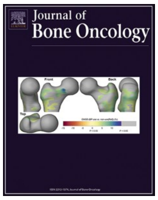
J BONE ONCOL 44 (2024) 100520

Contents lists available at ScienceDirect

Journal of Bone Oncology

journal homepage: www.elsevier.com/locate/jbo



GHA-DenseNet prediction and diagnosis of malignancy in femoral bone tumors using magnetic resonance imaging

Y. Yuguang [a,b,c], Yusi Chen [a,b,c], Daxin Zhu [a,b,c], Yifeng Huang [d], Ying Huang [d], Xiadong Li [e,*], Jianbing Xiahou [a,b,c,*]

[a] *Faculty of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou 362000, China*
[b] *Fujian Provincial Key Laboratory of Data Intensive Computing, Quanzhou 362000, China*
[c] *Key Laboratory of Intelligent Computing and Information Processing, Fujian Province University, Quanzhou 362000, China*
[d] *Department of Diagnostic Radiology, Huaqiao University Affiliated Strait Hospital, Quanzhou, Fujian 362000, China*
[e] *Department of Radiation Oncology, Children's Hospital of Zhejiang University School of Medicine, Zhejiang University, Hangzhou 310052, China*

HIGHLIGHTS

- The study uses specialized pre-trained CNN models for accurate classification of T1 and T2 MRI bone tumor images.
- GHA-DenseNet is a novel variant of the DenseNet architecture that improves malignancy classification particularly when dealing with a limited number of MRI samples.
- The bone tumor prediction model achieves over 80% accuracy by combining classifier outputs with patient-specific data.
- Overfitting issues are addressed through connectivity pruning and dropout methods due to the small dataset.
- Future research directions and improvements are suggested, reflecting a commitment to progress in the bone oncology field.

ARTICLE INFO

Keywords:
Deep convolutional network model
Attention module
Magnetic resonance imaging
Bone tumour
GHA-DenseNet

ABSTRACT

Background and objective: Due to their aggressive nature and poor prognosis, malignant femoral bone tumors present considerable hurdles. Early treatment commencement is essential for enhancing vital and practical outcomes. In this investigation, deep learning algorithms will be used to analyze magnetic resonance imaging (MRI) data to identify bone tumors that are malignant.

Methodology: The study cohort included 44 patients, with ages ranging from 17 to 78 (22 women and 22 males). To categorize T1 and T2 weighted MRI data, this paper presents an improved DenseNet network model for the classification of bone tumor MRI, which is named GHA-DenseNet. Based on the original DenseNet model, the attention module is added to solve the problem that the deep convolutional model can reduce the loss of key features when capturing the location and content information of femoral bone tumor tissue due to the limitation of local receptive field. In addition, the sparse connection mode is used to prune the connection mode of the original model, so as to remove unnecessary and retain more useful fast connection mode, and alleviate the overfitting problem caused by small dataset size and image characteristics. In a clinical model designed to anticipate tumor malignancy, the utilization of T1 and T2 classifier output values, in combination with patient-specific clinical information, was a crucial component.

Results: The T1 classifier's accuracy during the training phase was 92.88% whereas the T2 classifier's accuracy was 87.03%. Both classifiers demonstrated accuracy of 95.24% throughout the validation phase. During training and validation, the clinical model's accuracy was 82.17% and 81.51%, respectively. The clinical model's receiver operating characteristic (ROC) curve demonstrated its capacity to separate classes.

Conclusions: The proposed method does not require manual segmentation of MRI scans because it makes use of pretrained deep learning classifiers. These algorithms have the ability to predict tumor malignancy and shorten the diagnostic and therapeutic turnaround times. Although the procedure only needs a little amount of radiologists' involvement, more testing on a larger patient cohort is required to confirm its efficacy.

---

* Corresponding authors at: Faculty of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou 362000, China (J. Xiahou).
E-mail addresses: lxd2019@zju.edu.cn (X. Li), jbxiahou@qztc.edu.cn (J. Xiahou).

https://doi.org/10.1016/j.jbo.2023.100520
Received 17 October 2023; Received in revised form 23 December 2023; Accepted 26 December 2023

1. Introduction

Radiography is the primary approach for identifying femoral bone tumors and lesions that resemble tumors [1]. When a lesion is unclear or has aggressive characteristics, magnetic resonance imaging (MRI) is indicated for further characterization [1]. It significantly enhances diagnostic capabilities by visualizing various tissue components such as cartilage, vascular tissue, fat, fluids, and hemosiderin. MRI plays a pivotal role in narrowing down potential diagnoses, even in cases where a definitive diagnosis is challenging. The transition towards adopting a multimodal imaging approach, which combines radiographs with MRI, is motivated by several key factors. These include the growing demand for precise diagnoses, technological advancements, a patient-centric approach to healthcare, and the substantial improvements in diagnostic accuracy.

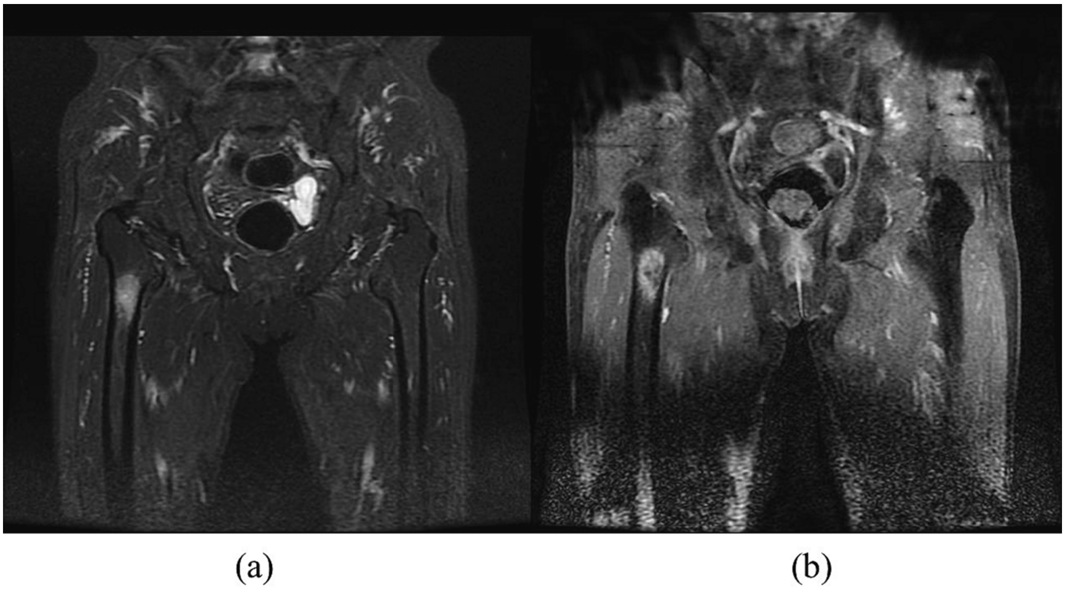
It can be challenging to see small lytic or sclerotic bone lesions on radiographs alone. When it comes to spotting bone marrow lesions, MRI is more effective than other imaging modalities [2]. Before the preprocessing step, each image was reviewed by a radiologist to detect anomalies in the exported images. Sample image tiles are presented in Fig. 1.

Aggressive, unsure cases need histological confirmation prior to staging and treatment plan selection. This track needs to be a component of the surgically removed zone, according to contamination [3]. Contrast-enhanced MRI (CEMRI) allows for the visualization of the tumor's most vascularized regions, and MRI guidance enables the avoidance of necrotic area biopsies [2]. MRI serves a pivotal role in the process of local staging and surgical planning by providing invaluable insights into the assessment of intramedullary extension and the potential invasion into neighboring structures, encompassing bones, joints, muscle compartments, and intricate neurovascular bundles. Furthermore, MRI's significance extends to post-treatment scenarios, making it an indispensable tool for re-evaluating the disease following neoadjuvant therapy and for ongoing patient monitoring during the post-treatment follow-up phase.

This research presents a novel method for assessing the malignancy of bone tumors using MRI scans. The dataset includes both MRI images and corresponding histological analyses, ensuring robust diagnosis validation. The classification of MRI images is performed using two pretrained residual convolutional neural networks (CNNs), with each network specialized in analyzing T1 and T2 MRI images, respectively. By employing pretrained CNNs, the model may train on a small dataset and converge quickly. The images are supplied to the training pipeline as a whole in our study instead of manually segmenting them to forecast cancer based on the complete image, contrary to past research that proposed manually segmenting the images [4]. The two residual CNN models' predicted values are also included in this proposed clinical model, along with the tumor's position within the bone.

2. Materials and methods

Muscle is primarily the source of T1 signal intensity in musculoskeletal MRI, however fat can also be useful in some anatomical regions, such as the fingers and toes [1]. Bone cancers show up as low-signal lesions against fatty marrow, hence T1-weighted imaging is crucial for assessing bone marrow [5]. The contrast between the cortex, marrow, and surrounding tissues is excellent in T1-weighted images [5]. Fat suppression in MRI confirms the presence or absence of fat in a lesion, assisting in the diagnosis of diseases such lipomas and haemangiomas [6]. Following contrast injection, suppressing the fat signal in T1-weighted images improves the evaluation of tumor vascularization [5]. Chemical shift-based fat suppression increases sensitivity for identifying metastases and myeloma lesions, specificity for ambiguous marrow abnormalities, and evaluation of bone marrow infiltration [5,6]. On T2-weighted images, cartilaginous tissue, which appears hyperintense, is present in the majority of bone tumors and tumor-like lesions [5]. In order to assess vascularity and differentiate between edema and living tumors, contrast-enhanced MRI helps distinguish between solid and fluid-containing lesions [5,7]. Distinguishing between benign and malignant tumors with the assistance of dynamic contrast-enhanced imaging can be a challenging task. However, this imaging technique provides a valuable opportunity for targeted biopsies and non-invasive evaluation of treatment responses. It does so by highlighting the more pronounced enhancement in viable tumor tissue compared to non-tumoral or necrotic regions, enabling healthcare professionals to make more informed diagnostic and treatment decisions [7]. Water molecule diffusion limits in malignant tumors are shown by diffusion-weighted imaging, which also allows for the assessment of therapy efficacy [5,8,9]. By identifying choline metabolites, MR spectroscopy has the potential to discriminate between malignant and benign tumors [1,11]. When early sequences are equivocal, it's critical for each institution to develop guidelines for MSK imaging evaluation



(a) (b)

Fig. 1. A solid (fibrous dysplasia) or liquid (solitary bone cyst) femoral lesion with a high signal on T2FS WI (a) may be present. The diagnosis of a solitary bone cyst was confirmed by the peripheral increase on the T1FS gadolinium-based contrast medium-enhanced sequence (b), which is indicative of liquid content.

and take contrast-enhanced MRI into consideration [12]. Many important suspicious imaging features can be assessed using MRI (as shown in Table 1).

Note that 63 MRI scans of individuals with proven malignant and benign tumors based on histological examinations were used in our study. Men and women between the ages of 17 and 78 made up the participants, who were observed between 2019 and 2021. Our dataset contains data on the patients’ names, the quantity of MRI scans conducted, the patients’ ages, genders, the bones implicated, the location, and the tumor kind. We retrieved images from each MRI scan for both T1 weighted and T2 weighted images in all three planes (sagittal, coronal, and transverse) in order to pre-process the data. This pre-processing step’s goal was to divide the datasets into two distinct ones, one with T1 weighted images of malignant and benign tumors and the other with T2 weighted images. The pre-processing resulted in the acquisition of 1677 T1 weighted images and 534 T2 weighted images, both of which contained benign and malignant tumors. A radiologist thoroughly inspected each image before pre-processing to spot any irregularities.

MRI scans were selectively chosen to include only those with lesions, and their data were stored in the DICOM format. These images were meticulously organized into distinct folders based on lesion type, with ’0′ representing benign tumors and ’1′ signifying malignant ones. Building a robust model relies heavily on a well-structured dataset. In this research, the dataset underwent extensive image processing procedures with the primary goal of amassing ample data for effective model training. Data augmentation, as a preprocessing method, involves selecting a sample from the dataset and generating additional samples by applying various methods to it. These extra samples are similar to the original sample and do not differ enough to warrant changing the input’s label. Using data augmentation approaches, a deep learning model can prevent overfitting the training set of data [13]. The following processes were performed to generate new data for each image in the collection: brightness adjustment, contrast adjustment, horizontal or vertical flipping, rotation (0–15 degrees), width adjustment, height adjustment, and shearing.

2.1. Proposed method

This paper proposes two improvement methods for the classification of two types of MRI images, benign tumors, and malignant tumors:

1. Introduction of a Global Hybrid Attention Block (GHAB): The GHAB module primarily utilizes convolution operations to extract feature maps to obtain global and channel information. It aims to reduce the loss of key features when capturing the location and content information of bone tumors, thereby improving the accuracy of the model’s classification.
2. Improvement of the network architecture of DenseNet: Due to the inherent consistency in the distribution of medical image data, overfitting is a significant concern, especially with limited datasets. In response to this challenge, the paper proposes a solution involving a restructuring of the DenseNet network’s connectivity. More precisely, it entails the removal of intermediate layer connections, leaving only those connections that carry crucial features at both ends intact. This revised DenseNet configuration is further enhanced with the integration of the GHA-Block, resulting in the creation of a modified network known as GHA-DenseNet. This modification aims to optimize the network’s performance and its ability to extract essential features.

Table 1
A list of radiological and clinical traits that are used to distinguish between chondrosarcoma and enchondroma.

Enchondroma	Chondrosarcoma
More typical in the extremities.	Greater prevalence in the axial skeleton
No pain (only if a fracture is present)	Can be painful
There is no periostitis	May be associated with periostitis
There is no growth after skeletal maturity	Untainted by skeletal maturity and rapid growth
No soft tissue is involved	Associated with localized thickening of the cortex
Bone oedema is not seen	Various signal intensity patterns
The normal size is 4 cm	Bone destruction
	Follow-up calcification loss
	Component with soft tissue
	Bone marrow swelling
	Usually, >4 cm in size

2.1.1. GHA-DenseNet

Deep convolutional neural network models can suffer from issues such as gradient vanishing and exploding gradients as the network depth increases, leading to decreased model performance.

DenseNet networks are renowned for their innovative connectivity strategy, which addresses these issues by seamlessly passing features to subsequent layers. This feature has made DenseNet a popular choice for various tasks. Nevertheless, deep convolutional networks tend to have a high number of trainable parameters and typically demand ample data samples to attain reliable classification outcomes. In the case of this paper, where the available MRI image samples of bone tumors are limited, employing the original model as is could result in significant overfitting issues.

To help reduce the loss of important tumor information and address the issues related to limited data, a Global Hybrid Attention Block is embedded within the model. Consequently, this paper proposes the GHA-DenseNet network model, as depicted in Fig. 2.

The GHA-DenseNet network model takes as input a bone tumor MRI image and outputs the probabilities of the image belonging to the categories of benign tumor and malignant tumor. It starts by employing a convolutional layer with a 7x7 filter and a stride of 2 to perform a 2x down-sampling on the input image, reducing the dimension of the feature map to half of its original size. Batch normalization (BN) is applied to alleviate overfitting, followed by ReLU activation to leverage gradient information. Subsequently, another down-sampling operation is carried out using a 3x3 filter and a stride of 2. This further reduces the dimension and computational parameters. The main body of the network consists of 4 Dense Blocks, 7 GHA Blocks, and 3 Translation Layers, arranged in the order shown in Fig. 2. Each Dense Block contains a specific number of neurons, organized in a bottleneck format comprising BN-Relu-Conv(1x1) + BN-Relu-Conv(3x3) for improved computational efficiency and effectiveness. Translation Layers consist of BN-Relu-Conv(1x1) and an average pooling layer to reduce the overall parameter count. Following each Dense Block and Translation Layer, the designed Global Hybrid Attention Block is embedded to enhance the network’s feature extraction capabilities, emphasizing spatial and channel domains. To combat overfitting, Dropout is systematically implemented across all model components except for the GHA Blocks. In the model’s concluding phase, it leverages global average pooling to capture spatial information from the last output feature map. This information is subsequently channeled through fully connected layers, culminating in the employment of the Softmax activation function to yield the final classification outcomes. Further elaboration on the intricate connectivity mechanisms found within Dense Blocks and GHA Blocks will be provided in the following section. This approach enhances model generalization and feature extraction while preserving crucial details.

2.1.2. Global hybrid attention module

The application of attention mechanisms in convolutional neural networks helps the network focus more on regions of interest, facilitating the extraction of crucial features from the image while reducing attention to less important features. The global hybrid attention module proposed in this paper is illustrated in Fig. 3.

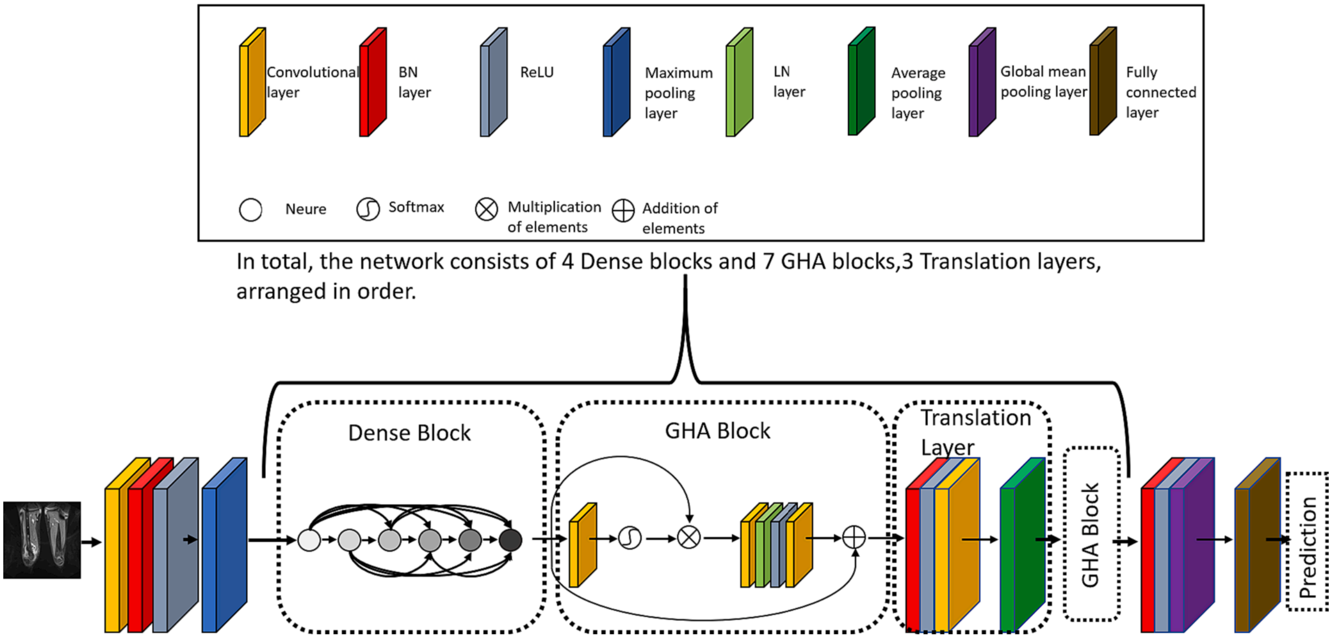


Fig. 2. GHA-DenseNet Network Architecture.

From Fig. 3, it can be observed that the global hybrid attention module consists of three components: the context modeling module, feature transformation module, and fusion module.

The context modeling module is inspired by non-local mean filtering [14], and its specific implementation steps are as follows: First, the input feature map $X \in \Re^{H\times W\times C}$ is passed through a 1x1 convolutional filter to obtain a weighted positional feature map $f(x_j) \in \Re^{HAx\times 1\times 1}$, where x_j denotes the position used for weighting. The obtained weighted positional feature map is then transposed and dimensionally reduced to obtain $\theta(x_j) \in \Re^{kHHW}$, and a global spatial weighting vector $U_s \in \Re^{1\times HW}$ is obtained through a Softmax activation function. This global spatial weighting vector is then multiplied with the input feature map, which is reduced from 3D space to 2D space, resulting in the global spatial feature identifier $I_s \in \Re^{|x|\times C}$:

$$X = \sum_{j=1}^{N_p} x_j \tag{1}$$

$$f(x_j) = W_s \cdot \sum_{j=1}^{Np} x_j \tag{2}$$

$$U_s = \sum_{j=1}^{N_p} \frac{\exp(\theta(x_j))}{\sum_{m=1}^{N_p} \exp(\theta(x_m))} \tag{3}$$

$$I_s = U_s \times X \tag{4}$$

In the equation: $W_s \in \Re^{IXC}$ represents the weights of the convolution. From the aforementioned equation steps, it can be understood that the context modeling module computes a weighted position only once on the input feature map to obtain the global spatial feature identifier, I_s. This is a spatial-level dot product that aggregates all point-wise information to a single point, effectively representing the global information of the input image and providing a global perspective.

The feature transformation module is responsible for channel-level feature fusion at each position. This process leverages two convolution operations to capture intricate non-linear relationships within the channels. To break it down further, the first convolution layer initially reduces the dimension with a reduction ratio of 16. Subsequently, the second convolution layer restores the dimension back to the original input feature dimension. The global spatial feature identifier is used as the input to the first convolution layer. The output is normalized using layer normalization (LN) [15] to normalize the channel features. Subsequently, a ReLU activation function is applied to capture the non-linear relationships at the channel level, and this is used as the input to the second convolution layer to obtain the channel identifier $I_c \in \Re^{|1\times 1\times C}$:

$$I_{GHA} = I_c = W_{c_2} \cdot relu(LN(W_{c_1} \cdot I_s)) \tag{5}$$

In the equation: $I_{GHA} \in \Re^{1\times 1\times C}$ represents the obtained global hybrid attention identifier, where C is the number of channels. $W_{c_1} \in \Re^{\frac{C\times C}{C}}$ and $W_{c_2} \in \Re^{C\times\frac{C}{r}}$ represent the weights for the first and second convolutions, respectively. The fusion module involves adding the obtained attention

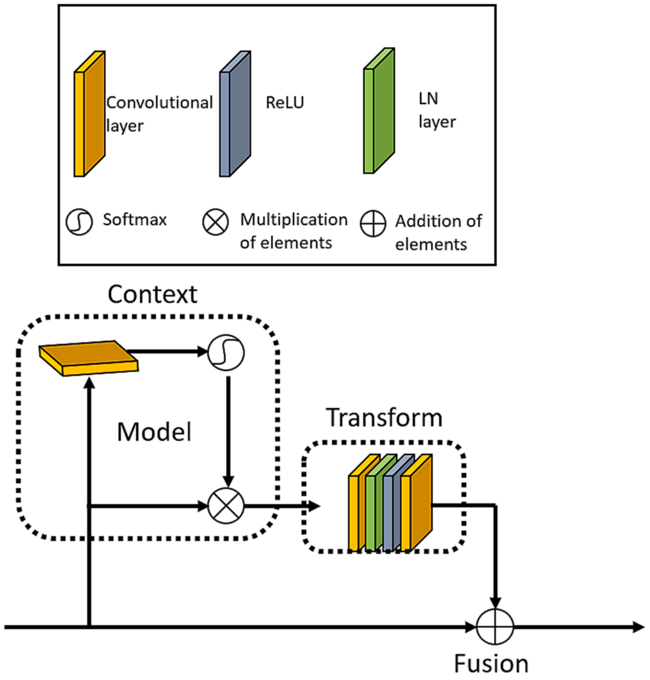


Fig. 3. Proposed methodology used in this study.

identifier to the input feature map to obtain the final attention feature map, denoted as $X_{GHA} \in \Re^{H \times W \times C}$:

$$X_{GHA} = X + I_{GHA} \tag{6}$$

2.1.3. *Dense block pruning*

DenseNet has achieved notable performance in tasks like natural image classification, such as the ImageNet dataset. However, when it comes to classification tasks with small-scale datasets like medical images, it is challenging to achieve results comparable to those in natural image classification. The primary issue in small medical image datasets is the problem of overfitting. This paper proposes a sparse connectivity approach, different from the Log-DenseNet introduced by Hu et al. [16], which, while empirically more effective than DenseNet, retains only deep-layer connections. The design in this paper involves removing connections from intermediate layers and preserving connections that contain important features at both ends, thereby reducing the original connectivity to mitigate overfitting.

The original DenseNet connectivity involves each layer's input being a merge of the outputs from all previous layers, as described by the equation:

$$x_l = H_l[x_0, x_1, \cdots, x_{l-1}] \tag{7}$$

In the improved sparse connectivity approach, the equation is as follows:

$$x_l = H_l[x_0, x_1, \cdots, x_{\lfloor \log(l-1) \rfloor .-1}, x_{l-1-(\log(l-1) \rfloor .-1)}, \cdots, x_{l-1} \tag{8}$$

In Eqs. (7) and (8), x_l represents the output of the l-th layer, and H_l is the non-linear function for each layer. As Eq. (8) shows, the pruned connectivity indicates that the input for the l-th layer equals the connection from the previous $\log(l-1)$ layers and the output from the subsequent $\log(l-1)$ layers. This approach retains the connections between shallow and deep layers, reducing unnecessary connections in the model. This adaptation addresses the requirement of extracting features from deep networks while accommodating the constraints imposed by small-scale medical image datasets.

2.2. *Data sets and data preprocessing*

The dataset experienced a haphazard torn apart into three segments: 75 % for training, 25 % for testing, and 12 % for authentication. The GHA-DenseNet model was then proficient for a total of 32 epochs, using a batch size of 22 and booming out 17 steps per epoch, along with 6 authentication steps. In fixing the dataset for the clinical model, the demographic data pertaining to the study participants, including gender, bone type, and lesion location, was converted into numerical codes, as illustrated in Table 2. Additionally, for every MRI scan, a minimum of two arbitrary images (comprising a T1 weighted image and a T2 weighted image) were selected as accompanying inputs for the clinical model. The selected images were extracted from the dataset designed specifically for the GHA-DenseNet classifiers. The dataset exclusively comprised images portraying various lesions, whether they were malignant or benign. These images were subsequently fed into the GHA-DenseNet classifier to determine the expected malignancy level for each one. Following this, the next phase involved computing the average of the predicted malignancy scores generated by the GHA-DenseNet model. These averaged results were then incorporated into the clinical prototype dataset, serving as the predicted output for both T1 and T2 images.

Due to the absence of T2 imaging in eight MRI scans, the clinical model dataset featured gaps in the column pertaining to T2 predicted malignancy. To preserve the integrity of the final output, the missing data for T2 were handled by using the standard deviation of the anticipated values. This standard deviation is represented as Eq. (9), with 'x' denoting the arithmetic mean of the observations in the column for T2 projected malignancy.

Table 2
Encoded values of clinical data.

Dataset Column	Value	Encoded Value
Gender	Male	0
	Female	1
Bone	Tibia	1
	Femur	2
	Humerus	3
	Phalanges	4
	Scapula	5
	Ilium	6
	Sacrum	7
	Fibula	8
Location	Axial	0
	Coronal	1
	Saggital	2

$$StdDev = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{9}$$

Before initiating the training process for the clinical model, a strategic oversampling method was employed to bolster the dataset. This technique entailed duplicating all existing samples within the clinical model dataset, ensuring a robust reservoir of samples for the subsequent training and validation stages. During the oversampling phase, a total of 327 new samples were meticulously generated. These additional samples were then judiciously distributed, with 75 % earmarked for training and the remaining 25 % designated for validation. This approach effectively addressed the potential limitations arising from an initially limited dataset, enhancing the model's training and assessment phases. The model's training was carried out using a batch size of 156 over the course of 47 epochs.

2.3. *Performance metrics*

In classification tasks, analyzing a model's performance is essential for determining its efficacy. The effectiveness of a classification model is evaluated using a variety of indicators. Accuracy, precision, and recall are the three main measures this paper focuses on.

A classification model's accuracy is determined by computing the percentage of true positives and true negatives it correctly predicted out of all the instances in the dataset.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

The accuracy of the model's successful predictions is measured by precision. Out of all positive predictions (true positives and false positives), it calculates the percentage of true positive predictions.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

The model's capacity to accurately detect positive cases is measured by recall, also known as sensitivity or true positive rate. Out of all real positive instances (true positives and false negatives), the fraction of true positive predictions is calculated.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

3. Results and discussion

In this study, 63 MRI scans made up the dataset used by the suggested approach. These scans revealed tumors that were both malignant and benign, as determined by histological examinations. The 44 patients, who ranged in age from 17 to 78 and comprised both men and women, had their MRI images taken. Several measures, including accuracy, precision, recall, and area under the curve, were used to evaluate the effectiveness of the image classifiers, particularly the GHA-DenseNet models. Table 3 displays the classifiers' performance outcomes for T1 and T2 weighted images. A threshold of 0.6 has been chosen after dividing the values into training and validation stages. During the authentication phase, image classification for T1-weighted images attained correctness of 97.00 %, a recall of 95.65 %, and an area under the curve of 1.00. The T2 weighted image classifier's authentication phase accuracy, recall, and area under the curve were 95.0 %, 95.52 %, and 0.9923, respectively.

The clinical model's accuracy, recall, area under the curve, and receiver operating characteristic (ROC) curve all showed similar outcomes. The values found in relation to the clinical archetypal are shown in Table 4. The mismatch in the number of T1 and T2 images provided was cited as the reason for the study's recognition of the performance gap of the GHA-DenseNet classifiers. Based on the dataset utilized, the suggested technique performed well, with the T1 classifier displaying higher precision. The ROC analysis revealed that the clinical model had the ability to separate classes. However, the performance of the final model was impacted by the unbalanced dataset. The study included an illustration of a benign tumor that the T1classifier properly identified as non-malignant.

Although it is still below the 50 % criterion, the T2 classifier indicated that the tumor in the matching T2 image was 41.05 % likely to be malignant. This implies that the T2 GHA-DenseNet classifier might produce predictions that are falsely optimistic. Each of the images in Fig. 4(a) (i) (ii) and Fig. 4(b) (i) (ii) were subjected to class activation maps (CAM). Fig. 4(a) (iii) (iv) and Fig. 4(b) (iii) (iv) of the generated CAM images demonstration that the proficient GHA-DenseNet models in the suggested method were successful in extracting pertinent features associated to the tumor from the MRI scans.

The ROC curve in Fig. 5(a and b), which depicts the effectiveness of a binary classifier, was created by plotting the confusion matrix of the proposed system as well as the true positives and false positives at various threshold values. The area under the ROC curve for a binary classifier is 0.5. The ROC curve shows that the generated model performs satisfactorily at various thresholds on the study's data.

However, the overall profile of the estimated measures is comparable to what was previously reported in tumour necrosis estimation absolute errors. The MRI subgroups with the highest levels of similarity are CONV + DCE-q, CONV + DW + DCE-q, and CONV + DW + DCE-q + DCE-s. For instance, using min-avg optimization, the dice and overlap coefficients for CONV + DW + DCE-q are (0.69, 0.04) and (0.91, 0.01) based on $(n = 7)$, respectively. On the other side, the similarity between the tumor viability maps generated by CONV, DW, and CONV + DW and those predicted by histology is the lowest. The CONV instance's dice and overlap coefficients are (0.61, 0.07) and (0.90, 0.08) based on $(n = 10)$, respectively. As would be expected, the dice coefficient is far less than the overlap coefficient. This shows that the sizes of A_necL and B_necL are actually significantly different. The overlap coefficient, however, demonstrates that the two sets are primarily contained within one another.

Table 3
Performance metrics for image classifiers.

Train Step	Accuracy	Recall	Precision	Area under the Curve
T1 classifier	92.88 %	94.27 %	96.35 %	0.9639
T2 classifier	87.03 %	84.05 %	88.19 %	0.8911
Validation Step	Accuracy	Recall	Precision	Area under the Curve
T1 classifier	95.24 %	94.78 %	97.13 %	0.9976
T2 classifier	95.24 %	99.71 %	86.42 %	0.9987

Table 4
Measures of performance for a medical model.

	Accuracy	Recall	Precision	Area under the Curve
T1 classifier	82.17 %	94.52 %	75.66 %	0.8905
T2 classifier	81.51 %	93.12 %	76.74 %	0.8843

MRI plays a crucial role in providing qualitative insights into the composition of tumors. It accomplishes this by distinguishing between fat, water, and blood within the tumor. Malignant tumors often exhibit distinct characteristics, such as larger size, irregular borders, varying signal patterns, signs of edema, necrosis, hemorrhage, fascial penetration, bone alterations, and involvement of neurovascular structures. Additional indicators of malignancy include peritumoral edema, heterogeneous signal intensity in T1-weighted images exceeding 5.3 cm, and elevated signal intensity in T2-weighted images. Specific features like tumor necrosis, a diameter exceeding 8 cm, and evidence of neurovascular involvement further enhance the diagnostic specificity in MRI assessments. [17]. According to a study [18], bone tumor malignancy can be predicted using MRI scans and deep learning algorithms. The study harnessed pretrained residual network models alongside a strategically designed fully connected neural network, which was tailored to clinical relevance. These models operated synergistically, producing a continuous numerical output ranging between 0 and 1, serving as an indicator of the anticipated malignancy. Impressively, the T1 classifier showcased remarkable performance by accurately classifying both true positive and true negative cases, exemplified by its impressive accuracy of 92.88 % and an impressive recall rate of 94.27 %. The absence of T2 images within the MRI scans did introduce a minor limitation, with the T2 classifier displaying marginally lower accuracy and recall scores. It nevertheless performed well, with an accuracy and recall of 87.03 % and 84.05 %, respectively. High area under the curve values (>0.90) for both classifiers indicated accurate classification. The bone tumor prediction model has an accuracy of 82.17 % and a recall of 94.52 % after being trained using clinical data and the T1 and T2 classifiers' predicted values.

Some studies have used deep learning algorithms to categorize bone lesions in the field of orthopaedics. The method put forth by Feyisope et al. [4] requires prior image segmentation for training, but the method provided in this study does not. The suggested method, on the other hand, employs a pretrained GHA-DenseNet classifier to categorize all T1 and T2 images as benign or malignant. Additionally, the bone tumor prediction model used is a classifier neural network with input and output layers that include sex, age, bone type, bone location, and predicted malignancy values from the T1 and T2 classifiers. With just 44 patients and 63 MRI images, this study contained fewer cases than other studies. The majority of the participants were 22 women between the ages of 16 and 74. While the clinical model demonstrated commendable performance with an 81.52 % accuracy rate during validation, it's important to acknowledge the ongoing challenge of class imbalance in the dataset. This issue arises due to the relatively lower occurrence of bone tumors compared to other tumor types. This study, being a pilot trial, underscores the need for further validation with a more extensive and diverse patient population to enhance the model's robustness. To achieve more precise orthopedic analyses, it is advisable to consider future implementations of both computed tomography (CT) in addition to MRI, as indicated in reference [19]. These modalities can collectively contribute to more accurate diagnostic assessments in the field of orthopedics.

4. Limitations and future work

Despite the positive findings of this study, it is vital to recognize the limitations that need to be resolved in subsequent investigations. The relatively small sample size employed in this study is one drawback that might restrict the generalizability of the results. In order to determine

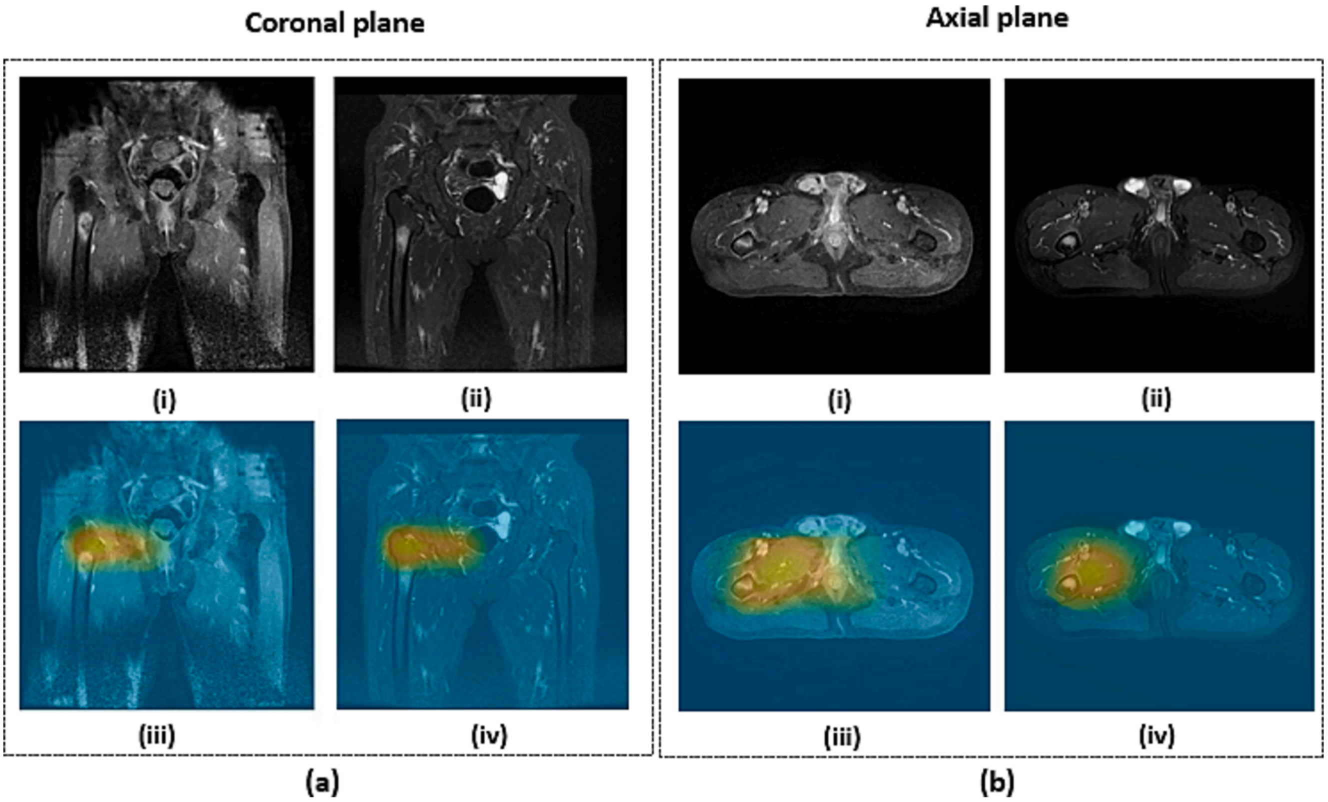
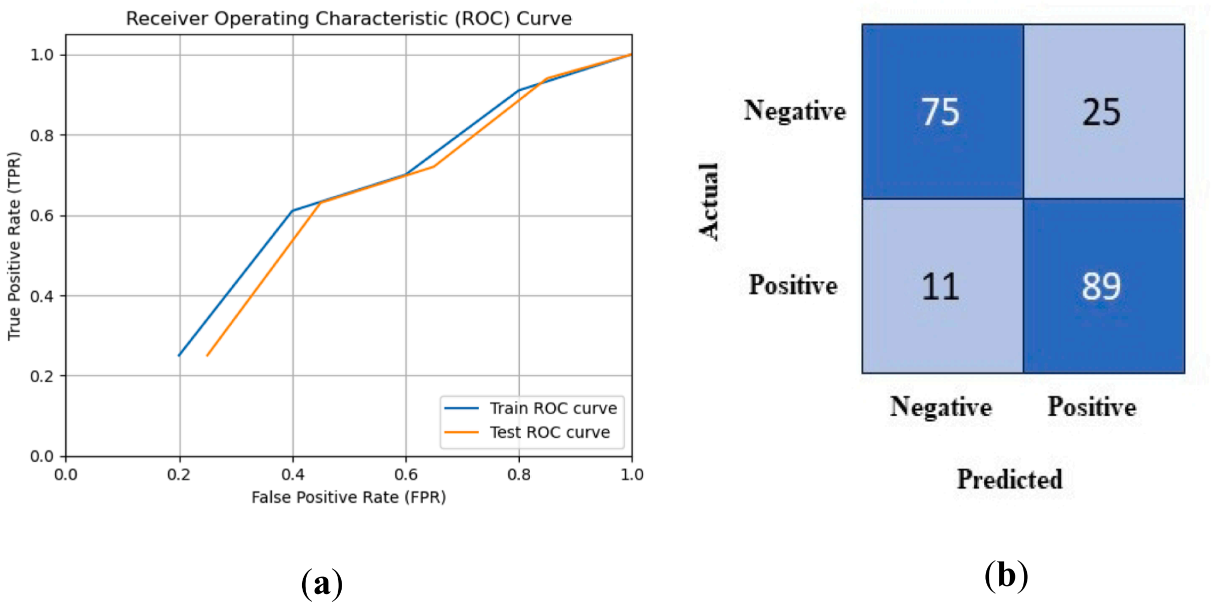


Fig. 4. MRI scans of the femur that pertain to (a) coronal plane and (b) axial plane, depicting (i) malignant tumour—T1 FS with weighted image; (ii) benign tumour—T2 weighted image; and (iii, iv) corresponding class activation maps.



	Negative	Positive
Negative	75	25
Positive	11	89

Fig. 5. Performance evaluation of GHA-DenseNet model based on (a) receiver operating characteristic curve and (b) confusion matrix.

the generalizability and robustness of the suggested strategy, additional validation and testing on a larger and more varied cohort of patients is needed.

The current investigation included patient clinical data and concentrated on T1 and T2-weighted MRI images. To further improve the GHA-DenseNet model's predictive accuracy, future studies can investigate the incorporation of different imaging modalities or new clinical variables. Additionally, external validation of distinct datasets from various institutions would offer more thorough insights into the model's functionality and generalizability.

Moreover, there is a strong case for conducting a prospective study to evaluate the practical utility of the developed model in real-world clinical settings. Such a study should encompass a comprehensive investigation into how this model influences clinical decision-making, aids in treatment planning, and ultimately impacts patient outcomes. This suggested approach is expected to evolve and improve over time as these critical aspects are addressed. Consequently, its continued development holds the promise of significantly benefiting individuals diagnosed with malignant bone tumors, as well as pushing the boundaries of medical imaging and deep learning-based tumor classification within the broader medical community.

Artificial intelligence based diagnostics [20] can be utilized for orthopedics applications. Neural architecture search [21] based on Pareto optimality [22,23] of the best error loss from the permutation of all possible neural weights settings can be implemented in future research. In addition, this work is a preliminary exploration of the classification of MRI images of bone tumors. In the subsequent work, we will try to use generative adversarial network to further explore the network model based on the abovementioned framework.

5. Conclusion

This paper introduced, a state-of-the-art method for predicting the malignancy of bone cancers using magnetic resonance imaging (MRI) images and neural network models. During the training phase, our pretrained GHA-DenseNet image classifiers classified T1 and T2-weighted MRI images with high accuracy, obtaining 92.88 % and 87.03 %, respectively. These classifiers performed well, maintaining a high accuracy level of 95.24 % during validation.

Moreover, the clinical model, which incorporated both patient clinical data and the outcomes of the T1 and T2 classifiers, exhibited commendable accuracy. Specifically, it achieved an accuracy rate of 82.17 % during the training phase and maintained a robust performance with an accuracy of 81.51 % during the validation process. Notably, the clinical model effectively distinguished between different classes, a fact substantiated by the comprehensive receiver operating characteristic (ROC) curve analysis, reinforcing its diagnostic capabilities and potential clinical relevance.

The suggested approach has a number of benefits, including the use of deep learning classifiers that have already been trained, which eliminates the need for manual segmentation of MRI data. This method simplifies the procedure for determining a tumor's malignancy, which may shorten the time between diagnosis and treatment. According to the study's findings, the created model may help with the timely start of therapy, improving vital and functional outcomes for patients with malignant bone tumors.

CRediT authorship contribution statement

Y. Yuguang: Writing – original draft, Resources, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Yusi Chen:** Validation, Resources, Funding acquisition. **Daxin Zhu:** Visualization, Validation, Software, Investigation, Conceptualization. **Yifeng Huang:** Validation, Software, Conceptualization. **Ying Huang:** Software, Resources, Investigation. **Xiadong Li:** Validation, Supervision, Methodology, Conceptualization. **Jianbing Xiahou:** .

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by the Science and Technology Program of Quanzhou (No.2021CT0010). The authors also acknowledge the support by Fujian Provincial Key Laboratory of Data-Intensive Computing, Fujian University Laboratory of Intelligent Computing and Information Processing, and Fujian Provincial Big Data Research Institute of Intelligent Manufacturing. Special thanks to The Huaqiao University Affiliated Strait Hospital and the Children's Hospital of Zhejiang University School of Medicine for providing the MRI data used in this research.

References

[1] American College of Radiology (2013) ACR Appropriateness Criteria. Primary Bone Tumours. American College of Radiology. Available via http://www.acr.org/~/media/ACR/Documents/AppCriteria/Diagnostic/PrimaryBoneTumours.pdf. Accessed 9 Apr 2014.

[2] R. Ojala, R.B. Sequeiros, R. Klemola, E. Vahala, L. Jyrkinen, O. Tervonen, MR-guided bone biopsy: preliminary report of a new guiding method, J MagnReson Imaging 15 (1) (2002) 82–86.

[3] M.B. Ribeiro, C.R. Oliveira, R.Z. Filippi, et al., Estudohistopatológico do trajeto de biópsia de Tumouresmusculoesqueléticosmalignos, Acta OrtopédicaBrasileira 17 (5) (2009) 279–281.

[4] Eweje F.R., Bao B., Wu J., Dalal D., Liao W.-H., He Y., Luo Y., Lu S., Zhang P., Peng X., et al. Deep Learning for Classification of Bone Lesions on Routine MRI. EBioMedicine. 2021. 68. 103402.

[5] A.M. Davies, M. Sundaram, S.L.J. James, Imaging of Bone Tumours and Tumour-Like Lesions (Techniques and Applications), Springer, Berlin Heidelberg, 2009.

[6] D.G. Disler, T.R. McCauley, L.M. Ratner, C.D. Kesack, J.A. Cooper, In-phase and out-of-phase MR imaging of bone marrow: prediction of neoplasia based on the detection of coexistent fat and water, AJR Am J Roentgenol 169 (5) (1997) 1439–1447.

[7] T.H. Berquist, Magnetic resonance imaging of primary skeletal neoplasms, Radiol Clin North Am 31 (2) (1993) 411–424.

[8] Y. Hayashida, T. Yakushiji, K. Awai, et al., Monitoring therapeutic response of primary bone Tumours by diffusion weighted image: initial results, EurRadiol 16 (12) (2006) 2637–2643.

[9] T.A. Bley, O. Wieben, M. Uhl, Diffusion-weighted MR imaging in musculoskeletal radiology: applications in trauma, Tumours, and inflammation, MagnReson Imaging Clin N Am 17 (2) (2009) 263–275.

[11] C.K. Wang, C.W. Li, T.J. Hsieh, S.H. Chien, G.C. Liu, K.B. Tsai, Characterization of bone and soft-tissue Tumours with in vivo 1H MR spectroscopy: initial results, Radiology 232 (2) (2004) 599–605.

[12] D.A. May, R.B. Good, D.K. Smith, T.W. Parsons, MR imaging of musculoskeletal Tumours and Tumour mimickers with intravenous gadolinium: experience with 242 patients, Skeletal Radiol 26 (1) (1997) 2–15.

[13] C. Shorten, T.M. Khoshgoftaar, A survey on Image Data Augmentation for Deep Learning, J. Big Data 6 (2019) 60.

[14] A. Buades, B. Coll, J.M. Morel, Non-local means denoising[J], Image Processing Online 1 (2011) 208–212.

[15] K. Taesup, S. Inchul, B. Yoshua, Dynamic layer normalization for adaptive neural acoustic modeling in speech recognition[C]//Conference of the International Speech Communication Association, INTERSPEECH, Stockholm, 2017, pp. 2411–2415.

[16] H. Hu, D. Dey, A. Del Giorno, et al., Log-densenet: how to sparsify a densenet[C]//International Conference on Machine Learning, ICML, Stockholm, 2018, pp. 348–355.

[17] R.A. Schoot, K. McHugh, R.R. Van Rijn, L.C.M. Kremer, J.C. Chisholm, H.N. Caron, J.H.M. Merks, Response Assessment in Pediatric Rhabdomyosarcoma: Can Response Evaluation Criteria in Solid Tumours Replace Three-dimensional Volume Assessments? Radiology 269 (2013) 870–878.

[18] V. Georgeanu, G. Golesteanu, T. Atasiei, D. Gartonea, B. Shahzad, S. Cristea, The Therapeutic Approach of Soft—Tissue Sarcomas, A 12 Cases Analysis. Rom. J. Orthop. Surg. Traumatol. 1 (2018) 58.

[19] X. Deng, Y. Zhu, S. Wang, Y. Zhang, H. Han, D. Zheng, Z. Ding, K.K.L. Wong, CT and MRI Determination of Intermuscular Space within Lumbar Paraspinal Muscles at Different Intervertebral Disc Level, PLoS ONE 10 (10) (2015) e0140315.

[20] K.K.L. Wong, G. Fortino, D. Abbott, Deep learning-based cardiovascular image diagnosis: A promising challenge, Future Generation Computer Systems 110 (2020) 802–811.

[21] K.K.L. Wong. Cybernetical Intelligence: Engineering Cybernetics with Machine Intelligence, John Wiley & Sons, Inc., Hoboken, New Jersey, ISBN: 9781394217489. 2023.

[22] K.K.L. Wong, Bridging game theory and knapsack problem, Journal of Engineering Mathematics 91 (1) (2015) 177–192.

[23] K.K.L. Wong, A geometrical perspective for the bargaining problem, PLoS ONE 5 (4) (2010) e10331.

Further reading

[10] American College of Radiology (2011). ACR Appropriateness Criteria. Follow-up of Malignant or Aggressive Musculoskeletal Tumours.