Computerized Medical Imaging and Graphics 102 (2022) 102141

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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

Bone tumor necrosis rate detection in few-shot X-rays based on deep learning

Zhiyuan Xu [a,1], Kai Niu [a,1], Shun Tang [b,1], Tianqi Song [a], Yue Rong [c], Wei Guo [b,*], Zhiqiang He [a,*]

[a] Key Laboratory of Universal Wireless Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China
[b] Musculoskeletal Tumor Center, Peking University People's Hospital, Beijing 100044, China
[c] Department of Electrical and Computer Engineering, Curtin University of Technology, Bentley, WA 6102, Australia

ARTICLE INFO

Keywords:
Bone tumor necrosis
Deep learning
Few-shot samples
Generative adversarial network
Time series

ABSTRACT

Although biopsy-based necrosis rate is a golden standard for reflecting the sensitivity of bone tumor and guiding postoperative chemotherapy, it requires biopsy which is invasive and time-consuming. In this paper, we develop a new necrosis rate detection method using time series X-ray images instead of biopsy. To overcome the limitations of few-shot samples, the proposed method utilizes a Generative Adversarial Network with Long Short-term Memory to generate time series X-ray images. For further data expansion, an image-to-image translation network is applied for producing the initial images. These augmented data are treated as the training set of a 3D-Convolutional Neural Network classification model. Our method expands the few-shot bone tumor X-rays by 10 times, and approaches the necrotic rate classification result of biopsy, which is the state-of-the-art technique in the detection of few-shot bone tumor necrosis rate. Furthermore, it provides an efficient method to investigate the bone tumor necrosis rate in few-shot samples.

1. Introduction

Primary malignant bone tumors are a group of highly malignant tumors, represented by osteosarcoma, Ewing's sarcoma, and undifferentiated sarcoma (malignant fibrous histiocytoma) etc. Among them, the most common one is Osteosarcoma (OS), which has an insidious onset and rapid growth rate. It is the second universal malignant tumor in children and adolescents (Dorfman and Czerniak, 1995; Ottaviani and Jaffe, 2009). Although great progress has been made in the study of bone tumor (Grignani et al., 2015; Lee et al., 2016; Zhang et al., 2018), a certain percentage of patients in clinical practice have primary or secondary resistance to chemotherapy, and the prognosis of such patients is poor (Ferrari et al., 2003; Fagioli et al., 2008; Duchman et al., 2015). As a consequence, accurate and timely diagnosis of the efficacy of chemotherapy is the key to improving the survival rate and prognosis of bone tumors. Necrosis rate is a widely adopted criterion to measure the sensitivity of osteosarcoma to chemotherapy and predict tumor outcome (Sami et al., 2008). Although applying biopsy to measure the necrosis rate is very effective, this invasive operation brings some risks (Interiano et al., 2016). Whether the surgeon plans to remove the entire tumor at the time of the biopsy also affects the choice of the type of the biopsy. Without the need to remove all or part of the limb containing the tumor, an incorrect biopsy can sometimes make it difficult for the surgeon to remove all tumors later. Partially removing the tumor may accelerate the spread of cancer. In addition, waiting for biopsy results can be distressing.

In recent years, deep learning has shown strong capabilities in solving medical image segmentation and classification tasks (Liu et al., 2021; Singh et al., 2021; Hansen et al., 2022; Hatamizadeh et al., 2022). In this paper, we introduce the 3D-Convolutional Neural Network (CNN) (Jin et al., 2017) for time series image classification to obtain necrosis rate results similar to those by biopsy. Considering that the value of necrosis rate is an indicator of the effect of chemotherapy, the image we use for classification is a time series diagram composed of X-ray images at different chemotherapy stages. We develop a new classification model to find the correlation between the chemotherapy effect over time and its characteristic changes on X-ray images. Thus the task of biopsy can be partially replaced by using time series X-ray images during chemotherapy instead of biopsy images(Fig. 1c). Even though this method does not give a specific necrosis rate value, a classification result of the necrosis rate with suitable threshold can be obtained.

Necrosis rate ranges from 0% to 100% and in the medical field 90% is usually considered as a threshold point that is very useful for the follow-up treatment of bone tumor (Kang et al., 2017). However, the medical imaging data with a necrosis rate above 90% is limited (Kumar and Gupta, 2016; Miller et al., 2018). The imbalance of two types of data with 90% as the threshold will cause instability of the deep

* Corresponding authors.
E-mail addresses: bonetumor@163.com (W. Guo), hezq@bupt.edu.cn (Z. He).
[1] Zhiyuan Xu, Kai Niu, and Shun Tang are co-first authors.

https://doi.org/10.1016/j.compmedimag.2022.102141
Received 30 June 2022; Received in revised form 29 October 2022; Accepted 6 November 2022
Available online 11 November 2022

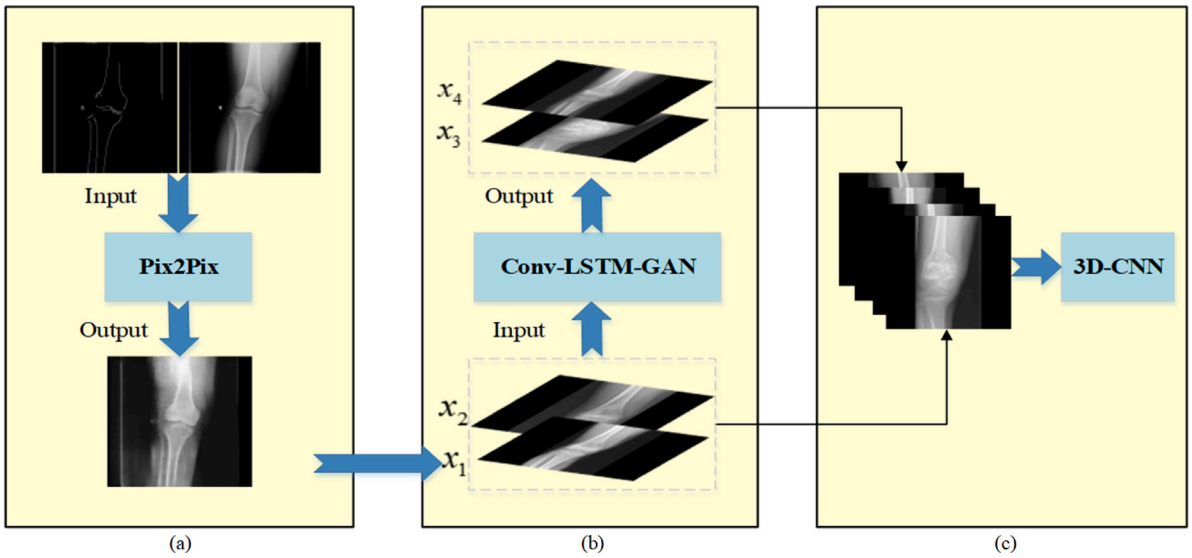

Fig. 1. Architecture of the proposed model. (a) Generating an X-ray image of the virtual patient by the Pix2Pix model. (b) Generating time series lesion images by the Conv-LSTM-GAN model. (c) Classification of time series bone tumor images by the 3D-CNN model.

learning model. To solve this problem, we divide the patients' necrosis rates into two categories with the threshold of 80% (Wang et al., 2016). The threshold is considered reasonable by orthopedic specialists of Peking University People's Hospital. Then the bone tumor necrosis rate detection problem becomes a binary classification task with a threshold as 80%.

From the analysis above, we need to track the effect of chemotherapy through time series images. However, the number of time series X-ray images of bone tumors is very limited due to patients' privacy and rarity of the bone tumor disease. This will seriously affect the accuracy of the 3D-CNN classification model and lead to over-fitting of the model caused by few-shot datasets, which is a common problem in deep learning (Shi et al., 2015; Sun et al., 2017). In order to cope with this significant challenge in the application of artificial intelligence in the medical field due to few-shot datasets, this paper presents a new data generation model named Convolutional Long Short-term Memory Generative Adversarial Network (Conv-LSTM-GAN) to construct time series X-ray images for predicting the effect of chemotherapy.

Current solution to solve the problem of small training samples is mainly data augmentation (Perez and Wang, 2017; Chen et al., 2021a). Recently, Generative Adversarial Network (GAN) is widely used in image enhancement by generating high quality and diverse images (Goodfellow et al., 2014; Zhan et al., 2021; Guan et al., 2022). Not only has it achieved good results in the enhancement of natural images, but also it has increasing applications in multimodal medical image generation and classification (Frid-Adar et al., 2018; Zhan et al., 2021; Chen et al., 2021b).

Hence in this paper, we propose a new Conv-LSTM-GAN model to generate time series X-ray images during chemotherapy with the starting images as input to augment the samples (Fig. 1b). Therefore the entire time series images are composed of the initial input images of Conv-LSTM-GAN and the subsequent generated time series images.

For the initial input image, we use another GAN model for image-to-image translation named Pix2Pix to generate a single X-ray image (Isola et al., 2017), which provides an X-ray image of the virtual patient and serves as the first input image of Conv-LSTM-GAN(Fig. 1a). Then the 3D-CNN model uses the enhanced time series images generated by the above two models as training set to obtain the final classification result.

In this way, we construct a model for the necrosis rate detection based on these three modules. Our contributions are the following:

- An image-to-image translation GAN is utilized to learn the mapping relationship between tumor lesion image and its contour. The lesion information with necrosis rate category is superimposed on the normal bone contour extraction image, then the pre-chemotherapy tumor image with the necrosis rate category label is generated, which significantly expands the input data of the Conv-LSTM-GAN model to generate more samples of the time series images.
- We propose the Conv-LSTM-GAN model to take advantage of the time dimension characteristics of bone tumor X-ray images. The model exploits the time correlation from real time series tumor images before chemotherapy and in chemotherapy. The Conv-LSTM-GAN takes the real and generated pre-chemotherapy image as the first image of overall time series images, generating subsequent image in chemotherapy.
- We input the tumor time series images generated into a 3D-CNN, which extracts the most representative tumor time series image lesion features, performing classification based on the category label of the necrosis rate. We demonstrate that the classification results of bone tumor necrosis rate obtained from images generated by the proposed approach are close to those of biopsy.

To the best of our knowledge, the method in this paper is applied to bone tumor medical image synthesis for the first time.

2. Method

The overall model presented in this article consists of serial modules as shown in Fig. 1. First, we give a general overview of the connection relationships and data interaction methods between all three sub-modules. Next, we introduce the structure and workflow of each sub-module in detail.

2.1. Overview of the model

As shown in Fig. 1, the three sub-modules are connected in a way that one module provides the generated image data for its consecutive module. First, the image-to-image translation conditional GAN (Pix2Pix) model (Isola et al., 2017) is adopted to translate normal bone contour into bone tumor lesion image. Then for the generation of time series image, we combine the traditional GAN with the Long Short-Term Memory network (LSTM) to generate the sequence images of starting images given by Pix2Pix (Sherstinsky, 2020). Finally, as enhanced input data, the generated time series bone tumor images are

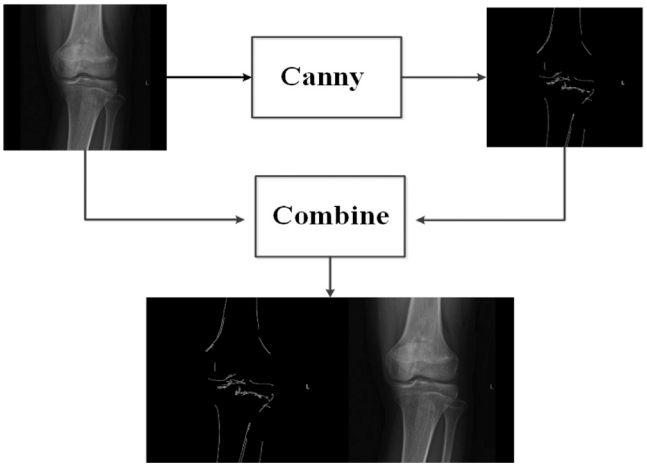

Fig. 2. The result of extracting the contour of an image from the training set or the test set and merging it with its original image.

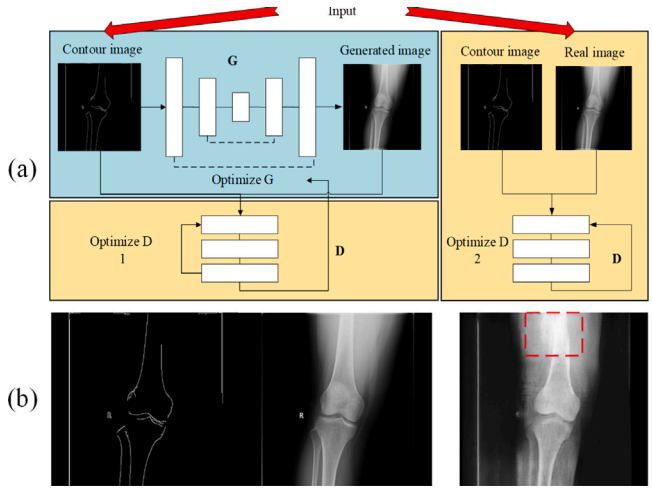

Fig. 3. Flowchart and output of the Pix2Pix model. (a) Optimization process of Generator and Discriminator. (b) From left to right: contour of the original image, original normal bone image during Pix2Pix model 1 test period, generated bone tumor image before chemotherapy with category 0.

sent into deep convolutional neural network of 3D-CNN (Sun et al., 2017) to obtain the final necrosis rate classification result.

Pre-chemotherapy image generation model. The training sets are the pre-chemotherapy bone tumor lesion image and its contour after pre-processing. During the training process, the model has learned the mapping relationship between the bone contour and the tumor lesion image. Then the test data of normal bone contour extraction images can be generated as pre-chemotherapy tumor images.

Time series image generation model. The time series images of real bone tumors at different chemotherapy stages are used as training images of the model. After the training process, the Generator can finally generate sufficiently realistic time series X-ray bone tumor images. Then in the process of Generating, the pre-chemotherapy bone tumor images generated by the Pix2Pix model are sent into the Encoder of the trained model as the given starting images of the time series images. After this, the output of the trained Decoder in the Generator is combined into the final generated time series images. In addition, the category label of the overall complete images for the following classification task is consistent with the input pre-chemotherapy bone tumor image.

Classification of time series bone tumor images. We construct a 3D-CNN network for the classification of the generated time series bone tumor images. The network parameters are trained based on the necrosis rate category labels.

2.2. Pix2Pix model generate pre-chemotherapy lesion images

2.2.1. Extract bone contour as pre-processing

We extract the bone contour by using the Canny edge detection operator in the opencv module of python3. Canny edge detection method extracting contour mainly includes the steps of graying, filtering, and calculating the gradient magnitude and direction of the image using the Canny operator.

As shown in Fig. 2, this operation is used for the original pre-chemotherapy tumor lesion image and the normal image with no tumor. Moreover, the paired pre-chemotherapy images include necrosis rates above 80% and below 80%, which are used as training sets for the Pix2Pix model twice respectively. Also, the combined images of the normal image (next referred to as category normal) and its contour are split into two batches, which are used as test sets for two trained models to generate two types of pre-chemotherapy bone tumor lesion images.

2.2.2. Pre-chemotherapy lesion images synthesis

The Pix2Pix model includes a Generator (G) and a Discriminator (D) as shown in Fig. 3a. As their names imply, the Generator generates an image, and the Discriminator determines whether it is real or fake. We send the combined images into the Pix2Pix model and select the left half of the paired image, which is the contour, and send it into G. The contour is first convolved to obtain the picture features, and then deconvolved to generate an image. Next, the generated image and its original input contour image are combined and sent into D.

Discriminator D judges whether this pair of images is real or fake and calculates the loss of this result with the preset "real" label to optimize the Generator. On the other hand, the real image and its original input contour image are also combined and sent into D besides the paired image of the generated image and its contour to optimize the Discriminator. Unlike G, the loss calculation in this optimization process includes two parts. One is the loss of D's "real or fake" result and the preset "fake" label for the generated image pair whereas the preset label to optimize G is "real", and the other loss is the "real or fake" result of D's real image pair and its preset "real" tag. The Generator and Discriminator constantly gamble to obtain the optimal model.

2.3. Conv-LSTM-GAN model generate time series lesion images

In the training stage, we first input real time-series images before and during tumor chemotherapy to train Conv-LSTM-GAN. In the generation stage, the generated and real pre-chemotherapy tumor images are combined as the input of the model, to generate time series images in chemotherapy.

2.3.1. Change the contrast and brightness to expand the train data

Because there are too few real time series images, there is not enough training data for an adversarial network to generate the time series images. Some common data enhancement methods are adopted for the training data. In this paper, the data is initially expanded by changing the contrast and brightness based on the original dataset.

2.3.2. Generator

The generation network is composed of two layers of convolutional LSTM. The input of the generation network consists of two parts, which are the memory state and the first stage (before chemotherapy) images in the time series.

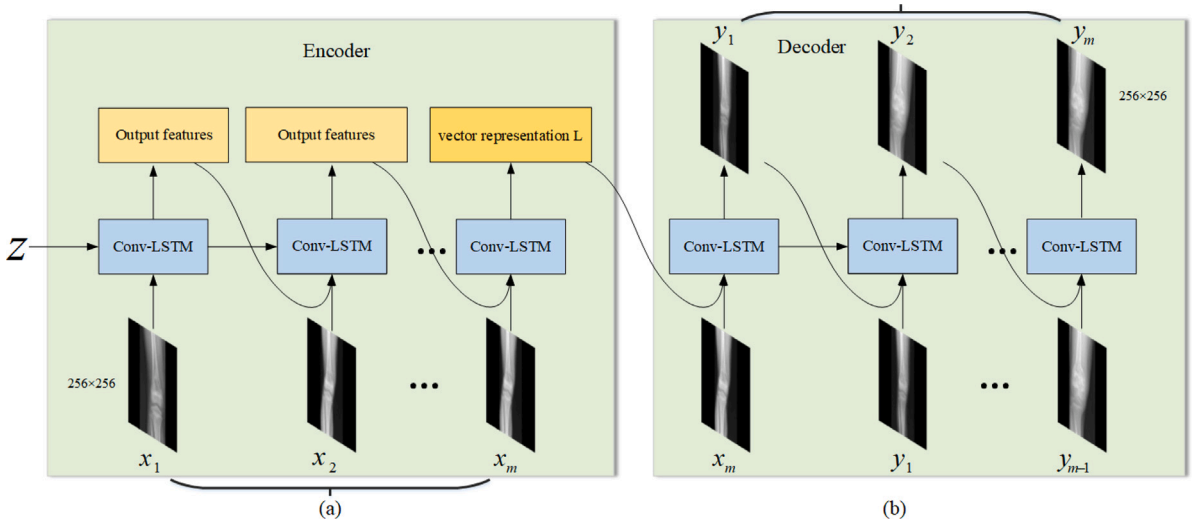

Fig. 4. Encoder and Decoder structure of Generator.

We aim to generate time series changes in the chemotherapy stage with given starting pre-chemotherapy bone tumor images. Specifically, x_i denotes the grayscale map of each bone tumor image, $X \in \mathbb{R}^{W\times H\times m}$ denotes the sequence of the first m images of real time series bone tumor images as shown in (1). Next we send this image sequence and the initial state into the Encoder as shown in (2). The Encoder transforms the input m images into a vector representation L through a non-linear transformation F, where W and H denote the width and height of an image, respectively.

$$X \in \mathbb{R}^{W\times H\times m} = \langle x_1, x_2 \cdots x_m \rangle \tag{1}$$

$$L = \mathrm{F}(z, X \in \mathbb{R}^{W\times H\times m}) \tag{2}$$

For the decoder as shown in (3), its task is to generate the image y_i at time i according to the intermediate vector representation L of the input m images and the historical information $y_1, y_2, \ldots, y_{i-1}$ that has been generated before as shown in Fig. 4. In particular, the first generated image y_1 is obtained by taking the last image x_m of the input time series as input.

$$y_i = \begin{cases} Decoder(L, y_1, y_2, \ldots y_{i-1}) \ , if \ i \geq 2 \\ Decoder(L, x_m) \ , \ if \ i{=}1 \end{cases} \tag{3}$$

Significantly, the basic structures of the Encoder and Decoder are a two-layer ConvLSTM network. By repeating this neural network module, a chain structure is formed as a whole Encoder or Decoder as shown in Fig. 4. The mapping process of the Encoder can be expressed as follows:

$$\left[H_t, C_t\right] = \begin{cases} ConvLSTM\left(x_t, \left[H_{t-1}, C_{t-1}\right]\right) \ , \ if \ 2 \leq t \leq m \\ ConvLSTM\left(x_1, \left[z^k, z^k\right]\right) \ , \ if \ t{=}1 \end{cases} \tag{4}$$

Where $ConvLSTM$ denotes the ConvLSTM cell of the Encoder, H_t and C_t represent the cell state and the hidden state of the ConvLSTM cell at time t, respectively, noise z makes up the initial state variables in the ConvLSTM cell of the Encoder network, both the cell state C_1 and the hidden state H_1 are linearly transformed by z, the dimension of z is k, and $x_i \in \mathbb{R}^{W\times H}$ denotes the input image in this ConvLSTM cell at the i_{th} time. In this way, we can obtain the final vector representation L of the Encoder, which is equivalent to $\left[H_m, C_m\right]$.

In contrast, the Decoder needs to save the output at each point in time as the generated image while the Encoder is only used to save the output state of the last point to the Decoder. (5) indicates the process of each generated image by the ConvLSTM cell at time t. Similar to the Encoder, the input of each cell is $\left(y_t, \left[H_{t-1}, C_{t-1}\right]\right)$ at time t, y_t denotes the generated bone tumor input image at time t. In particular, the input for generating the image in the first time phase is the last frame x_m of the input time series and vector representation L obtained by the encoder.

$$\left[H_t, C_t\right] = \begin{cases} ConvLSTM\left(y_t, \left[H_{t-1}, C_{t-1}\right]\right) \ , \ if \ t > m+1 \\ ConvLSTM\left(x_m, L\right) \ , \ if \ t = m+1 \end{cases} \tag{5}$$

$$G_{img} = \sum_{t=1}^{n} y_t = \sum_{t=1}^{n} Conv(H_t) \tag{6}$$

We save the output of each time point y_t to make up the generated time series images as shown in (6). G_{img} denotes the generated time series images, n is the length of the time series. Each generated image y_t is obtained by the convolution operation of the hidden state H_t of the corresponding time period in the ConvLSTM cell, where $Conv$ represents a multi-layer convolution operation. Finally, the generated images are combined with the corresponding previous time series images input of the Encoder to obtain the entire time series images.

2.3.3. Discriminator

On the other hand, the network structure of the Discriminator is similar to that of the Encoder. Because the role of the Discriminator is equivalent to that of a classifier, it extracts the whole image sequence features to obtain the final "real or fake" classification result to optimize itself and the Generator. Hence only the Encoder is needed to achieve this purpose. As with the Encoder in the Generator, we send time series images into the Discriminator, expand them in the time sequence through the same ConvLSTM module, continuously update the variable parameters of the ConvLSTM module, and finally obtain the "real or fake" classification result. The input image and state variable form are consistent with the Encoder network in Generator as shown in (7),

$$\left[H_t, C_t\right] = \begin{cases} ConvLSTM\left(x'_t, \left[H_{t-1}, C_{t-1}\right]\right), \ if \ 2 \leq t \leq n \\ ConvLSTM\left(x'_1, \left[v^{label}, v^{label}\right]\right) \ , \ if \ t{=}1 \end{cases} \tag{7}$$

where x'_t represents a real image or a generated image when optimizing D, n is the length of the time series, which is also consistent with the subsequent length of the real sequence except the input sequence. Different from G, the initial state vector settings here are not converted by the noise vector z, but all-zero vectors with the same dimensions as the number of label categories $label$. The state vector output at the last stage is transformed into a vector of "real or fake" through a linear transformation. So the specific function of the Discriminator is: $(X \in \mathbb{R}^{W\times H\times n}) \rightarrow A \in [0, 1]$, which can distinguish the real time series (ideal result is 1) images from the generated time series images (ideal result is 0) as shown in Fig. 5a. In other words, the output of the last stage of the LSTM network is linearly transformed to determine whether the input time series is real or fake.

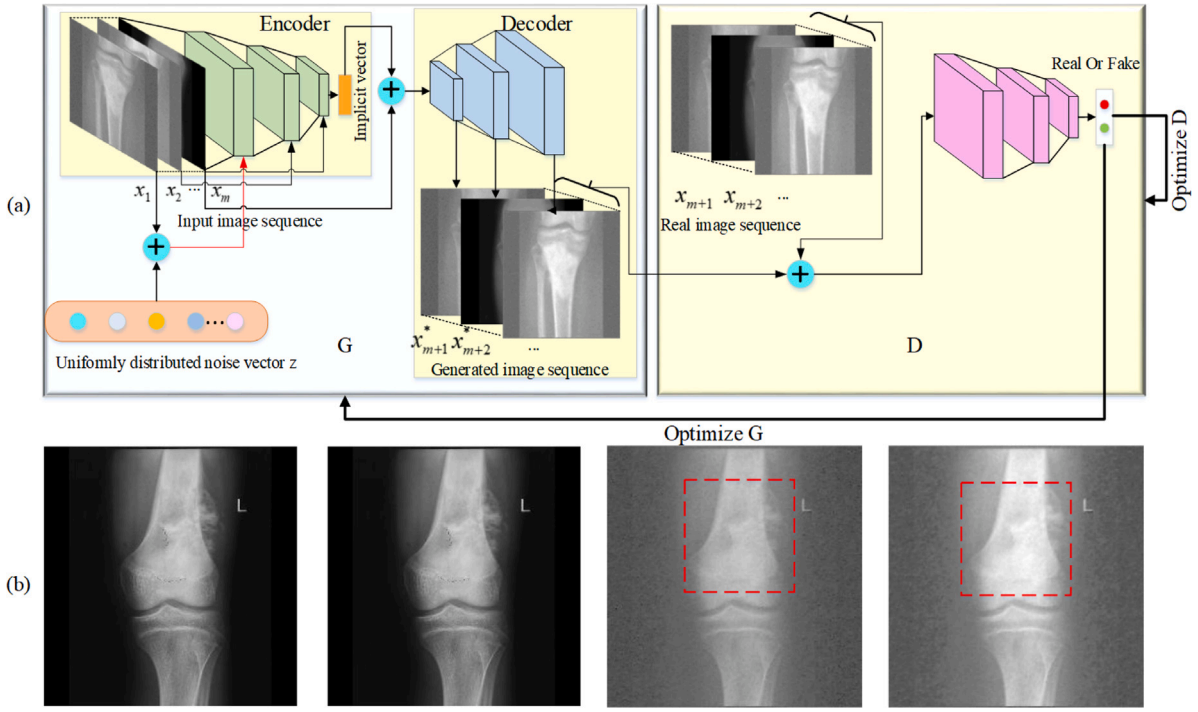

Fig. 5. Flowchart and output of the Conv-LSTM-GAN. (a) Optimization process of Generator and Discriminator. (b) Two on the left: the input time series images; Two on the right: the generated time series images.

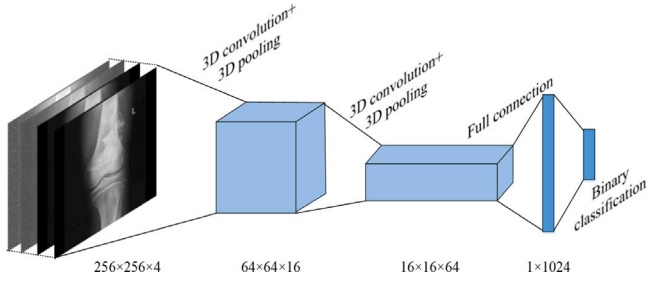

Fig. 6. 3D-CNN classification.

2.3.4. Objective function

We propose a combination of two objective functions to enhance the model's generation ability:

$$\min_G \max_D V(D,G) = E_{X_i^* \sim p_{data}(X_i^*)}[\log(D(X_i^*))] + E_{X_i \sim P_{data}(X_i), z \sim p_z(z)}[\log(1 - D(G(X_i, z)))] \quad (8)$$

where X_i^* is a real time series bone tumor image sequence from the training dataset, X_i is an input time series images sequence of the generated sequence, z is the initial state variable vector of the ConvLSTM cell in the Encoder of our proposed Conv-LSTM-GAN model.

2.4. Classification of time series bone tumor images

We send the generated images to the 3D convolutional neural network for binary classification of bone tumor necrosis rate. The threshold of the binary classification necrosis rate is 80%. The deep convolutional neural network includes two 3D convolutional layers, two 3D pooling layers, and two fully connected layers as shown in Fig. 6.

The generated time series bone tumor image of each patient is regarded as a whole sample and sent to the above mentioned deep convolutional neural network, and this classifier is trained based on the tumor necrosis rate label. The parameters of the network are trained and adjusted to extract the most representative features of the time series bone tumor lesion images.

3. Experiment settings and results

In this section, we first present the details of our bone tumor X-ray image dataset, then report the image generation results and experiment settings, finally analyze the classification results of the generated datasets by the 3D-CNN model in comparison with different datasets. All the experiments in this paper have been implemented in Tensorflow on NVIDIA RTX2080Ti GPU with 11 GB memory. About 1 GB of GPU memory is required to store model parameters, and about 10 GB of GPU memory is consumed when "batch size" is set to 24 during training with the input image size of 256 × 256. It takes about 12 h to train the model until the model converges, and it only takes less than 1 min for the trained model to detect a single image.

3.1. Dataset and classification threshold

We collect a dataset which contains bone tumor X-ray images of 119 patients from Peking University People's Hospital. For each patient, the bone tumor necrosis rate is diagnosed by orthopedic specialists. The dataset also includes time-series chemotherapy images of 33 patients, and each patient has 2 images before chemotherapy and 2 images in chemotherapy. The ratio of males to females in the above dataset is 78:41, and the age distribution is 10 to 83 years old. The tube voltage used in radiology is 100 kVp. In addition, we collected 130 extra bone contour images for Pix2Pix model to generate bone tumor images.

Especially, a large amount of evidence points out that the extremely low incidence of bone tumors has led to a small number of X-ray images collected. According to Centers for Disease Control and Prevention, Primary bone cancer is rare, which accounts for only about 0.5% of all cancers in the U.S. Seer, the authoritative source for cancer statistics in the US, points out that the rate of new cases of bone and joint cancer was 1.0 per 100,000 men and women per year. In 2018, an estimated

Table 1
Description of the Pix2Pix dataset.

Category	Training(pair)	Test(pair)	Generated(single)
0	68	130	130
1	51	130	130

3,450 new cases of primary bone cancer will be diagnosed in the United States (Miller et al., 2018). Taking into account the image specification issues caused by different X-ray equipment and the impact of noise interference, we can actually use very few images on the deep learning model. In China, we have collected as many bone tumor images as possible for the dataset.

X-ray images of bone tumors are already very scarce, and the samples with a necrosis rate higher than 90% is even rarer. In our dataset, If we use 90% as the classification threshold for necrosis rate, the available input sample number ratio is 40:79 (above 90%: less than 90%) with a threshold of 90% used as the classification criterion. The imbalance of two types of data with 90% as the threshold will cause instability of the deep learning model. To solve this problem, we set the threshold to 80%, which is feasible in the view of orthopedic specialists in Peking University People's Hospital.

3.2. Generation results of the Pix2Pix network

The input of the Pix2Pix model is the combination of paired images, which are real bone tumor images before chemotherapy with necrosis rate labels and their corresponding contour images. The Generator (G) of the model can learn the mapping relationship of each pair of contour image to the real pre-chemotherapy image after optimized by the Discriminator (D). Then the generator can generate pre-chemotherapy images with the same label as the real image based on a given contour image (Fig. 1a).

Specifically, the Pix2Pix model is applied twice to generate bone tumor images before chemotherapy with a necrosis rate below 80% (here after referred to as label 0) and the necrosis rate above or equal to 80% (here after referred to as label 1). The first Pix2Pix model is trained by 68 pairs of bone tumor images before chemotherapy with label 0 and their contour images. Similarly, 51 pairs of bone tumor images before chemotherapy with label 1 and their contour images serve as the input to train the second Pix2Pix model (Table 1).

Since the Discriminator can distinguish the generated and real lesion image as much as possible after the optimization of G and D, the trained Generator learns sufficiently credible lesion information from the normal contour to the target lesion image. Hence the target pre-chemotherapy lesion images with two necrosis rate category labels can be approximated by the trained Generator. An example of the generated pre-chemotherapy lesion image with label 0 is shown in Fig. 3b. The normal bone contour images are divided into two batches as the test data of two trained Generators of the models respectively. Then we can obtain the generated pre-chemotherapy bone tumor images with different categories(Table 1) .

3.3. Generation results of Conv-LSTM-GAN

After we obtain the preamble dataset of time series images in the Pix2Pix model, the generated prediction images can be obtained by the trained Generator of the Conv-LSTM-GAN model. In this paper, the input to this Generator consists of two parts: 119 single real pre-chemotherapy images and 260 pre-chemotherapy images generated by Pix2Pix as shown in Table 2. In this paper, the length of the input bone tumor image sequence is 2, and the output prediction sequence length is also 2. A generated image sequence in the test period is shown in Fig. 5b, whose input is a real pre-chemotherapy bone tumor image with a necrosis rate below 80%.

During the training process, the available training data are obtained from 33 patients with 4 X-ray images for each patient. We first expand this training data by traditional data enhancement methods, hence the amount of data is expanded to 7 times as described in Table 2. The input and output image size of Conv-LSTM-GAN are both 256 × 256. The time series length of the LSTM is set to 4.

As shown in Fig. 5a, the enhanced data serve as training data to the Generator, which is composed of an Encoder and a Decoder (Cho et al., 2014). We set a 1 × 256 noise vector which is uniformly distributed as the initial state. This vector together with the first image x_1 serves as the whole input at the first time stage to a ConvLSTM cell of the encoder as shown in Fig. 4. Then the input image of the next time stage is combined with the cell state information passed by the current time stage as the whole next input to the next ConvLSTM cell and so on (Fig. 5a). We complete the coding of the input image sequence in this way.

After the nonlinear transformation, an implicit vector representation of the input sequence is obtained as shown in Fig. 5a. Based on the implicit vector representation and the last original image x_m, the generated image at the next moment x^*_{m+1} can be predicted by the Decoder, where m is 2 in this paper. The image x^*_{m+2}, which is the fourth image can be generated in the same manner, based on the cell state passed in the time dimension of the ConvLSTM cell in the Decoder and the third generated image.

In the test period, the length of the test input image sequence is only 1 (x_1) for both the 119 single real pre-chemotherapy images and the 260 pre-chemotherapy images. In order to make the input sequence length greater than 1 to meet the coding requirements, all test input images are copied to obtain the second time stage images x_2 (Figs. 1a and 5a).

3.4. Classification results

3.4.1. Classification results using the generated image data

After applying the above two generation models, we have generated sufficient time series bone tumor images for the following binary classification task. This binary classification task takes the necrosis rate of 80% as a threshold, trains the 3D-CNN network to distinguish the image features of the two classes, and obtains results that are as consistent as possible with the labels.

To improve the robustness of the network, we augmented the dataset by traditional data augmentation methods before classification. Both the training and test sets are expanded to 7 times. In addition, the ratio of the training set to the test set is 2: 1 in all experiments for comparison. We use the average value of each indicator after a 3-fold cross-validation for all datasets as the final result. As illustrated in Table 3, the final test accuracy of the generated data is stable at 90%.

3.4.2. Further verification of the classification results

In addition to the accuracy of the classification, the test results of Receiver Operating Characteristic (ROC) are shown in Fig. 7, including the ROC curves of the 3-fold cross-validated test dataset and the average ROC value of the three curves. The abscissa indicates the false positive rate (FPR), while the ordinate shows the true positive rate (TPR). They are calculate as $TPR = TP/(TP + FN)$ and $FPR = FP/(TN+FP)$, where TP, FP, TN and FN represent the value of true positive, false positive, true negative and false negative respectively. The X-ray image with a necrosis rate below 80% indicates positive type, otherwise it is negative. We use the converged classification model to calculate the probability of "positive" prediction results for all test samples. By comparing it with the label, we can obtain the FPR and TPR values of each point on the ROC curve. Besides, we also denote the curve formed by the average ROC value as the center and the standard deviation of TPR as the upper and lower boundaries. It can be seen from Fig. 7 that the difference between most of the TPR values and their average value is small, so the TPR value we obtained can be considered as stable. More important, provided the ROC curve, the mean Area under Curve (AUC) result of 0.97 indicates our trained classifier performs well.

Table 2
Statistics of the generated time series images.

Model	Training data		Test data		Generated data
	Real	After augment	Real before chemotherapy	Generated by Pix2Pix	Output of the test data composed of two sources
Conv-LSTM-GAN	33×4	33×4×7	89×2	260×2	(89+260)×4

Table 3
Classification results of different training datasets.

Threshold	Train source	Train data	GAN model	Test source	Test data	ACC	REC	PRE	F1-score
80%	Real	89×7		Real	30×7	0.948	0.983	0.445	0.613
80%	Real+Generated	(89+260)×7	cycleGAN	Real	30×7	0.943	0.961	0.457	0.619
80%	Real+Generated	(89+260)×7	DCGAN	Real	30×7	0.921	0.941	0.583	0.720
80%	Real+Generated	(89+260)×7	Conv-LSTM-GAN	Real	30×7	0.916	0.948	0.796	0.865
90%	Real+Generated	(89+260)×7	Conv-LSTM-GAN	Real	30×7	0.877	0.868	0.782	0.823

3.4.3. Comparison of classification results from different data sources

Table 3 summarizes the classification results of two different datasets, which specifically involve the following indicators: accuracy (ACC), recall (REC) and precision (PRE). $ACC = (TP + TN)/(TP + FP + TN + FN)$, $REC = TPR$, $PRE = TP/(TP + FP)$. We apply the generated images to a 3D-CNN for the binary classification task as a comparative experiment. The dataset for contrast consist of real pre and post chemotherapy bone tumor images without any generated images. This dataset contains X-ray images from 119 patients with two images for each patient, where one is the pre-chemotherapy bone tumor image while the other is the post-chemotherapy one. The data of the experimental group is 379 generated bone tumor image sequences as shown in Table 3. Each generated sequence includes two real input images and two generated images.

In addition, we supplemented the experiment with 90% as the classification threshold, and compared the performance with the original experiment (Table 3). The results show that using 90% as the classification threshold can still achieve good performance, which proves the effectiveness of our proposed model. At the same time, due to the imbalance of the dataset, the performance of the model trained with 90% as the threshold is worse than the model we proposed with 80%.

3.4.4. Fine-grained classification control experiment

In order to further illustrate the superiority of the combination of bone tumor time series images and 3D-CNN model, we utilize a fine-grained classification model (Wang et al., 2018) instead of 3D-CNN to focus on the lesions. The classification results of this model is compared with the results of 3D-CNN applied to the same experimental data. The dataset for the control experiment comes from 119 real patient, including 61 people with a necrosis rate less than 80%, and 58 people with a necrosis rate greater than 80%. The time series images of each patient contain one image before chemotherapy and one image during chemotherapy.

In the fine-grained classification algorithm, the time series correlation cannot be directly exploited, therefore two time series images are directly spliced into one image to add the dimension of the plane image. After the convolution operation on the input image, a feature map representation of dimension $C \times H \times W$ (C, H, W represent the channel, height and width of images respectively)is obtained. An additional $C \times 1 \times 1$ feature extractor is trained to extract the feature map response of the specified parts of different channels. Finally, through the maximum pooling, the most responsive feature map part of the entire feature map can be determined.

The experimental results show that due to the lack of time series information, the performance of fine-grained classification model under the same dataset conditions is still weaker than the 3D-CNN time series classification model in this paper (see Table 4).

Table 4
Image generation evaluation metrics.

Model	Category	FID	KID
Conv-LSTM-GAN	Positive	47.93	0.0016
Conv-LSTM-GAN	Negative	47.98	0.0016
cycleGAN	Positive	163.90	0.1130
cycleGAN	Negative	176.55	0.1130
DCGAN	Positive	87.14	0.0476
DCGAN	Negative	89.73	0.0591

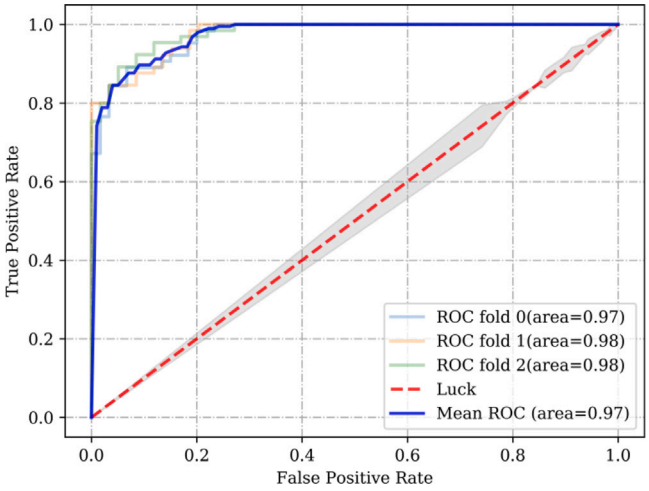

Fig. 7. ROC curves of the 3-fold cross-validated test datasets, their average ROC value (dark blue bold) and the standard deviation. (shaded region).

4. Discussion

A small sample bone tumor necrosis rate detection method based on deep learning proposed in this paper combines generative adversarial network and deep convolutional neural network. The results show that the method of generating images to expand the dataset and the approach of classifying the necrosis rate detection can effectively simulate the necrosis rate results obtained by biopsy.

4.1. Feasibility analysis of our scheme

From Table 3, it can be seen that the accuracy of classification using the generated dataset is 90%, which is good enough to indicate that the proposed model have learned a distribution similar to the original data. And we use cross-validation to further determine the stability of the results. On the other hand, each point on the ROC curve in Fig. 7 represents a pair of sensitivity and specificity under a certain threshold,

the threshold selection rule is to select positive probability value corresponding to the predicted value of each sample in order from large to small, which comprehensively reflects the changes in sensitivity and specificity of bone tumors at different critical values. The area value of the mean ROC is 0.97, which proves that our method has reliable diagnostic values. At the same time, the PRE value in Table 3 is 0.979. It shows that the false positive rate of the final test result is very low, which further proves the validity of the AUC value. The above results of our proposed 3D-CNN model in time series bone tumor images are close to the biopsy results indicating the effectiveness of our necrosis detection method. Besides, it effectively supports the reliability of using sequential medical images instead of biopsy operations.

The classification results of necrosis rate we obtained can assist doctors in diagnosing the effects of chemotherapy, which helps to improve the overall diagnostic level of bone tumor diseases in hospitals. Thus, patients can obtain better survival prognosis and quality of life.

Although important discoveries are revealed by our studies, there are also limitations. The overall classification results of the generated image performs well, and the false positive rate is very low when the model is stable. However, the true positive rate of 0.814 in Table 3 is not satisfactory, hence the future work needs to focus on improving the accuracy of detecting positive samples. More importantly, if the length of each time-series training dataset we can obtain is sufficiently long in the future, we can try to generate longer time series images for each patient, which will predict richer information about chemotherapy effects.

4.2. Time correlation advantage of generated images

It can be seen from Table 4 that the classification performance of the generated image is not as good as that of the real image. The reason may be that the features of the generated image is not as rich as the real image. Note that the time length of each sample of the generated image is 4, which is twice the length of the real image sequence(Table 2). Hence although the dataset of real pre and post chemotherapy bone tumor images has achieved a higher accuracy, the length of the time span is limited. So this dataset does not fully reflect the change of the lesion in the time dimension. Meanwhile, the number of samples of the generated image is more than 3 times that of the real image, which will be more conducive to the stability and robustness of the classification model.

In summary, the method we proposed for generating time series bone tumor images during chemotherapy reveals the temporal correlation of bone tumor images effectively. By learning the time series features in real bone tumor time series images, a subsequent time series of the input initial time-phase images can be generated. So we can predict the development trend of the lesion over time, which is important for the evaluation of the effect of chemotherapy. For example, reduction in flocculent lesions over time can be found in Fig. 5b, from which the effectiveness of chemotherapy can be inferred.

4.3. A significant increase in the number of generated images

In addition to using the time correlation between images in the chemotherapy phase to enhance the data, this paper also uses the Pix2Pix network to generate the pre-sequence images of the time sequence to further increase the number of datasets. This method provides a more powerful data support for the training stability and the generalization ability of the deep learning model. As shown in Table 1, the Pix2Pix model makes a major contribution in increasing the number of pre-sequence samples for time series images, which produces two types of pre-chemotherapy samples for a total of 260 virtual patients. In this way, the number of generated images is approximately 10 times that of the real 4-sequence image. Hence we complete the first step of bone tumor image data enhancement.

Moreover, we use the Pix2Pix model to learn the mapping relationship between the real pre-chemotherapy bone tumor image and its contour, so as to obtain the pre-chemotherapy image from the normal bone contour. As shown in Fig. 3b, in comparison with the normal bone image with no tumor, we can clearly see the generated lesion in the generated bone tumor image.

Prior works have noted the importance of biopsy-based necrosis rate detection (Goodfellow et al., 2014; Frid-Adar et al., 2018). However, this operation brings problems such as infections of invading tissues. In this paper, it has been demonstrated that this problem can be resolved by a novel approach for the detection of necrosis rate using time series X-ray images instead of biopsy. In reviewing the literature, no data was found on the link between time series X-ray images generated by GAN and the necrosis rate detection. On the other hand, the results of our study corroborate the findings of previous works of applying GAN in the medical image field. Last but not least, more research on this topic needs to be undertaken before the link between medical images and the necrosis rate detection is more clearly understood. At the same time, research on GAN in few-shot medical images will attract lots of interests.

5. Conclusions

Primary malignant bone tumors are a group of highly malignant tumors. The current method of detecting bone tumor necrosis rate relies on the invasive and time-consuming biopsy. In this study, we propose a non-contact method to detect bone tumor necrosis rate with few-shot X-rays images based on deep learning. It expands the few-shot X-rays by 10 times, and achieves the necrotic rate classification results similar to biopsy. Our method translate normal bone contour into bone tumor lesion image to expand the rare bone tumor dataset based on the mapping relationship between tumor lesion image and its contour. Then it exploits the time correlation from the real time series tumor images to generate the subsequent images in chemotherapy for given pre-chemotherapy images. Finally, the real and generated time series images are sent into 3D-CNN to obtain the final necrosis rate classification result. It is a new approach for the study of small sample medical images. In the future, We will further improve the scalability and generalization of our model so that it can be applied to solve more clinical problems, such as the tumor detection of CT, MRI and PET images.

CRediT authorship contribution statement

Zhiyuan Xu: Conceptualization, Methodology, Software, Visualization, Investigation, Writing – original draft. **Kai Niu:** Conceptualization, Methodology, Formal analysis, Writing – review & editing. **Shun Tang:** Validation, Resources, Data curation. **Tianqi Song:** Conceptualization, Methodology, Software, Writing – original draft. **Yue Rong:** Methodology, Writing – review & editing. **Wei Guo:** Resources, Data curation, Supervision. **Zhiqiang He:** Conceptualization, Methodology, Supervision, Writing – review & editing, Project administration.

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.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported by the National Key Research and Development Program of China (2021YFE0205300), Fundamental Research Funds for the Central Universities, China (2020XD-A02-3) and Capital's Funds for Health Improvement and Research (2020-2-4079).

References

Chen, X., Lian, C., Wang, L., Deng, H., Kuang, T., Fung, S.H., Gateno, J., Shen, D., Xia, J.J., Yap, P.-T., 2021a. Diverse data augmentation for learning image segmentation with cross-modality annotations. Med. Image Anal. 71, 102060.

Chen, J., Yang, G., Khan, H., Zhang, H., Zhang, Y., Zhao, S., Mohiaddin, R., Wong, T., Firmin, D., Keegan, J., 2021b. JAS-gan: Generative adversarial network based joint atrium and scar segmentation on unbalanced atrial targets. IEEE J. Biomed. Health Inf..

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

Dorfman, H.D., Czerniak, B., 1995. Bone cancers. Cancer 75 (S1), 203–210.

Duchman, K.R., Gao, Y., Miller, B.J., 2015. Prognostic factors for survival in patients with high-grade osteosarcoma using the surveillance, epidemiology, and end results (SEER) program database. Cancer Epidemiol. 39 (4), 593–599.

Fagioli, F., Biasin, E., Mereuta, O., Muraro, M., Luksch, R., Ferrari, S., Aglietta, M., Madon, E., 2008. Poor prognosis osteosarcoma: new therapeutic approach. Bone Marrow Transplant. 41 (2), S131–S134.

Ferrari, S., Briccoli, A., Mercuri, M., Bertoni, F., Picci, P., Tienghi, A., Del Prever, A.B., Fagioli, F., Comandone, A., Bacci, G., 2003. Postrelapse survival in osteosarcoma of the extremities: prognostic factors for long-term survival. J. Clin. Oncol..

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H., 2018. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 321, 321–331.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2672–2680.

Grignani, G., Palmerini, E., Ferraresi, V., D'Ambrosio, L., Bertulli, R., Asaftei, S.D., Tamburini, A., Pignochino, Y., Sangiolo, D., Marchesi, E., et al., 2015. Sorafenib and everolimus for patients with unresectable high-grade osteosarcoma progressing after standard treatment: a non-randomised phase 2 clinical trial. Lancet Oncol. 16 (1), 98–107.

Guan, Q., Chen, Y., Wei, Z., Heidari, A.A., Hu, H., Yang, X.-H., Zheng, J., Zhou, Q., Chen, H., Chen, F., 2022. Medical image augmentation for lesion detection using a texture-constrained multichannel progressive GAN. Comput. Biol. Med. 145, 105444.

Hansen, S., Gautam, S., Jenssen, R., Kampffmeyer, M., 2022. Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. Med. Image Anal. 78, 102385.

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D., 2022. Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 574–584.

Interiano, R.B., Malkan, A.D., Loh, A.H., Hinkle, N., Wahid, F.N., Bahrami, A., Mao, S., Wu, J., Bishop, M.W., Neel, M.D., et al., 2016. Initial diagnostic management of pediatric bone tumors. J. Pediatr. Surg. 51 (6), 981–985.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1125–1134.

Jin, T., Cui, H., Zeng, S., Wang, X., 2017. Learning deep spatial lung features by 3D convolutional neural network for early cancer detection. In: 2017 International Conference on Digital Image Computing: Techniques and Applications. DICTA, IEEE, pp. 1–6.

Kang, J.-W., Shin, S.H., Choi, J.H., Moon, K.C., Koh, J.S., kwon Jung, C., Park, Y.-K., Lee, K.B., Chung, Y.-G., 2017. Inter-and intra-observer reliability in histologic evaluation of necrosis rate induced by neo-adjuvant chemotherapy for osteosarcoma. Int. J. Clin. Exp. Pathol. 10 (1), 359–367.

Kumar, N., Gupta, B., 2016. Global incidence of primary malignant bone tumors. Curr. Orthopaedic Pract. 27 (5), 530–534.

Lee, J.A., Paik, E.K., Seo, J., Kim, D.H., Lim, J.S., Yoo, J.Y., Kim, M.-S., 2016. Radiotherapy and gemcitabine–docetaxel chemotherapy in children and adolescents with unresectable recurrent or refractory osteosarcoma. Jpn. J. Clin. Oncol. 46 (2), 138–143.

Liu, Q., Gaeta, I.M., Zhao, M., Deng, R., Jha, A., Millis, B.A., Mahadevan-Jansen, A., Tyska, M.J., Huo, Y., 2021. ASIST: Annotation-free synthetic instance segmentation and tracking by adversarial simulations. Comput. Biol. Med. 134, 104501.

Miller, K.D., Goding Sauer, A., Ortiz, A.P., Fedewa, S.A., Pinheiro, P.S., Tortolero-Luna, G., Martinez-Tyson, D., Jemal, A., Siegel, R.L., 2018. Cancer statistics for hispanics/latinos, 2018. CA: Cancer J. Clin. 68 (6), 425–445.

Ottaviani, G., Jaffe, N., 2009. The epidemiology of osteosarcoma. In: Pediatric and Adolescent Osteosarcoma. Springer, pp. 3–13.

Perez, L., Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621.

Sami, S.H., Rafati, A.H., Hodjat, P., 2008. Tissue necrosis after chemotherapy in osteosarcoma as the important prognostic factor. Saudi Med. J. 29 (8), 1124–1129.

Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D 404, 132306.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-c., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Adv. Neural Inf. Process. Syst. 28, 802–810.

Singh, R., Bharti, V., Purohit, V., Kumar, A., Singh, A.K., Singh, S.K., 2021. MetaMed: Few-shot medical image classification using gradient-based meta-learning. Pattern Recognit. 120, 108111.

Sun, C., Shrivastava, A., Singh, S., Gupta, A., 2017. Revisiting unreasonable effectiveness of data in deep learning era. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 843–852.

Wang, S., Liu, W., Wu, J., Cao, L., Meng, Q., Kennedy, P.J., 2016. Training deep neural networks on imbalanced data sets. In: 2016 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 4368–4374.

Wang, Y., Morariu, V.I., Davis, L.S., 2018. Learning a discriminative filter bank within a cnn for fine-grained recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4148–4157.

Zhan, B., Li, D., Wu, X., Zhou, J., Wang, Y., 2021. Multi-modal MRI image synthesis via GAN with multi-scale gate mergence. IEEE J. Biomed. Health Inf..

Zhang, Y., Yang, J., Zhao, N., Wang, C., Kamar, S., Zhou, Y., He, Z., Yang, J., Sun, B., Shi, X., et al., 2018. Progress in the chemotherapeutic treatment of osteosarcoma. Oncol. Lett. 16 (5), 6228–6237.