisting of 32 filters of 3×3 with a stride of 1, followed by a leaky ReLU and a max pooling layer of 2×2 (stride of 2), while CNN-2 includes such a stack and added a similar stack but with 64 filters. CNN-3, besides two stacks as CNN-2, is added a more similar stack with 128 filters. Next, the output from such stacks of all three architectures is flattened, then fetched into a neural layer with 128 units, followed by a ReLU, and provides two outputs corresponding to bone detection tasks and four outputs for human body part classification. All networks receive an image input with the size of 512×512.

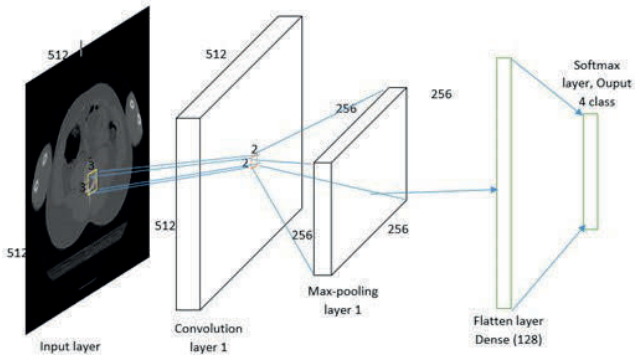


Fig. 2. The proposed CNN architecture for body part classification task.

Figure 2 illustrates the classification of CNN-1 for human body parts with four outputs. First, the network receives input as grey images of size 512×512. Then the image is passed through a convolutional layer with 32 3×3 filters, and the convolution stride is one pixel. The spatial padding of this layer is set to preserve the image size when the convolution is complete. Specifically, we add 1-pixel zero-padding to the input image before convolving with 3×3 filters. The activation function used here is Relu. Next, the image size is halved (reduced to 256×256) as it passes through the max-pooling layer with a 2×2 window, and the stride is 2 pixels. Two fully connected layers are following those layers. The first one has 128 neurons, followed by a ReLu. The second one uses softmax as the activation function and has four neurons corresponding to the four data classes. For parameters for compiling the models, we use categorical cross-entropy as the loss function and Adam as the optimizer.

It is similar to the two and three-stack architecture, except that they will have the corresponding number of stacks of convolutional (3×3 filters) and max-pooling (2×2 kernel) layer blocks. The number of convolution filters in each block doubles the one preceding it, beginning from 32.

The training process is illustrated in Fig. 3. First, training images are loaded using the pydicom library, and the pixel array attribute of each DICOM file is fed to the model. Each image is labelled from 0 to 3, representing images of the head, hip, pelvis, and shoulder. Then we train the model for 50 epochs.

Body part image classification: Figure 4 describes the classification process highlighted in the red box. We use the most accurate model to classify body parts images.

We put 1258 testing images into an input folder. Those images are used to evaluate the model accuracy and are entirely different from those used to train the model.

- The input images will be converted into a pixel array, forming a matrix array, and fed into our trained model.
- The feature extraction from this matrix array is similar to the model training process. It will be converted to a single array to map features to the output layer at the fully connected layer. This layer will have a voting process to determine which body part those features belong to based on the features the trained model has learned. The part of the most voted body will have the highest probability and become the final output.

The bone detection: We do the experiments on the same architectures of CNN as in section III-C, except that the number of neurons of the output layer is two. From 1284 original training images, augmentation methods enhance training data to 7404 images. Next, based on the value of each pixel in grayscale (0 to 255), we separate bones (the pixels with a value greater than 150) and the remaining of each image into two images. They will be assigned a label corresponding to 0 to 1 representing bones and the remaining. Then the training process is conducted similarly to training the body part classification models.

D. Highlighting the Bones in CT Image Using Grad-CAM

Grad-CAM proposed by [34] stands for Gradient-weighted Class Activation Mapping, an algorithm allowing us to visualize an object in any image through a model trained to identify that object. Grad-CAM uses the final convolution layer in a pre-trained model combined with the gradient in this layer to detect an object in the image. The Grad-CAM output is a visualized heat map for the label we set up for the object of interest in the image. Combining the Grad-CAM output with the original image makes the object stand out. Grad-CAM is applied in many recent studies with numerous practical applications such as studies of [35]–[40].

Deploying Grad-CAM technique for highlighting the bones in the body part images: In order to accurately highlight the bones in CT images, we choose the most accurate bone and non-bone classification model for the Grad-CAM algorithm. Next, we combine the original images with the model final convolution layer to compute gradients and generate heat map images. Finally, the original images and their heat map images are combined to generate Grad-CAM images with the bones highlighted.

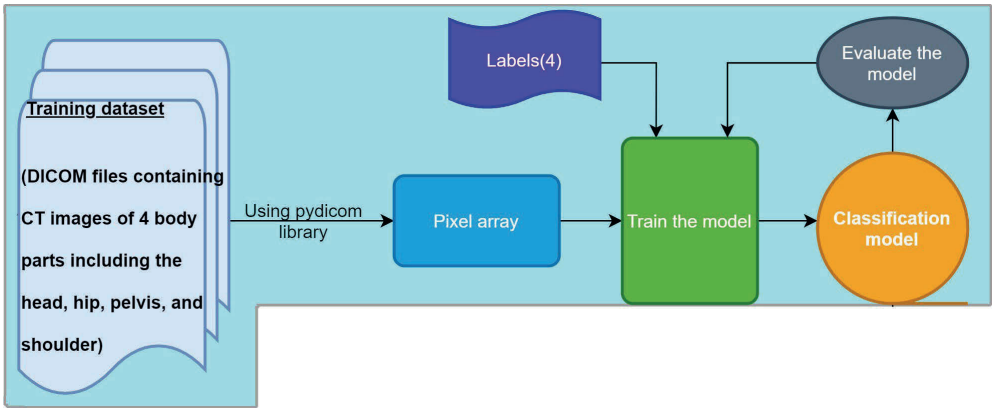


Fig. 3. The flowchart for body part classification tasks.

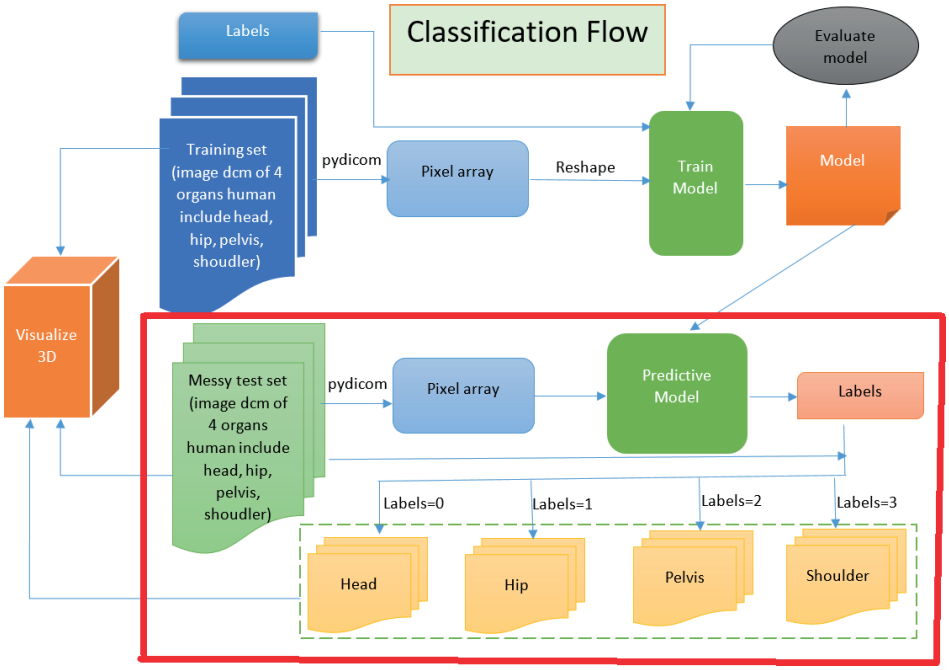


Fig. 4. The classification combining visualization tasks.

The process of creating a heat map is as follows:

- We first create an input image mapping model of the final convolutional layer of the model.
- Next, we build a classification model from the last transition layer named "max_pooling2d_11" and class "flatten_2" (the class that converts the feature matrix array to a single array, making it easier to map the final output layer of the model).
- Then, we compute the gradient to detect the bone in the input image: The input image is converted as a pixel array and mapped into the model created in step one. The pixel array values are mapped from the model on the current state as a matrix array, so the flatten_2 class will convert them to single arrays for easy computation. Next, we use the same estimation method $y' = f(w, x) + bias$ to the classification model. The classification model will calculate the function gradient y' to predict the image expected values for the bone class. In this way, we obtain an array of feature pixels roughly equal to the actual pixel value of the bone class features in the image.
- Finally, the array of pixels above will be used with the average pooling method, whose output is multiplied by the image pixel values mapped in step one, creating a heat map that highlights the typical values of the bone class.

E. 3D Visualization from 2D images of the Body Parts

Toolkits for 3D visualization: In this research, we used two main libraries, namely Insight Toolkit[2], introduced by [5] to load DICOM images: slice images of a particular patient's body part for displaying and visualizing them into a 3D model that can be viewed from any angle.

Insight Segmentation and Registration Toolkit (ITK)[3] is an open-source, cross-platform software package. ITK is used in

[2] https://www.insight-journal.org/browse/publication/923, accessed on 20 May 2021

[3] https://itk.org/, accessed on 20 May 2021

image segmentation and registration applications, while image segmentation is data classification from visual image datasets, usually from medical devices such as CT or MRI. Registration image is the arrangement of corresponding data together.

VTK (Visualization Toolkit) [4] is a WebGL-based visualization library to visualize images into 3D models in a web browser. It has become the most popular tool for image visualization purposes. VTK supports many different algorithms, including scalar method, vector, tensor, cut, reduce polygon, etc.

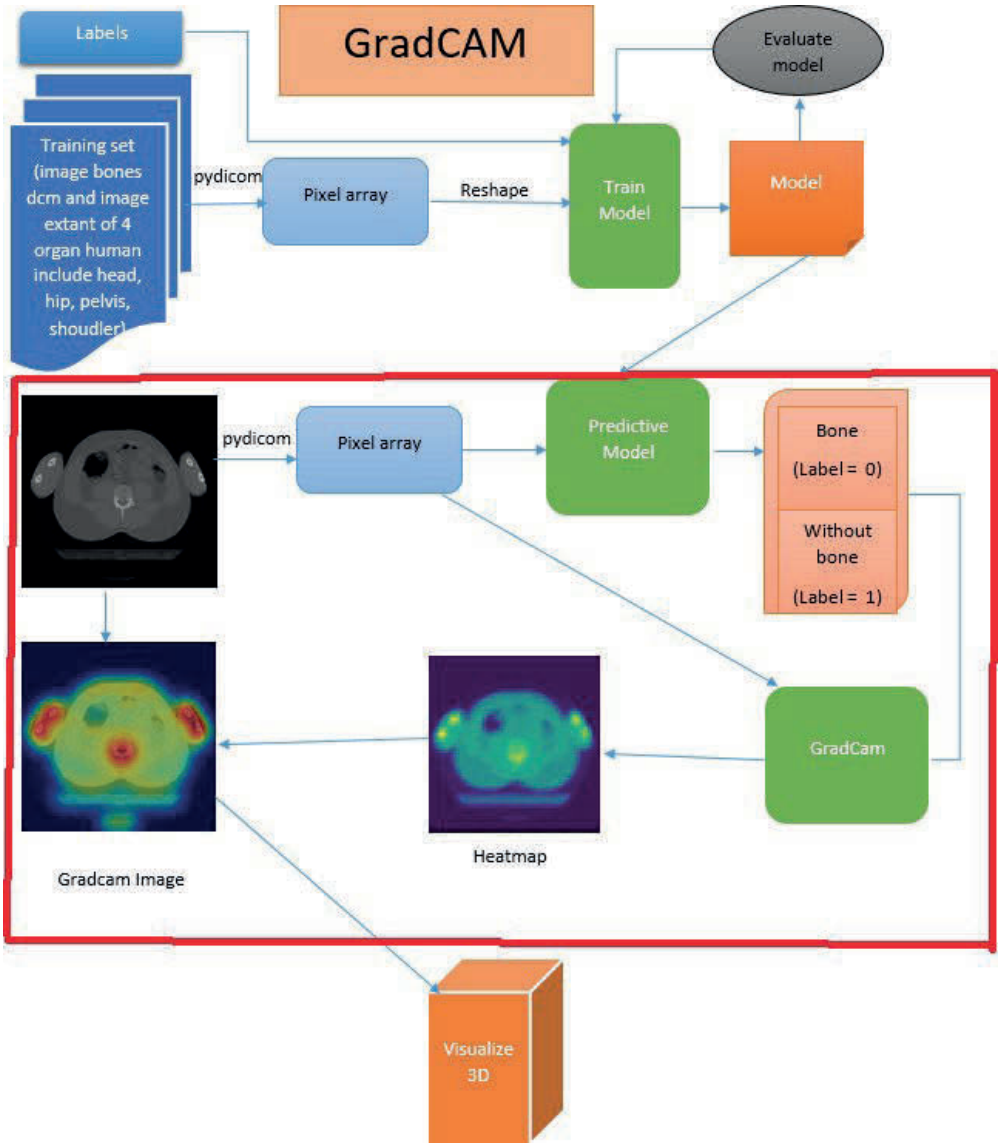


Fig. 5. The Grad-CAM flowchart for highlighting bone.

Steps of the 3D Visualization Process: To visualize 2D images as a 3D model, we use the famous medical imaging 3D visualization libraries such as ITK.js and VTK.js. ITK/VTK viewer is a great software system for medical and scientific image, mesh, and point set visualization. Also, this is an open-source web-based viewer for 2D/3D images. We can contribute and discuss bugs with the authors.

We focus on handling the Dicom (.dcm) image as a slice. The step-by-step visualization from 2D to 3D:

- Step 1: Choose the folder that contains 2D medical image slices.
- Step 2: Use the ITK.js library to load DICOM files and read information.
- Step 3: We utilize the VTK library to visualize 2D/3D images based on that information.

Combining 2D slice images with a 3D model makes it easier to observe abnormalities. Moreover, the generated 3D model can be sliced and viewed along the x, y, and z axes, making finding the abnormality position clearer.

[4] https://vtk.org/, accessed on 20 May 2021

IV. Results

A. Metrics for Comparison

We use a confusion matrix for model evaluation to compute the overall accuracy. Based on this method, the overall rating of the classifiers is observed objectively and clearly. The common metrics are:

- *Accuracy*: This allows us to evaluate the model prediction performance on a dataset. The higher the accuracy, the more accurate the model is; it is less than or equal to 1.
- *Precision*: The number of positive predicted cases is confirmed. The higher the precision, the better our model classifies the BAD profile (BAD is the positive group).
- *Recall*: It measures the rate of accurate prediction of positive cases on all positive samples.
- *F1 score*: It is the balance between the precision and recall values. It, therefore, concurrently represents the evaluation of precision on both the precision and recall values.

Precision and Recall are both important metrics. Due to the inconvenience in choosing between the Recall and Precision values to evaluate the model, the F1-score value can be an appropriate choice (1).

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (1)$$

The training phase can be very time-consuming. Therefore, grey images of size 512×512 were used in the experiments, with inference time taking about 3 hours. We trained and tested this system on a Google Colab machine with 12GB of memory.

B. Comparison of Various Shallow Architectures

We have deployed three various shallow architectures for comparison. The first architecture (CNN-1) includes a stack of 32 filters with the size of 3×3 using a stride of 1, followed by a max pooling of 2×2 with stride 2, using Leaky ReLU proposed by [41] after each convolution. CNN-2 and CNN-3 include two stacks and three stacks, respectively.

TABLE II

Performance Comparison of Various Architectures on the Bone Detection Task

		Precision	Recall	F1-score
CNN-1	Bone	1.00	0.97	0.99
	Non-Bone	0.97	1.00	0.99
	Average	0.99	0.99	0.99
CNN-2	Bone	1.00	0.42	0.60
	Non-Bone	0.63	1.00	0.78
	Average	0.82	0.71	0.69
CNN-3	Bone	1.00	0.97	0.98
	Non-Bone	0.97	1.00	0.99
	Average	0.99	0.98	0.98

Table II and Fig. 6 exhibit accuracy comparison among those architectures. As observed, CNN-1 and CNN-3 reveal good performance with precision, recall, and f1-score values not less than 0.97 for both bone and non-bone classes, and their overall accuracy is 0.99 and 0.98, respectively. In contrast, the model using two stacks achieves poor results. The recall value for class bone is just 0.42, the precision for class non-bone is 0.63, and the overall accuracy is 0.71.

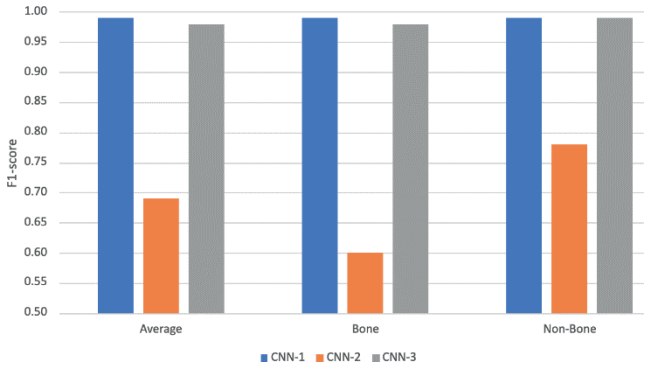


Fig. 6. Performance comparison of various shallow convolutional neural network architectures.

C. Results of Human Body Part Classification

To evaluate the models, we use a dataset of 1258 images of four body parts, including the head, hip, pelvis, and shoulder, with the best architecture, CNN-1, selected from results shown in Table II. The overall accuracy is 98 % and is shown in Table III. Generally, all metrics are not less than 0.96. In terms of precision, class head and shoulder have the maximum value of 1.00, while the case for recall and F1-score metrics belongs to class hip and head, respectively.

TABLE III

The Evaluation of the Proposed Body Part Classification CNN

	Precision	Recall	F1-score
Head	1.00	0.99	1.00
Hip	0.97	1.00	0.98
Pelvis	0.96	0.99	0.97
Shoulder	1.00	0.97	0.98
Macro Average	0.98	0.99	0.98
Weighted Average	0.98	0.98	0.98

Table IV shows the classification results for each class. The class hip has the best results, with 400 out of 400 images classified correctly. Class head also has pretty good results, with 243 out of 245 images classified. About 3 % of the class shoulder images are misclassified, while class pelvis has about nearly 14 % of images classified to other classes. The training section is always time-consuming. We used our CNN model with grey images of size 512×512 as the input in this research. The training section finished after 3 hours. We trained and tested this system on a Google Colab machine with 12GB of memory.

TABLE IV

The Results in Confusion Matrix of Human Body Part Classification

		Predicted class			
		Head	**Hip**	**Pelvis**	**Shoulder**
Actual class	Head	243	2	0	0
	Hip	0	400	0	0
	Pelvis	0	2	138	20
	Shoulder	0	10	6	457

We also obtain consistent results in comparison with various architectures as Section IV-B. Table V exhibits the classification performance on four considered body parts. As observed, CNN-1 can reach 0.995 in head recognition, which is higher than CNN-2 with 0.990, and CNN-3 with only 0.982.

For detecting hip, CNN-1 achieves an accuracy of 0.985, while CNN-2 and CNN-3 obtain 0.773 and 0.918, respectively. Moreover, CNN-1 gets the best score on Pelvis detection with 0.974, while the worst obtains 0.914. In overall performance, CNN-1 performs the best, while CNN-3 exhibits the lowest with 0.870. These results are similar to the bone detection task.

TABLE V
THE COMPARISON OF THREE ARCHITECTURES ON HUMAN BODY PART CLASSIFICATION

	Head	Hip	Pelvis	Shoulder	Average
CNN-1	0.995	0.985	0.974	0.984	0.980
CNN-2	0.990	0.773	0.918	0.870	0.870
CNN-3	0.982	0.918	0.914	0.928	0.930

D. Visualization with Grad-CAM

An example of the heat map (above) and Grad-CAM images (below) of the models with the architectures of CNN-1, CNN-2, and CNN-3 are shown in Fig. 7 from left to right, and visualization of various parts in Fig. 8. The model that uses three convolutional layers gives the best Grad-CAM image in which the bones are highlighted. Hence, it generates Grad-CAM images for the 3D visualization experiment.

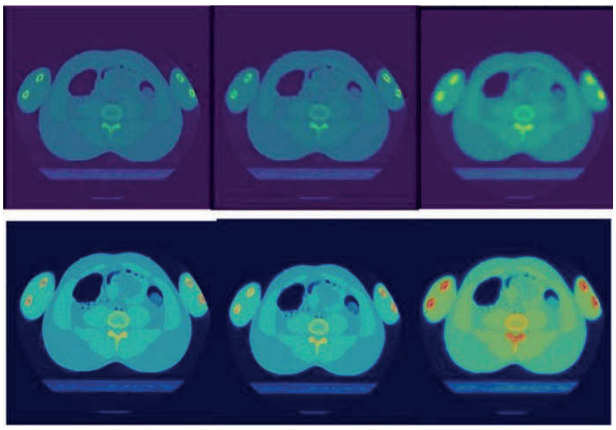


Fig. 7. Heat map and Grad-CAM image of CNN-1, CNN-2, and CNN-3 from left to right.

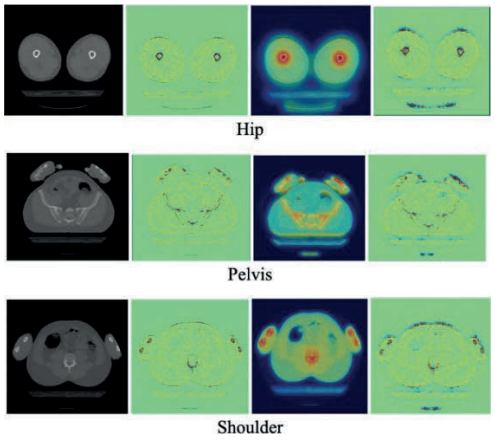


Fig. 8. Grad-CAM highlighted bone areas in various human body parts.

E. 3D Visualization Results

We built a web application for visualization tasks transforming and combining 2D slices of a body part into a 3D visualization (on the upper right). Moreover, the visualization can be observed from any perspective by clicking, dragging, and zooming in and out, as shown in Fig. 9 and visualized in 3D with Grad-CAM in Fig. 10.

Moreover, we can view the 3D model slices along the X (from left to right and left to right), Y (from front to back and back to front), and Z (from top to bottom and bottom to top) axes by adjusting their position on the three bars on the upper-left. Those three slices along the X, Y, and Z axes are displayed on the bottom, respectively, from left to right. Moreover, the 3D model will be updated as its position is changed.

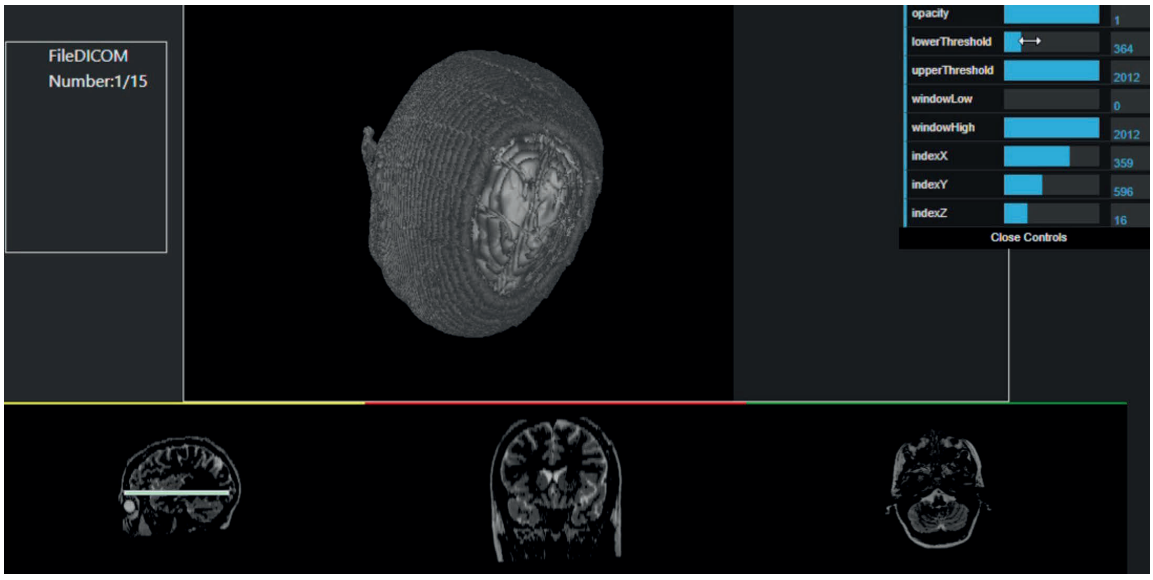


Fig. 9. Visualization in 3D of slides capturing a head.

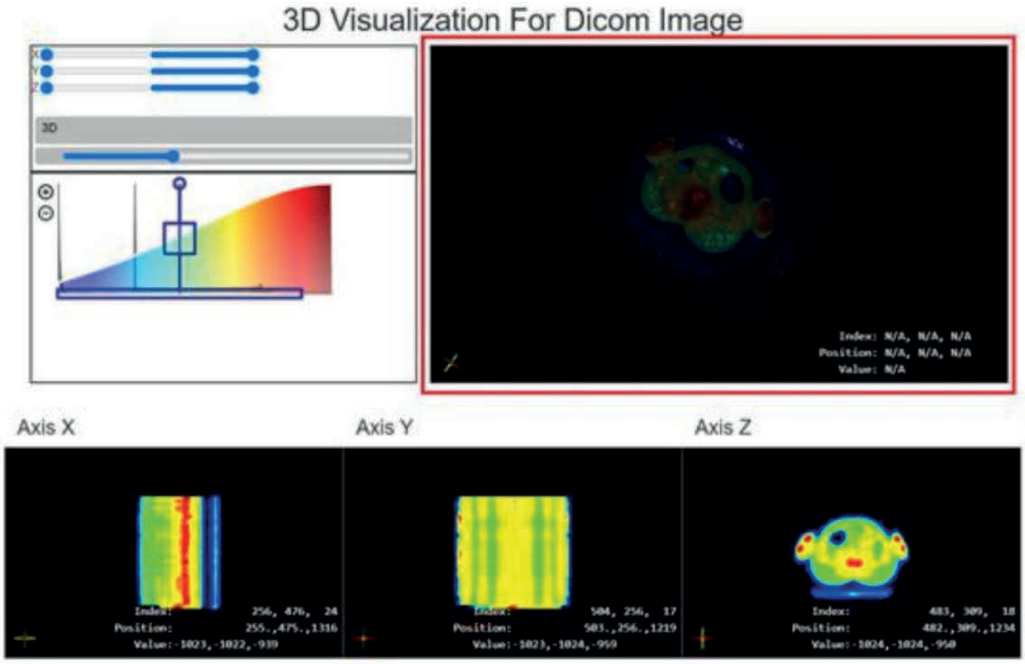


Fig. 10. Visualization in 3D with Grad-Cam.

V. Conclusion

We have presented the CNN model to classify CT images of body parts, including the head, shoulder, hip, and pelvis, to detect and segment bones in the medical slices shaped in CT images and visualize them in 3D. As shown by the results, we can accurately categorize messy images of body parts into corresponding parts. However, it can be seen that the performance of CNN-1 is better than other architectures. The architecture of the most accurate is with one stack of a convolutional layer, a ReLU, followed by a max pooling layer, and a neural layer with 128 neurons can obtain an accuracy of 98 % in human body part classification. Therefore, we found that the difference between four considered human body parts, including the head, pelvis, hip, and shoulder, when captured in CT images between a woman and a man was trivial. However, the pelvis can be the most misclassed part compared to the others. We also presented highlighting the bones in CT images of the body parts. The inputs use grayscale images, and after using the Grad-CAM algorithm, their colours are in RGB format. The red-shaded areas represent the bones, the yellow areas are just the gradients, and the blue ones represent non-bone areas. That means the model has correctly marked the bones in the image, creating a heat map highlighting the bone areas in that image. This makes it easier to detect the bones of the body parts.

Furthermore, instead of viewing each CT image of a body part separately, we have proposed a method using the Insight Toolkit and the Visualization Toolkit libraries to visualize them into an interactive 3D model making it more intuitive and clear to detect abnormalities inside the patient's body. In education, the proposal in this study provides a tool for medical students to be more convenient in classifying and identifying CT images of human body parts and accurately detecting the bones in those images. In addition, our approach is expected to help doctors improve their diagnostic results and aid in developing generations of autonomous robots in surgery in real life.

For further research, we can collect more data for deep analysis on comparing the difference between human body parts and male and female bone. Furthermore, we can focus on combining famous segmentation algorithms and model explanation techniques to improve bone detection and visualization performance.

Declarations

Availability of Data, Code, and Material

The data supporting this study findings are available by [32].

Conflict of Interest statement

All authors declare that they have no conflicts of interest.

References

[1] X. Wang, J. Yu, Q. Zhu, S. Li, Z. Zhao, B. Yang, and J. Pu, "Potential of deep learning in assessing pneumoconiosis depicted on digital chest radiography," *Occupational and Environmental Medicine*, vol. 77, no. 9, pp. 597–602, 2020. https://doi.org/10.1136/oemed-2019-106386

[2] H. Tang and Z. Hu, "Research on medical image classification based on machine learning," *IEEE Access*, vol. 8, pp. 93145–93154, 2020. https://doi.org/10.1109/ACCESS.2020.2993887

[3] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Jan. 2017. https://doi.org/10.1038/nature21056

[4] L. Song, T. Xing, Z. Zhu, W. Han, G. Fan, J. Li, H. Du, W. Song, Z. Jin, and G. Zhang, "Hybrid clinical-radiomics model for precisely predicting the invasiveness of lung adenocarcinoma manifesting as pure ground-glass nodule," *Academic Radiology*, vol. 28, no. 9, Sep. 2021. https://doi.org/10.1016/j.acra.2020.05.004

[5] L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide and the Insight Software Consortium: updated for ITK version 2.4*. Erscheinungsort nicht ermittelbar: Kitware Inc, 2005. https://www.igb.illinois.edu/sites/default/files/upload/core/PDF/ItkSoftwareGuide-2.4.0.pdf

[6] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, Feb. 2020. https://doi.org/10.1007/s11263-019-01228-7

[7] A. A. Giannopoulos and T. Pietila, "Post-processing of DICOM Images," in *3D Printing in Medicine: A Practical Guide for Medical Professionals*, F. J. Rybicki and G. T. Grant, Eds. Cham: Springer International Publishing, 2017, pp. 23–34. https://doi.org/10.1007/978-3-319-61924-8_3

[8] C. Li, R. Huang, Z. Ding, C. Gatenby, D. Metaxas, and J. Gore, "A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 20, pp. 2007–2016, Jul. 2011. https://doi.org/10.1109/TIP.2011.2146190

[9] Y.-L. Ling &. A.-H. Zhong, "Research on the classification of body type and prototype of middle-aged women based on 3D scanning," *Journal of Fiber Bioengineering and Informatics*, vol. 13, no. 3, pp. 161–167, Nov. 2020. https://doi.org/10.3993/jfbim00344

[10] C. Hellmann, A. Bajrami, and W. Kraus, "Enhancing a robot gripper with haptic perception for risk mitigation in physical human robot interaction," in *2019 IEEE World Haptics Conference (WHC)*, Tokyo, Japan, Jul. 2019, pp. 253–258. https://doi.org/10.1109/whc.2019.8816109

[11] F. Li, "Classification of students' body shape based on deep neural network," in *Innovative Computing.Lecture Notes in Electrical Engineering*, C.T. Yang, Y. Pei, and J.W. Chang, Eds., vol. 675. Springer Singapore, 2020, pp. 549–557. https://doi.org/10.1007/978-981-15-5959-4_66

[12] J. F. Yu, L. Pung, H. Minami, K. Mueller, R. Khangura, R. Darflinger, S. W. Hetts, and D. L. Cooke, "Virtual 2D angiography from four-dimensional digital subtraction angiography (4D-DSA): A feasibility study," *Interv. Neuroradiol.*, vol. 27, no. 2, Sep. 2020. https://doi.org/10.1177/1591019920961604

[13] M. Boussif, N. Aloui, and A. Cherif, "DICOM imaging watermarking for hiding medical reports," *Medical and Biological Engineering and Computing*, vol. 58, no. 11, pp. 2905–2918, Sep. 2020. https://doi.org/10.1007/s11517-020-02269-8

[14] X. Jiang, Y. Zhang, Q. Yang, B. Deng, and H. Wang, "Millimeter-wave array radar-based human gait recognition using multi-channel three-dimensional convolutional neural network," *Sensors*, vol. 20, no. 19, Sep. 2020, Art. no. 5466. https://doi.org/10.3390/s20195466

[15] N. S. Chan, K. I. Chan, R. Tse, S.-K. Tang, and G. Pau, "ReSPEcT: privacy respecting thermal-based specific person recognition," in *Thirteenth International Conference on Digital Image Processing (ICDIP 2021)*, X. Jiang and H. Fujita, Eds., SPIE, Jun. 2021. https://doi.org/10.1117/12.2599271

[16] E. Sadek, N. A. Seada, and S. Ghoniemy, "Computer vision techniques for autism symptoms detection and recognition: A survey," *International Journal of Intelligent Computing and Information Sciences*, vol. 20, no. 2, pp. 89–111, Dec. 2020. https://doi.org/10.21608/ijicis.2020.46360.1034

[17] F. Zhou, H. Zhao, and Z. Nie, "Safety helmet detection based on YOLOv5," in *2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA)*, Shenyang, China, Jan. 2021, pp. 6–11. https://doi.org/10.1109/ICPECA51329.2021.9362711

[18] A. Ashraf, T. S. Gunawan, F. D. A. Rahman, and M. Kartiwi, "A summarization of image and video databases for emotion recognition," in *Recent Trends in Mechatronics Towards Industry 4.0. Lecture Notes in Electrical Engineering*, A.F. Ab. Nasir *et al*., Eds., vol. 730. Springer, Singapore, Jul. 2021, pp. 669–680. https://doi.org/10.1007/978-981-33-4597-3_60

[19] D. A. Clunie, "DICOM format and protocol standardization – a core requirement for digital pathology success," *Toxicologic Pathology*, vol. 49, no. 4, Oct. 2020. https://doi.org/10.1177/0192623320965893

[20] G. Kwon, J. Ryu, J. Oh, J. Lim, B.-k. Kang, C. Ahn, J. Bae, and D. K. Lee, "Deep learning algorithms for detecting and visualising intussusception on plain abdominal radiography in children: a retrospective multicenter study," *Scientific Reports*, vol. 10, no. 1, Oct. 2020, Art. no. 17582. https://doi.org/10.1038/s41598-020-74653-1

[21] I. Lavdas, B. Glocker, D. Rueckert, S. Taylor, E. Aboagye, and A. Rockall, "Machine learning in whole-body MRI: experiences and challenges from an applied study using multicentre data," *Clinical Radiology*, vol. 74, no. 5, pp. 346–356, Feb. 2019. https://doi.org/10.1016/j.crad.2019.01.012

[22] G. Li, X. Shen, J. Li, and J. Wang, "Diagonal-kernel convolutional neural networks for image classification," *Digital Signal Processing*, vol. 108, Jan. 2021, Art. no. 102898. https://doi.org/10.1016/j.dsp.2020.102898

[23] H. Barzekar and Z. Yu, "C-Net: A reliable convolutional neural network for biomedical image classification," *arXiv preprint,* arXiv:2011.00081, 2020. https://arxiv.org/pdf/2011.00081.pdf

[24] G. Jia, X. Huang, S. Tao, X. Zhang, Y. Zhao, H. Wang, J. He, J. Hao, B. Liu, J. Zhou, T. Li, X. Zhang, and J. Gao, "Artificial intelligence-based medical image segmentation for 3D printing and naked eye 3D visualization," *Intelligent Medicine*, vol. 2, no. 1, pp. 48–52, Feb. 2022. https://doi.org/10.1016/j.imed.2021.04.001

[25] Q. Duan, G. Wang, R. Wang, C. Fu, X. Li, M. Gong, X. Liu, Q. Xia, X. Huang, Z. Hu, N. Huang, and S. Zhang, "SenseCare: A research platform for medical image informatics and interactive 3D visualization," *ArXiv*, vol. abs/2004.07031, 2020. arxiv.org/pdf/2004.07031.pdf

[26] L. Cai, T. Long, Y. Dai, and Y. Huang, "Mask R-CNN-based detection and segmentation for pulmonary nodule 3d visualization diagnosis," *IEEE Access*, vol. 8, pp. 44400–44409, Feb. 2020. https://doi.org/10.1109/ACCESS.2020.2976432

[27] S. AlZu'bi, M. Shehab, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Parallel implementation for 3D medical volume fuzzy segmentation," *Pattern Recognition Letters*, vol. 130, pp. 312–318, Feb. 2020. https://doi.org/10.1016/j.patrec.2018.07.026

[28] D. Mitsouras, P. C. Liacouras, N. Wake, and F. J. Rybicki, "RadioGraphics update: Medical 3D printing for the radiologist," *RadioGraphics*, vol. 40, no. 4, pp. E21–E23, Jul. 2020. https://doi.org/10.1148/rg.2020190217

[29] M. Javaid and A. Haleem, "Virtual reality applications toward medical field," *Clinical Epidemiology and Global Health*, vol. 8, no. 2, pp. 600–605, Jun. 2020. https://doi.org/10.1016/j.cegh.2019.12.010

[30] X. Zhou, T. Hara, H. Fujita, Y. Ida, K. Katada, and K. Matsumoto, "Extraction and recognition of the thoracic organs based on 3D CT images and its application," in *CARS 2002 Computer Assisted Radiology and Surgery*, H.U. Lemke *et al*., Eds. Springer, Berlin, Heidelberg, 2002, pp. 776–781. https://doi.org/10.1007/978-3-642-56168-9_130

[31] M. S. M. Rahim, A. Norouzi, A. Rehman, and T. Saba, "3D bones segmentation based on CT images visualization," *Biomedical Research,* vol. 28, no. 8, pp. 3641–3644, 2017. https://www.researchgate.net/publication/317745097_3D_bones_segmentation_based_on_CT_images_visualization

[32] M. Ackerman, "The visible human project," *Proceedings of the IEEE*, vol. 86, no. 3, pp. 504–511, Mar. 1998. https://doi.org/10.1109/5.662875

[33] L. Friedli, D. Kloukos, G. Kanavakis, D. Halazonetis, and N. Gkantidis, "The effect of threshold level on bone segmentation of cranial base structures from CT and CBCT images," *Scientific Reports*, vol. 10, no. 1, Apr. 2020. https://doi.org/10.1038/s41598- 020-64383-9

[34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-Cam: Visual explanations from deep networks via gradient-based localization," *International Journal of Computer Vision*, vol. 128, no. 2, p. 336–359, Oct. 2019. http://doi.org/10.1007/s11263-019-01228-7

[35] H. Panwar, P. Gupta, M. K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, and V. Singh, "A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images," *Chaos, Solitons & Fractals*, vol. 140, Nov. 2020, Art. no. 110190. https://doi.org/10.1016/j.chaos.2020.110190

[36] R. Fu, Q. Hu, X. Dong, Y. Guo, Y. Gao, and B. Li, "Axiom-based Grad-CAM: Towards accurate visualization and explanation of CNNs," *arXiv*, vol. 2008.02312, 2020. https://arxiv.org/pdf/2008.02312.pdf

[37] P. Morbidelli, D. Carrera, B. Rossi, P. Fragneto, and G. Boracchi, "Augmented Grad-CAM: Heat-maps super resolution through augmentation," in *ICASSP 2020 – 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, May 2020, pp. 4067–4071. https://doi.org/10.1109/ICASSP40776.2020.9054416

[38] H. Jiang, J. Xu, R. Shi, K. Yang, D. Zhang, M. Gao, H. Ma, and W. Qian, "A multi-label deep learning model with interpretable Grad-CAM for diabetic retinopathy classification," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Montreal, QC, Canada, Jul. 2020, pp. 1560–1563. https://doi.org/10.1109/EMBC44109.2020.9175884
[39] Y. Zhang, D. Hong, D. McClement, O. Oladosu, G. Pridham, and G. Slaney, "Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging," *Journal of Neuroscience Methods*, vol. 353, Apr. 2021, Art. no. 109098. https://doi.org/10.1016/j.jneumeth.2021.109098
[40] J. Choi, J. H. Choi, and W. Rhee, "Interpreting neural ranking models using Grad-CAM," *ArXiv*, vol. abs/2005.05768, 2020. arxiv.org/pdf/2005.05768.pdf
[41] J. Xu, Z. Li, B. Du, M. Zhang, and J. Liu, "Reluplex made more practical: Leaky ReLU," in *2020 IEEE Symposium on Computers and Communications (ISCC)*, Rennes, France,, Jul. 2020, pp. 1–7. https://doi.org/10.1109/iscc50000.2020.9219587



Hai Thanh Nguyen is a Lecturer of the College of Information and Communication Technology, Can Tho University, Vietnam. He received his Master's degree in Computer Science and Engineering from National Chiao Tung University, Taiwan, in 2014, and his PhD degree in Computer Science from Sorbonne University, France, in 2018. His current research includes bioinformatics, healthcare systems, recommendation systems, and applications using computer vision, machine learning, and data analysis.
E-mail: nthai.cit@ctu.edu.vn
ORCID iD: https://orcid.org/0000-0002-1386-1390



My N. Nguyen is a Lecturer at Can Tho University. He received his B.S. degree in Computer Science from Can Tho University, Vietnam, in 2010, and the M.Sc. degree in Electrical Engineering and Computer Science from National Chiao Tung University, Taiwan, in 2015. He is currently a PhD student at Kyoto Institute of Technology, Japan. His research interests include medical image processing, IoT and bioinformatics.
E-mail: nnmy@cit.ctu.edu.vn



Bang Anh Nguyen was a student of the College of Information and Communication Technology, Can Tho University, Vietnam. He received his Bachelor degree in Information Technologies in 2021. Currently, he is a Fullstack Developer at Hybrid Technologies. His research interests include medical data analysis and applications in machine learning.
Email: bang98.2016@gmail.com

Linh Chi Nguyen was a student of the College of Information and Communication Technology, Can Tho University, Vietnam. His research relates to intelligent image processing and deep learning for healthcare systems.
Email: nchilinhith@gmail.com



Duong Linh Phung received his Bachelor degree in Computer Science at the College of Information and Communication Technology, Can Tho University, Vietnam, in 2021. Currently, he is a Master student at the College of Information Science and Engineering, Ritsumeikan University, Japan. His research relates to intelligent image processing and deep learning.
Email: linhphung15999@gmail.com