



Article

A Multimodal Auxiliary Classification System for Osteosarcoma Histopathological Images Based on Deep Active Learning

Fangfang Gou ¹, Jun Liu ^{2,*}, Jun Zhu ^{3,4} and Jia Wu ^{1,5,*}¹ School of Computer Science and Engineering, Central South University, Changsha 410083, China² The Second People's Hospital of Huaihua, Huaihua 418000, China³ The First People's Hospital of Huaihua, Huaihua 418000, China⁴ Collaborative Innovation Center for Medical Artificial Intelligence and Big Data Decision Making Assistance, Hunan University of Medicine, Huaihua 418000, China⁵ Research Center for Artificial Intelligence, Monash University, Melbourne, VIC 3800, Australia

* Correspondence: markliu8899@163.com (J.L.); jiawu5110@163.com (J.W.)

Abstract: Histopathological examination is an important criterion in the clinical diagnosis of osteosarcoma. With the improvement of hardware technology and computing power, pathological image analysis systems based on artificial intelligence have been widely used. However, classifying numerous intricate pathology images by hand is a tiresome task for pathologists. The lack of labeling data makes the system costly and difficult to build. This study constructs a classification assistance system (OHlcsA) based on active learning (AL) and a generative adversarial network (GAN). The system initially uses a small, labeled training set to train the classifier. Then, the most informative samples from the unlabeled images are selected for expert annotation. To retrain the network, the final chosen images are added to the initial labeled dataset. Experiments on real datasets show that our proposed method achieves high classification performance with an AUC value of 0.995 and an accuracy value of 0.989 using a small amount of labeled data. It reduces the cost of building a medical system. Clinical diagnosis can be aided by the system's findings, which can also increase the effectiveness and verifiable accuracy of doctors.

Keywords: medical system; active learning; auxiliary diagnosis of osteosarcoma; histopathological images; classification

MSC: 68T01

Citation: Gou, F.; Liu, J.; Zhu, J.; Wu, J. A Multimodal Auxiliary Classification System for Osteosarcoma Histopathological Images Based on Deep Active Learning. *Healthcare* **2022**, *10*, 2189. <https://doi.org/10.3390/healthcare10112189>

Academic Editor: Tin-Chih Toly Chen

Received: 25 August 2022

Accepted: 28 October 2022

Published: 31 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The incidence of osteosarcoma ranks first among malignant bone tumors [1]. As a highly malignant tumor, it is extremely harmful to human health. Adolescents and the elderly are the most affected groups. Although the incidence of osteosarcoma is only 3 to 10 per million, it accounts for 40% of primary malignant bone tumors [2]. Osteosarcoma has high mortality and morbidity rates, especially in developing countries [3]. Due to the large population, the death toll is high. Early screening and diagnosis of osteosarcoma are critical to improving patient survival [4].

Routine detection methods for osteosarcoma include symptomatic examination, imaging examination, and pathological examination [5]. Since many tumors do not have typical imaging features, it is difficult to determine the nature of the mass by relying solely on imaging examinations, and it is impossible to determine whether a patient has a bone tumor [6]. Osteosarcoma has multiple subtypes with different pathological features [7]. Therefore, the pathological examination is regarded by all experts as the “gold standard” in the diagnosis of osteosarcoma.

Digital pathology images are obtained by scanning pathology slides. The data volume of pathological images is very large, and one histopathological section contains millions of

cells [8]. The complex pathological features of osteosarcoma require extremely high professionalism and the experience of pathologists [9]. The number of specialized pathologists in a hospital is limited, and each specialist processes many slices in a day [10]. This is a very time-consuming and complicated job. In addition, the huge workload also makes doctors overworked [11]. Diagnosis results are subject to subjective influence, which can easily lead to missed diagnosis and misdiagnosis. Therefore, it is extremely important to develop a decision analysis system for histopathological images to assist pathologists in diagnosing osteosarcoma and alleviate the problems existing in hospitals.

With the development and popularization of artificial intelligence technology in the medical field, the neural network plays an increasingly important role in the field of medical image analysis with its powerful feature extraction ability [12–14], such as auxiliary staging of lung cancer, MRI segmentation of osteosarcoma [8], and other applications. However, many methods based on supervised learning still face the following problems when analyzing histopathological images of osteosarcoma:

- (1) The system construction lacks enough labeled data, and the initial training set is insufficient [15]. Insufficient labeled data will greatly limit the performance of intelligent diagnostic systems based on supervised learning.
- (2) Lots of unlabeled pathology images. The data volume of histopathological images in hospital databases is very large. However, labeled data are scarce.
- (3) The cost of labeling samples is high. Due to the complexity of histopathological sections of bone and flesh, non-professionals cannot distinguish them. Only relying on professional pathologists to mark samples is costly and difficult.

To solve the problem that there are few pathological images of osteosarcoma labeled during model training, and data are difficult to obtain, active learning selects informative samples from unlabeled data and sends them to pathologists for labeling [16,17]. Such datasets are then trained to maximize model performance within a limited labeling budget. The methods of minimum confidence, marginal sampling, entropy, etc., are all based on the uncertainty of the model for the sample [18]. Such methods are biased and complex, which can affect the representativeness of the sample. Even though these active-learning techniques lessen the need for labeled training data, there are still some issues with relying solely on one tactic. The uncertainty-based metric alone is overconfident in sample selection, and the sample's prior information is likely to be disregarded, resulting in unstable model performance. It may even be worse than the random sampling strategy. The selection of more samples with less information could result from only taking sample diversity into account, which would raise the cost of labeling.

Based on this, this study proposes an active-learning-based intelligent assisted classification system for osteosarcoma histopathological images (OHICsA). The system effectively improves the annotation gain of osteosarcoma pathological images by actively acquiring the most characteristic pathological images as annotated samples. First, we train the network to reconstruct more pseudo-samples using the labeled dataset. The more insightful pathological images are then chosen from the unlabeled samples to be labeled using a selection strategy. These labeled images are then put together to retrain the network. Through this iterative training, selection, annotation, and retraining method, the OHICsA method effectively alleviates the problems of high cost and difficulty in labeling osteosarcoma pathological images. In the clinical management of osteosarcoma, the results are used as a diagnostic aid, which can increase the precision and effectiveness of diagnosis and decrease the incidence of missed and misdiagnosis cases.

The main contributions of this research include the following:

- (1) This paper builds an osteosarcoma classification system based on active learning and histopathology images. The system utilizes the fitting ability of deep neural networks to achieve image classification. It not only improves the efficiency of pathologists but also contributes to the objective accuracy of diagnosis.
- (2) We combine generative adversarial networks (GAN) and active learning to obtain more high-value samples; GAN constructs high-quality pseudo-samples, and active

learning obtains more unlabeled images with the “richest” information. This approach maximizes the performance of the model with a small number of labeled images. It effectively reduces the cost of labeling images, which is valuable in areas where there are not enough pathologists.

- (3) A new strategy that integrates the query sample class diversity and uncertainty is proposed. It effectively reduces the problems of sampling bias that often result from uncertainty-based sampling alone and the increased tagging costs that can result from considering diversity alone. This approach effectively improves the ability of the model to select samples.
- (4) Experimental results with osteosarcoma pathology images show that our system can achieve good classification performance with only a small number of labeled images. The results processed by the system can be used as an objective reference for clinical diagnosis and improve the detection accuracy of physicians.

2. Related Works

With the advancement of deep-learning technology and the improvement of big data computing capabilities, there are now many medical assistant decision-making methods. We introduce some methods.

The effectiveness of deep-learning methods in disease identification and detection was first demonstrated [19,20]. He et al. [21] proposed a multi-view uncertainty measurement method. By performing a softmax operation on each middle layer, the final uncertainty measurement takes into account the parameters of the middle layer of the model. Sunpreet Kaur Nanda et al. [22] developed an intelligent medical system that can quickly and accurately detect whether people are wearing a mask or not by using YOLO and CNN models. It lays the foundation for developing intelligent medical image systems for the detection of various disease and tumor abnormalities. Shivani G. Dharmale et al. [23] used machine-learning techniques to identify faults and anomalies in medical sensors from the perspective of medical sensors. They are mainly implemented through the random forest, k-neighborhood algorithm, and analysis based on geographical and temporal information. Neha Yadav et al. [24] proposed a hemorrhoid detection method (LeakyReLU) based on the HSV model. It includes deep-learning techniques such as K-Means, texture analysis, an HSV-based segmentation model, and activation function with Leaky ReLU. The method achieves a high detection accuracy.

When the task lacks sufficient samples, data augmentation can be performed by generating the model. Commonly used generative models include GAN and VAE. Many outstanding generative models can generate samples that conform to the original distribution without introducing excessive noise, improving the robustness of the model without reducing its performance of the model. Generative models are widely used in data augmentation [25]. They are also closely integrated with active learning. When lacking unlabeled samples, Zhu et al. [26] used GAN to generate samples. These samples are selected based on a specific strategy and handed to experts for labeling. In [27], Tran et al. combine the reconstructed samples of the generated model with the selected samples and train the ACGAN model and classifier. In both data augmentation [28] and active learning, samples need to be selected. The main difference is that the former is selected to be similar to the original sample, because the generative model may produce a lot of data that do not conform to the original data distribution, while the latter is to try to select the sample that is not similar to the original sample to reduce epistemic uncertainty.

K George et al. [29] utilized a deep-learned kernel feature classifier to detect breast cancer from histopathology images. V Gupta et al. [30] implemented an adenocarcinoma histopathology image classification based on the ResNet model. P Wang et al. [31] proposed a pathological image classification method (Cd-dtffNet). These methods have a very good role in promoting the diagnosis of breast cancer. Subsequently, to more accurately predict the size of tumors in breast cancer, Ophir Nave et al. [32] combined mathematical models

and machine-learning algorithms. The study shows that employing mathematical models to discover image features can result in more precise forecasts.

Similarly, Sekhar et al. [33] used transfer learning to classify brain tumors in MRI images into three categories (glioma, meningioma, and pituitary). Wu et al. [34] designed an algorithm for segmenting lung nodules in chest CT images. The algorithm uses a 3D conditional random field (CRF) to continuously optimize training samples during training. The optimized samples accelerate the convergence of the 3D-UNet model and reduce the training time. Gur et al. [35] proposed a novel unsupervised method for vessel segmentation in pathological images. Based on edgeless morphological active contours, the method can label datasets and be applied to similar datasets.

Many researchers use computer vision technology to realize the auxiliary diagnosis of osteosarcoma. Long ago, Brian D Ragland et al. [36] performed a detailed analysis of osteosarcoma from a cytogenetic and biomolecular perspective, demonstrating that no cytogenetic and molecular markers have been identified to determine the diagnosis of osteosarcoma. After that, Rashika Mishra et al. [37] divided tumors into three groups based on CNN: living tumors, necrotic tumors, and non-tumors, and they also came up with a method to determine the amount of necrosis in pathological images. Anisuzzaman et al. [38] used transfer learning to supervise the metastases of bone tumors in bone and soft tissue. Yedong Shen et al. [39] proposed an OSGABN method based on guided aggregate bilateral networks. It mainly segmented MRI images through a fast bilateral segmentation network (FaBiNet), and the results showed that the method has high accuracy. Yu Fu [40] used a somatic neural network for the automatic classification of histological images of osteosarcoma, helping pathologists to improve the accuracy of diagnosis. To further improve the classification accuracy, Barzekar et al. [41] and D'Acunto et al. [42] used a machine-learning approach to differentiate between malignant and benign tumor cells.

3. System Design

With the rapid development of large-scale data-computing capabilities and the continuous improvement of hardware technology, the research and application scope of artificial intelligence in the medical field continues to expand [43–45]. Especially in the field of medical images, AI uses its computer vision technology to realize lesion identification and labeling, image segmentation, feature extraction, comparative analysis, 3D reconstruction, etc. [46–49]. Due to the complexity of histopathological images of osteosarcoma, it is difficult for non-professionals to distinguish them [50,51]. The development of an automated and accurate histopathology image analysis system is of great importance to assist pathologists and clinical specialists in their work. As shown in Figure 1, the processing flow of pathological data of patients with osteosarcoma includes several parts: patient, pathology, doctor, hospital cloud service center, and intelligent auxiliary classification system. The specific process includes: (1) The pathology department sends pathology images to the patient and uploads them to the cloud service center. (2) The clinical diagnostic features of the patient are sent to the doctor and fed back to the cloud service center for storage. (3) The doctor gives the diagnosis to the patient and uploads it to the cloud service center. (4) Patients, pathology departments, physicians, and intelligent classification assistance systems receive data from the cloud service center continuously. (5) The results of the system analysis are provided to the physicians and pathology departments as an auxiliary reference for diagnosis and uploaded to the cloud service center. (6) The pathology department and the doctor will feed the results of the diagnosis to the intelligence system again.

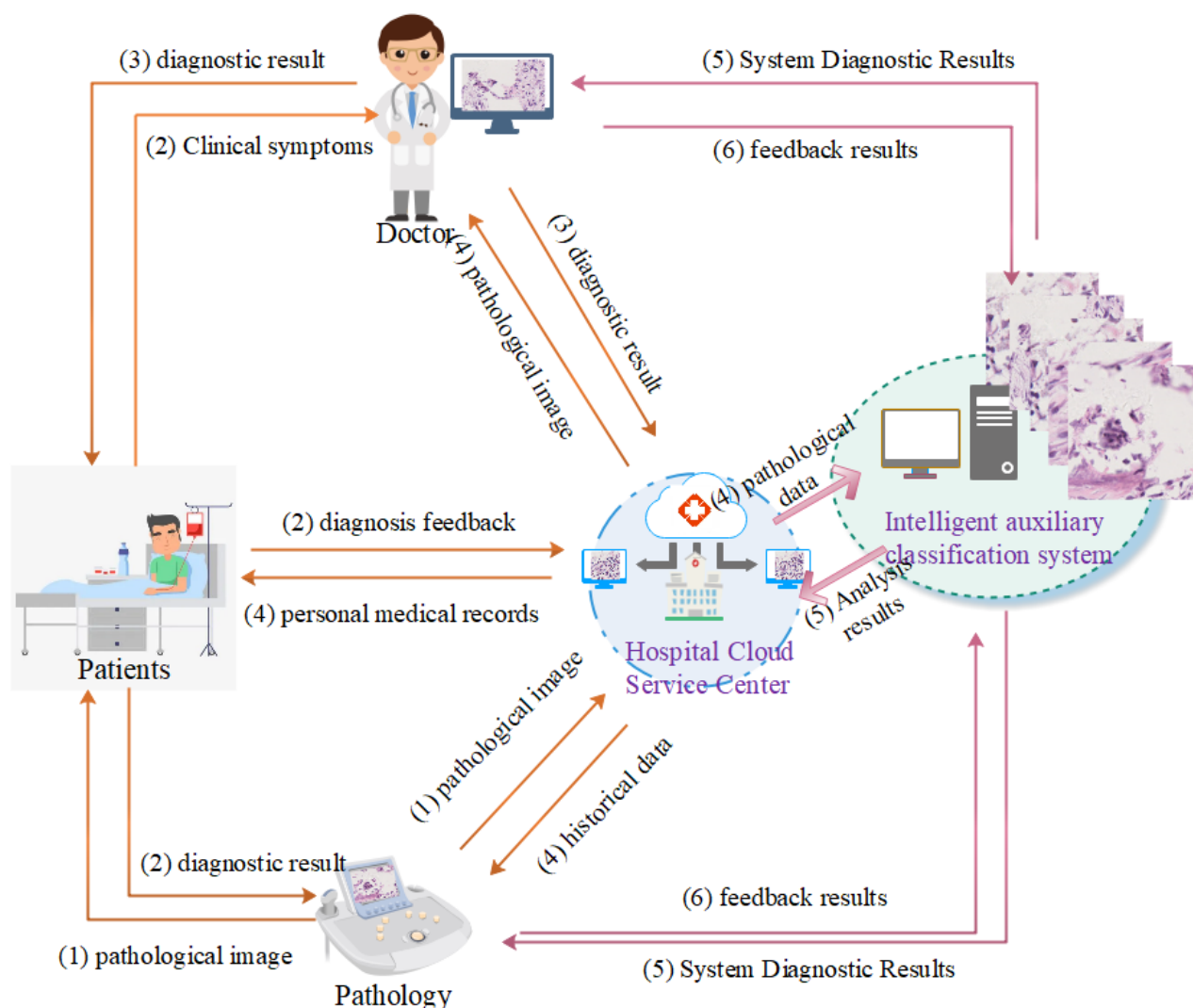


Figure 1. Patient pathology data decision-making process. The orange arrow is the workflow of the original medical system. The purple arrow is the workflow after adding the intelligent classification assistance system.

However, in the diagnosis and treatment of osteosarcoma, the only way to label pathological data is to rely on professional pathologists. Therefore, labeling data is expensive and difficult. Although the application of active learning effectively alleviates the problem of requiring a large number of labeled samples for supervised learning model training, existing methods based only on uncertainty are overconfident and often result in sampling bias and data redundancy. Considering only diversity in turn leads to an increase in the cost of labeling, after considering the above issues together, this study proposes an osteosarcoma auxiliary diagnosis system based on active learning and generative adversarial networks (OHICsA). It mainly improves the classification performance through the powerful fitting ability of the deep neural network. We combine active learning and generative adversarial networks. Active learning integrates the diversity and uncertainty of samples to obtain more informative samples. The adversarial model uses its powerful data generation capabilities to extend the original training set. This approach drives the model to obtain faster performance gains and maximizes the performance of the model using a small number of high-quality labeled samples.

The common symbols and their meanings in this section are shown in Table 1.

Table 1. Partial abbreviation vocabulary and basic symbols interpretation table.

Symbol	Meaning
x	The true sample
$\mathcal{D}^{\mathcal{L}}$	Labeled data
$\mathcal{D}^{\mathcal{U}}$	Unlabeled data
x_{re}	Reconstructed samples
x_p	Pseudo-samples
$KL(\cdot)$	Kullback–Leibler divergence operator
L_D	The loss of discriminator D
L_C	The loss of classifier \mathcal{M}
L_{KL}	The difference between distribution $\mathcal{N}(0, 1)$ and posterior probability distribution of z
L_G, L_{GD}, L_{GC}	The loss of generator G
$unc(x, \mathcal{M})$	The uncertainty of model \mathcal{M} to sample x
$H[\cdot]$	Shannon entropy
T	The number of dropout iterations
\hat{p}_c^t	The probability that the sample is the $c - th$ class in the $t - th$ round of prediction
div_i	The diversity score of the sample i
$d_{i,c}$	The distance from sample i to category c
θ	Trade-off factors for diversity and uncertainty

3.1. Models Design

In this section and the following section, we will describe how the proposed framework works in detail. The patterns in histopathological images are complex and require large amounts of labeled data to train the model [52–54]. In order not to increase the annotation burden, we constructed a new classification model that combines active learning and generative adversarial models, choosing a conditional variational self-encoder (CVAEGAN) as the main framework. CVAEGAN combines generating adversarial networks (GAN) and variable score auto-encoder (VAE). It combines the ability of active learning to acquire the most informative samples and the ability of GAN to learn data distribution to generate real pseudo-samples simultaneously. We achieved faster model performance improvements than when employing only the uncertainty-based approach by feeding both selected and generated samples into the fundamental model for training. We used a convolutional neural network as a model for the classification task. Osteosarcoma pathology images and their classes are denoted by x and y , respectively. By using the data generation capability of the GAN model to extend the original training set, the system obtains faster performance improvements. This means that we can obtain higher performance with the same labeling cost and the same training method. As shown in Figure 2, in this regard, the structure of each module is as follows:

Auto-encoder: It is divided into two parts: encoder E and generator G . E is in charge of encoding and mapping the sample x to a potential vector representation z . Then, $z = E(x)$, and the generator G is responsible for using E or reconstructing the original image based on the encoding result of z , to generate a new image x_{re} . Meanwhile, z_p is randomly sampled from a specific distribution (we use here a normal distribution with mean 0 and variance 1, denoted as $N(0,1)$) and reconstructed by the generator G to generate the pseudo-image x_p . Thus, we consider that the learned low-dimensional variable z obeys the same distribution over the labeled and unlabeled samples. The auto-encoder is trained using all labeled data

as well as unlabeled data. In turn, the model is guaranteed to process all samples fairly and without bias.

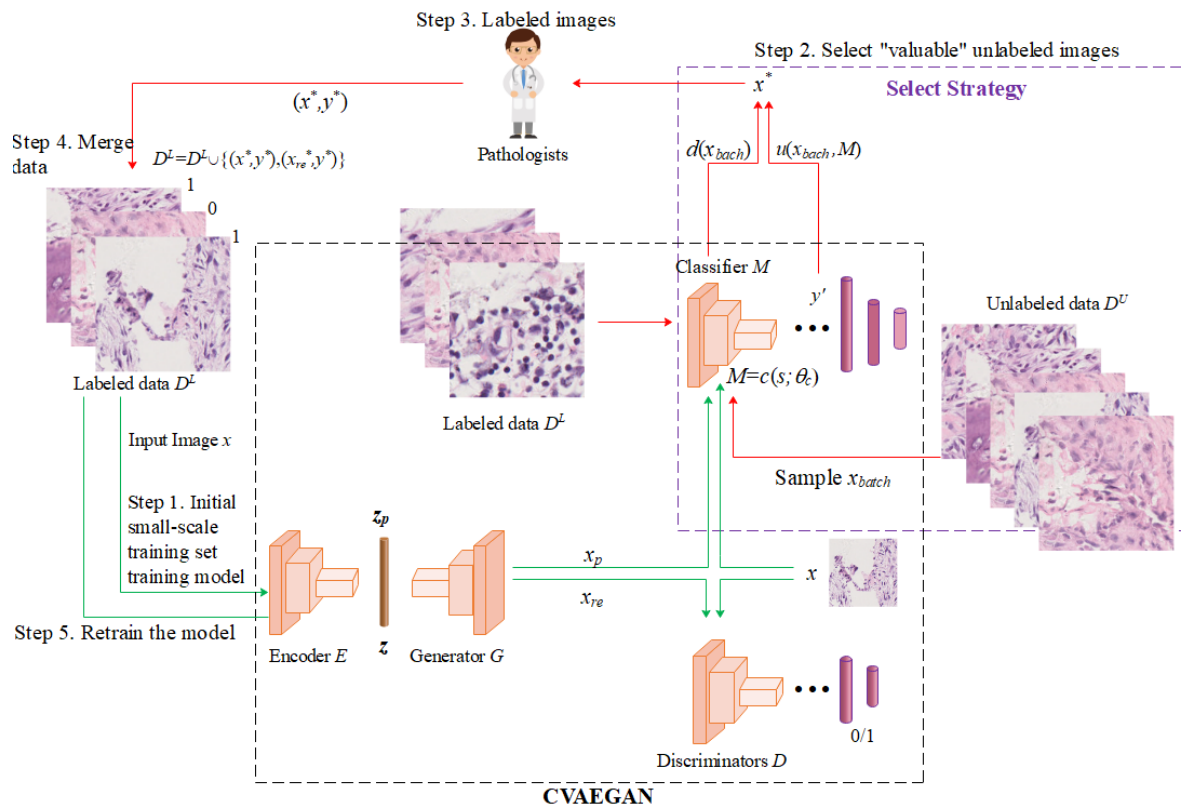


Figure 2. The detailed structure of the proposed deep active-learning framework. The boxes in the figure represent models, and the ellipses represent database entities. Oracle is usually a human expert. The arrows represent the data flow, the green arrows are the training process, and the red arrows are the filtering process. E , G , D , and M form the CVAEGAN model.

Discriminator D : It measures the difference in distribution between labeled data \mathcal{D}^L and unlabeled data \mathcal{D}^U and determines whether the image is a real image or a generated image. That is, the discriminator D is used to distinguish x from x_p and x_{re} . The large of \mathcal{D}^U is much larger than that of \mathcal{D}^L in the osteosarcoma pathology image.

Classifier M : In histopathological image classification, M classifies the labeled images to determine their classes. The parameters of the classifier are denoted by θ . M can be any model, which greatly increases the applicability of the framework.

Figure 2 in this paper shows the proposed deep active-learning model. Its main process consists of four steps:

- (1) Step 1: an initial small-scale labeled training set \mathcal{D}^L is used to train the neural network model, including encoder E , generator G , discriminator D , and classifier M .
- (2) Step 2: the most informative sample is selected from the large number of unlabeled samples.

The selection strategy will take into account both the class diversity of the samples and the uncertainty of the model on the samples. Moreover, it uses generative adversarial networks to reconstruct the selected samples to increase the diversity of the data. A subset of x_{batch} is sampled according to a specific strategy. The strategy will be described in Section 3.2.

- (3) Step 3: Pathology expert labeled data. After the pathology sample is labeled by the expert, it is put into the training set along with the reconstructed sample (x_{re} in Figure 2) for the next iteration of training.

- (4) Step 4 and Step 5: the selected sample data are merged into the original labeled dataset to retrain the classifier.

$$\mathcal{D}^{\mathcal{L}} = \mathcal{D}^{\mathcal{L}} \cup \{(x^*, y^*), (x_{re}^*, y^*)\} \tag{1}$$

where $\mathcal{D}^{\mathcal{L}}$ is the labeled data, (x^*, y^*) is the pathological sample labeled by the pathologist, and (x_{re}^*, y^*) is the real sample generated by automatically coding it.

With this iterative training, selection, annotation, and retraining approach, our proposed framework achieves similar performance compared to training with a full labeled dataset but using only about half of the labeled data. This significantly reduces the cost of building such a medical system.

Under the above description, the mathematical model describing the classification of osteosarcoma pathology images is a system based on the CVAEGAN model. In the model, our goal is to minimize the following loss functions to obtain superior classification results:

$$L_{CVAEGAN} = L_C + L_D + \lambda_1 L_{KL} + \lambda_2 L_G + \lambda_3 L_{GD} + \lambda_4 L_{GC} \tag{2}$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are hyperparameters to trade off different losses. $L_C, L_D, L_{KL}, L_G, L_{GD}, L_{GC}$ are losses for different optimization targets. The detailed expressions are given in Equations (3)–(8).

$$L_D = -(\log(D(x)) + \log(1 - D(x_{re})) + \log(1 - D(x_p))) \tag{3}$$

L_D is the loss of discriminator D . It aims to let D distinguish the true samples x from reconstructed samples x_{re} and generated pseudo-samples x_p .

$$L_C = -\log(P(\mathcal{M}(y|x))) \tag{4}$$

L_C is the loss of classifier \mathcal{M} . It aims to force classifier \mathcal{M} to output the true label y when giving a true sample x . In addition, the network tries to minimize the Softmax loss:

$$L_{KL} = KL(q(z|x, y) || \mathcal{N}(0, 1)) \tag{5}$$

L_{KL} denotes the difference between distribution $\mathcal{N}(0, 1)$ and posterior probability distribution of z when given x, y denoted by $q(z|x, y)$. $KL(\cdot)$ is the Kullback–Leibler divergence operator. This will help z , the output of encoder E obeys a normal distribution, and the sampled pseudo-latent vector z_p comes from the same sample distribution with true samples. In other words, it can help to make z_p more real, thus contributing to generating a realistic pseudo-image x_p , where $x_p = G(z_p)$.

$$L_G = \frac{1}{2} (\|x - x_{re}\|_2^2 + \|f_D(x) - f_D(x_{re})\|_2^2 + \|f_M(x) - f_M(x_{re})\|_2^2) \tag{6}$$

L_G denotes a part of the loss of generator G . It consists of three parts. The first one is the loss between the true sample (x) and the reconstructed sample (x_{re}). $x_{re} = G(E(x))$. This part forces the encoder and generator coordinate to reconstruct the image more precisely. In the second part and third part, $f_M(\cdot)$ and $f_D(\cdot)$ mean the second-to-last layer’s output. The latter two parts evaluate the difference of the original sample (x) and the reconstructed sample (x_{re}) in the view of discriminator D and classifier \mathcal{M} .

$$L_{GD} = \frac{1}{2} \left\| \frac{1}{m} \sum_i^m f_D(x) - \frac{1}{m} \sum_i^m f_D(x_p) \right\|_2^2 \tag{7}$$

$$L_{GC} = \frac{1}{2} \left\| \frac{1}{m} \sum_c f_M^{c_i}(x) - \frac{1}{m} \sum_c f_M^{c_i}(x_p) \right\|_2^2 \tag{8}$$

L_{GD}, L_{GC} are also losses of generator G . Different from L_G , the force generator generates pseudo-samples x_p that have close class centers to true samples x in the view of

discriminator D and classifier \mathcal{M} . The latter two items in L_G focus on the singular sample's difference between true sample x and pseudo-sample x_{re} , while L_{GD} , L_{GC} focus on the difference among batch samples.

Through the collaborative training of multiple models, it is ensured that the generator G generates high-quality samples and classifier \mathcal{M} also performs well. The main reason for choosing a conditional variational self-encoder (CVAEGAN) as the model in this paper is that it generates realistic images with a strong fit to the data distribution of the original samples and it can reconstruct x and generate x_{re} . x_{re} is similar to but different from the original image x . It can be regarded as the original image with added noise. Adding it to the training set can increase the number of the training set and help train a classifier that is robust to noise.

3.2. Selection Strategy

We visually verify the proposed strategy to obtain an intuitive sense of effectiveness. Multi-dimensional scaling (MDS) is a commonly used solution for dimensionality reduction and visualization of high-dimensional data [55,56]. It can map distances between samples in high dimensions into distances in low dimensions and keep the relative length relation and order. We use the second-to-last fully connected layers in the classifier to perform dimensionality reduction on the data and then perform MDS on the dimensionality reduction result. The visualization result is shown in Figure 3.

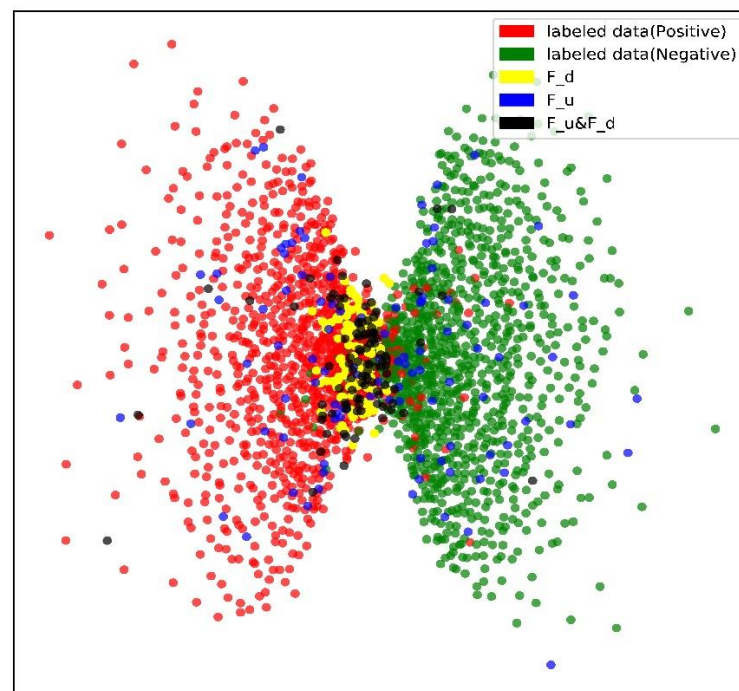


Figure 3. MDS visualization of three different strategies. Red points represent negative labeled data, and green points represent positive labeled data. The boundary between the two classes is basically clear. Blue points (F_u) represent samples selected based on uncertainty strategy in an iteration, yellow points (F_d) are samples selected based on diversity strategy (according to (12)), and black points ($F_u \& F_d$) represent selected samples that combine uncertainty and diversity.

Samples near the decision boundary often make the model make mistakes, so selecting and annotating the first will help speed up the improvement of model performance. This is also one of the motivations for us to introduce diversity. Many samples selected based on uncertainty strategy are falling at the center of the classes, not near the decision boundary of the sample, and this may result in a lot of not very effective annotation. The black points based on the two selection methods contain points near the decision boundary (where the

diversity score is high) and also contain points away from the decision boundary (where the uncertainty score is high).

3.2.1. Uncertainty

The uncertainty base strategy is widely used in active learning. It can be evaluated through the output y' of classifier \mathcal{M} . Bayesian active learning (BALD) [57] was proposed by researchers. It follows (9):

$$unc(x, \mathcal{M}) = H[y' | x, \mathcal{D}^{\mathcal{L}}] - \mathbb{E}_{\theta \sim p(\theta | \mathcal{D}^{\mathcal{L}})} H[y' | x, \theta] \tag{9}$$

where $unc(x, \mathcal{M})$ means the uncertainty of model \mathcal{M} to sample x , and $H[\cdot]$ is Shannon entropy. The Monte Carlo (MC) dropout [58] method measures model uncertainty by obtaining the output of multiple rounds of predictions. The expression is as follows (10):

$$unc(x, \mathcal{M}) \approx - \sum_c \left(\frac{1}{T} \sum_t \hat{p}_c^t \right) \log \left(\frac{1}{T} \sum_t \hat{p}_c^t \right) + \frac{1}{T} \sum_{c,t} \hat{p}_c^t \log(\hat{p}_c^t) \tag{10}$$

where T is the number of dropout iterations, and \hat{p}_c^t represents the probability that the sample is the c -th class in the t -th round of prediction.

3.2.2. Diversity

A selection strategy based on BALD is not quite sufficient and may overfit on the selected samples [59,60]. Therefore, a diversity strategy is introduced to further accelerate performance improvement. As shown in Figure 3, the classifier is often easy to distinguish samples at the center of the category and prone to making mistakes for points in the intersection area or disagreement area. For example, in handwritten digit recognition, the number one and the number seven are similar and may have more large overlapping areas in high dimension manifold [61]. Selecting samples located in the disagreement area will help to acquire clearer decision boundaries, thus obtaining a greater performance improvement.

We believe that samples at the boundary of the category cluster tend to confuse the classifier, so we formulated a diversity score to quantify this property, and its expression follows (11):

$$div_i = \sum_k softmax(Z(d_{i,c})) \log(softmax(Z(d_{i,c}))) \tag{11}$$

where div_i represents the diversity score of sample i , $Z(\cdot)$ represents the Z-score operation, $softmax(\cdot)$ represents the softmax operation, and $d_{i,c}$ represents the distance from sample i to category c . k is the truncation coefficient. After sorting $d_{i,c}$ in ascending order, the first k elements are intercepted and used to calculate div_i . $d_{i,c}$ is measured by (12):

$$d_{i,c} = \left| \frac{1}{|D_c|} \sum_{j \in D_c} z'_j - z'_i \right|_2 \tag{12}$$

where $d_{i,c}$ represents the distance from sample i to category c (we use the centroid of all samples as the category center), z'_i is the latent vector of sample i , D_c is the marked sample set of class c , and $|D_c|$ is the sample size. The larger the div_i , the more balanced the distance of the sample to all class centroids and the more likely the sample is in the intersection area, through comprehensive consideration of the uncertainty and diversity of the sample, that is, according to (13):

$$score = Scl(unc(x_i, \mathcal{M})) + (1 - \theta) Scl(div_i) \tag{13}$$

where score represents the comprehensive score of the sample, $Scl(\cdot)$ is the min-max scaling operation, and θ is the trade-off coefficient to weigh the effects of the two.

The entire model can be summarized as Algorithm 1.

Algorithm 1. Procedure of the proposed framework.

Require: labeled dataset $\mathcal{D}^{\mathcal{L}}$, unlabeled dataset $\mathcal{D}^{\mathcal{U}}$
Ensure: classifier \mathcal{M}

- 1: Construct and initialize class classifier \mathcal{M} , encoder E , generator G , real/fake classifier D
- 2: Train \mathcal{M}, E, D, G with $\mathcal{D}^{\mathcal{L}}$
- 3: **while** not satisfy condition **do**
- 4: randomly sample x_{batch} from $\mathcal{D}^{\mathcal{U}}$
- 5: calculate $score_i \forall i \in x_{batch}$ according to (13)
- 6: select n sample with the highest score
- 7: $D_{inc} \leftarrow \emptyset$
- 8: **for** x^* in x_{re} **do**
- 9: query the label of x^* from Oracle and get $y^* \quad x_{re}^* \leftarrow G(E(x^*))$
- 10: $x_{re}^* \leftarrow G(E(x^*))$
- 11: $D_{inc} \leftarrow D_{inc} \cup \{(x_{re}^*, y^*), (x^*, y^*)\}$
- 12: $\mathcal{D}^{\mathcal{L}} \leftarrow \mathcal{D}^{\mathcal{L}} \cup D_{inc}$
- 13: **end for**
- 14: retrain \mathcal{M}, E, D, C with $\mathcal{D}^{\mathcal{L}}$ by (2)–(8)
- 15: **return** \mathcal{M}

Assuming that the time for a round of forwarding propagation on b samples is t_f , and the time for backward propagation is t_b , and in each iteration, we train classifier e epochs. Then, the time consumed to train a normal classifier using n training samples is about $\frac{n}{b}e(t_f + t_b)$.

For the proposed strategy, we need to train four models: encoder, decoder, discriminator, and classifier in each query iteration. Assuming that the number of parameters of the four models is similar, the training time is about $4\frac{n}{b}e(t_f + t_b)$. When the training is over in an iteration, we need to query the unlabeled samples. When using MC Dropout to calculate the uncertainty, assume that the dropout is performed T times, and each time q samples are selected from p samples for labeling, the time cost is $4\frac{n}{b}e(t_f + t_b) + (T + 1)\frac{p}{b}t_f$, where the latter item indicates the time required in the query. The number of training samples n is $n_0, n_0 + q, n_0 + 2q, n_0 + 3q, \dots, n_0 + mq$ in each iteration, where m is the total number of iterations. The total training time required for the entire iteration process is $4\frac{(2n_0 + mq)(m + 1)}{b}e(t_f + t_b) + m(T + 1)\frac{p}{b}t_f$. Compared to the time spent by directly training the classifier, the proposed strategy training will be more time consuming.

However, the training time cost will not be an obstacle in its actual application. The first reason is that human experts cannot annotate the selected samples instantly; therefore, the high training cost will not be the bottleneck in terms of time cost. The second reason is that it is more expensive to hire doctors than hiring GPUs. The increase in training costs is beneficial compared to the decrease in annotation costs it brings.

4. Experiments and Results

4.1. Datasets and Configuration

The data in this experiment are from the Monash University Centre for Artificial Intelligence Research. The dataset includes more than 1000 histopathology scan images. After we process the dataset, there are a total of 17,000 color images. Directly using raw datasets to train models can impose serious burdens and costs on equipment. We resize the image to 200×200 to train the model more efficiently. After removing invalid data, there are 6826 valid pathological images. We set the ratio of the training and test sets to 7:3. The experiment was performed 10 times of cross-validation. Among them, the positive samples numbered 3904 or about 57%. The negative samples were 2922, accounting for approximately 43%. There was no sample imbalance. The difference in the proportion of positive and negative samples is small. To fairly evaluate the performance of each model, all models are experimented in the same environment using the same dataset.

We use convolutional neural networks as classification models. In the experiments, the number of samples (n in Section 3.2.2) selected for each iteration is 100, and the total number of iterations is 100. We use precision, recall, and AUC as evaluation metrics for the method [62]. In addition, we comparatively analyze the following methods:

- (1) Random selection (Random): the core idea of this strategy is that in each iteration, we randomly select samples to label.
- (2) BALD uncertainty-based strategy (denoted as BALD): in each iteration, we use BALD uncertainty to select samples for labeling.
- (3) Other strategies based on uncertainty, including entropy-based strategies (entropy) and confidence-based strategies (confidence) [18,63].
- (4) Strategy based on GAN reconstruction and BALD uncertainty (denoted as GAN + BALD): simultaneously train GAN and classifier, use BALD to select samples, and then augment the training set with reconstructed and selected images.
- (5) Strategy (our) based on the method proposed in this study: train the GAN and the classifier simultaneously, select samples using Equation (13), and then augment the training set with the reconstructed and selected images.
- (6) AIFT: It is a model that integrates active learning and transfer learning together. It is achieved by integrating active learning into fine-tuned CNNs in a continuous manner [64].
- (7) O-MedAL: It is a strategy that uses new labeled samples and a subset of previously labeled samples to increment the model and improve the MedAL model by introducing online learning methods [65].

4.2. Data Processing

In the osteosarcoma original pathology dataset, there are a total of 17,000 color images. There are many invalid images, as shown in Figure 4. Some color images do not include intact cells, or even only tissue fluid, or pathological sections are contaminated, resulting in unclear images, water spots, etc. To avoid ineffective resource waste, we screened the original osteosarcoma histopathology images. We removed invalid images from the dataset. After eliminating 10,174 invalid images, there are 6826 remaining valid color images. Of these, 4027 images were $10\times$ magnification of the original slice, and 2799 were $20\times$ magnification.

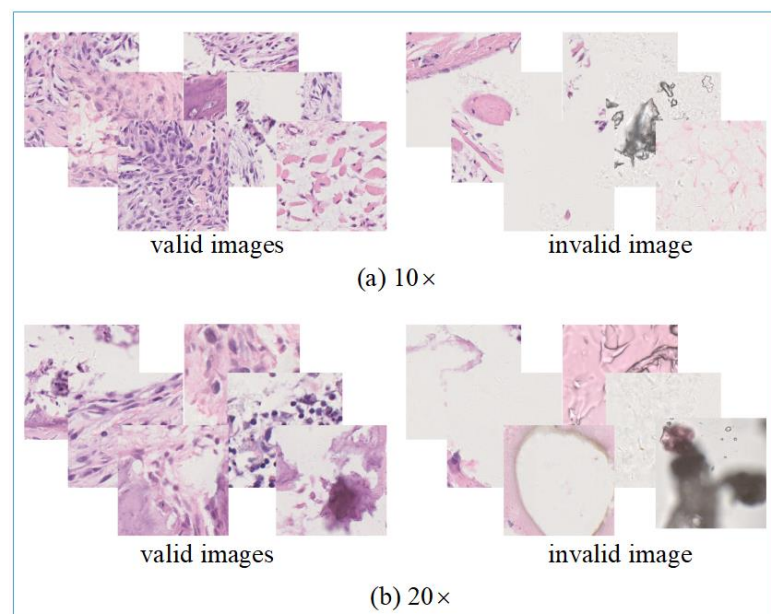


Figure 4. The example of data processing outcome. In the figure, the invalid images are the rejected samples, and the valid images are the sample data used for model training. Part (a) is the result of enlarging the original pathology image by $10\times$. Part (b) is the result of enlarging the original image by $20\times$.

To prevent the model's overfitting issue, we perform data enhancement after removing the invalid data. Firstly, each image is rotated by 90° , 180° , and 270° , and each image is flipped horizontally and vertically, at which time the amount of data is eight times that of the original dataset. Next, we select some images at random and add noise to them to reduce the overfitting phenomenon that occurs when the model learns high-frequency features. After data enhancement, the existing pathology samples reached 54,608.

4.3. Results

(1) Validation of hyperparameters

The hyperparameter θ is introduced in the selection strategy combined with diversity. When there is no prior knowledge of the data distribution, the optimal value of θ is not known. To choose the best value, we conduct experiments to test the effect of different θ values on the model performance, as shown in Figure 5. As θ changes, the classification accuracy of the model changes dynamically. Throughout the training process, accuracy takes the lead in a continuous rise, reaches a critical point, and then continues to decline. With $\theta < 0.5$, the model focuses more on the diversity of the sample. With $\theta > 0.5$, the model focuses more on the uncertainty about the sample. The results demonstrate that the model has equal importance based on diversity and uncertainty.

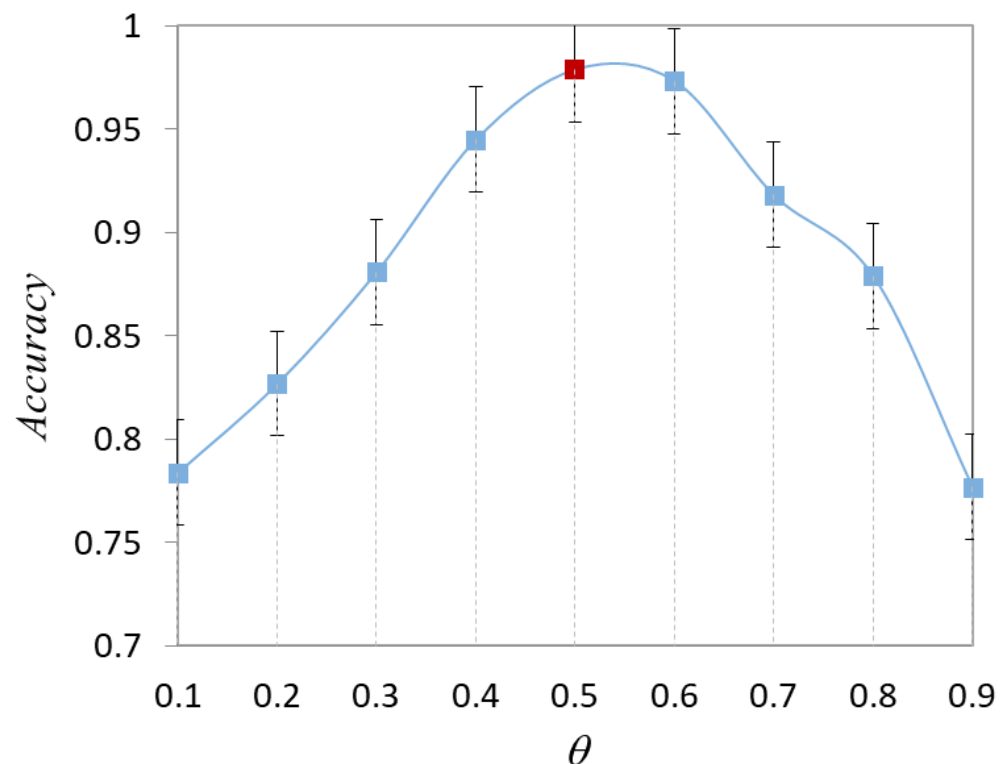


Figure 5. Effect of parameter θ on model accuracy.

The effect of the parameter θ on the recall of the model is shown in Figure 6. In the actual osteosarcoma pathology images, we are more concerned with model recall. This is mainly due to the more serious consequences of misclassifying abnormal images as normal images. The recall of the method has a significant change with the increase of θ . In particular, the recall rate reaches a critical value when the parameter θ increases from 0.3 to 0.7. The recall of the model is close to 0.9 when $\theta = 0.4$. It indicates that the model has a good classification effect at this time. With $\theta > 0.5$, the recall of the model gradually decreases.

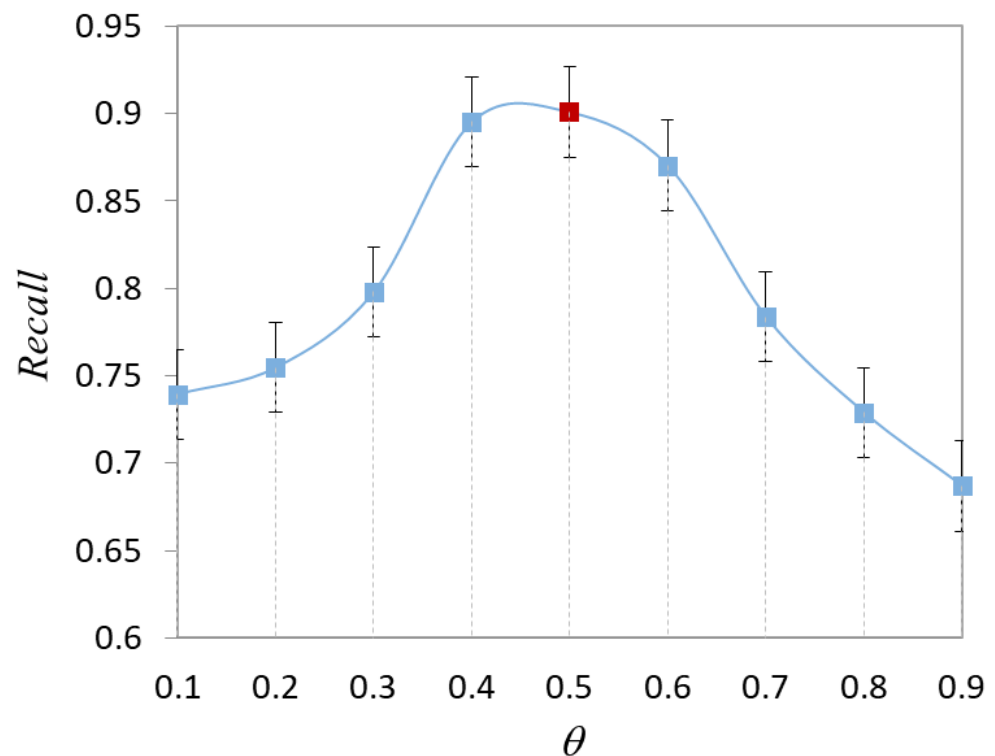


Figure 6. Effect of parameter θ on model recall.

After considering Figures 5 and 6 together, we set $\theta = 0.5$ for the experiments. As the number of training rounds increases and the experimental data changes, the optimal θ may also be other values. The value of θ may change dynamically during the subsequent iterations.

(2) Analysis of the final results of model classification

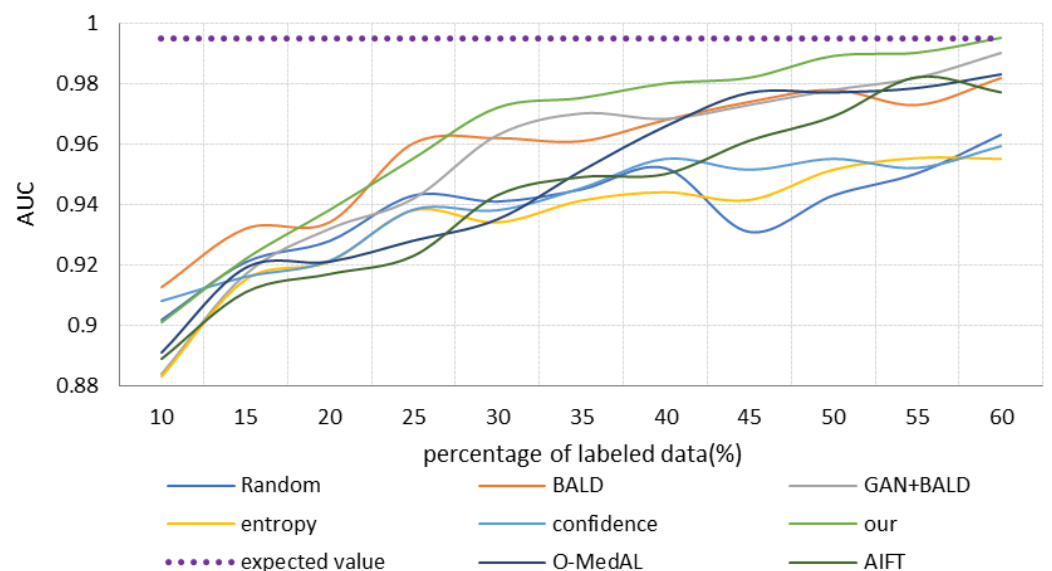
Based on the above dataset and environment configuration, we iteratively trained the model. Table 2 shows the performance of Random, BALD, GAN + BALD, Entropy, Confidence, and our method in terms of AUC, Accuracy, and Recall. The AUC value of the Random strategy is the smallest, and the recall rate is also the lowest, only 0.77. Relatively speaking, the performance of the Entropy and Confidence methods has been greatly improved. AUC increased by 2.605% on average, and recall increased by 10.78% on average. Although the AUC value of GAN + BALD reaches 0.99 and the accuracy reaches 0.98, its recall rate is low. Compared to the first five methods, the AIFT model has little difference in each metric. Although it uses active learning to fine-tune the CNN, it finally achieves an AUC value of 0.9807, an accuracy of 0.9832, and a recall of 0.8712. All metrics are lower than those of the GAN + BALD strategy. The AUC value of the O-MedAL model is also lower than that of the GAN + BALD strategy, but its recall and accuracy are better. In the classification of osteosarcoma histopathological images, this is disadvantageous for the classification of osteosarcoma pathological images. In hospital diagnosis, doctors and patients will pay more attention to the recall rate. According to the experimental results, the indicators of our proposed method are the best. Not only does it have a precision close to 0.99, but the recall also hits 0.90. In particular, its recall rate is about 17.5% higher than that of GAN + BALD.

Table 2. Performance of different strategies at the end of the iterative process.

Selection Strategy	AUC	Accuracy	Recall
Random	0.9546	0.9830	0.7729
BALD	0.9856	0.9829	0.7922
GAN + BALD	0.9931	0.9849	0.8743
Entropy	0.9786	0.9789	0.8598
Confidence	0.9804	0.9804	0.8526
AIFT	0.9807	0.9832	0.8712
O-MedAL	0.9918	0.9854	0.8894
Our	0.9950	0.9890	0.9080

(3) Results of comparing different scaled datasets

One aspect of this research is to propose a high-precision osteosarcoma pathological image-assisted classification system. Another important purpose is to use fewer labeled samples to achieve similar accuracy to all data training, demonstrating the low cost and effectiveness of the selection strategy. Therefore, we retrain each model in turn by selecting the same proportion of samples with labels. Specifically, we choose 10% of the total labeled samples as the initial training set and then keep adding new labeled images at a rate of 5% for experiments. To avoid imbalance in the dataset, each time we selected the positive and negative images of the patients separately according to the same ratio. The performance of different strategies is shown in Figures 6–8. The different colored lines represent several different selection strategies. Here, we use “convergence” to mean using part of the samples to train the classification model until achieving the same performance of using all the sample classification models, such as obtaining the same accuracy, recall, or AUC, that is, the purple dotted line is the result of training the base model with all the data. We end the iterative process with this value.

**Figure 7.** AUC with the increase of iterations.

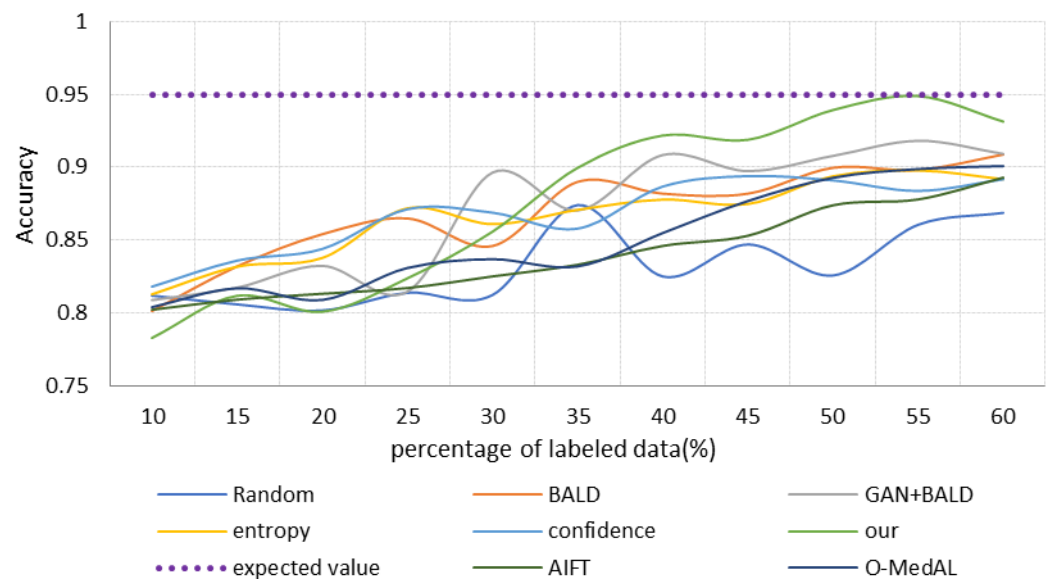


Figure 8. Accuracy with the increase of iterations.

The relationship between AUC and training samples is shown in Figure 7. With the increase of training samples, the AUC of all strategies increases gradually. The BALD strategy has the highest AUC value when the training sample is only 10% of the total dataset, followed by Confidence, our method, Random, O-MedAL, and AIFT models, in that order. The Entropy strategy and the GAN + BALD method have the worst performance. The performance of the BALD strategy is the best when the percentage of training samples is less than 25%. GAN + BALD and the method proposed in this study have the largest increase in AUC values. For the AIFT model and the O-MedAL model, the AUC values of both are still at a low level, despite their improvement. The performance of the Random strategy is very unstable when the training sample is greater than 25%, and the AUC value has a substantial decrease when the training sample increases from 40% to 45%. The AUC value of the AIFT model has a large increase, especially when the training sample is 55%, and the AUC reaches 0.98. The AUC value of our proposed method is always the highest, and the “convergence” value is reached with only 60% of the training data. This is mainly due to the combination of uncertainty and diversity in our proposed strategy. Furthermore, it can be clearly seen that throughout the iteration process, the BALD strategy is always better than simply randomly selecting samples for training.

Figure 8 is the relationship between the accuracy of different strategies and samples. The Random strategy achieves an accuracy of 0.88 when the labeled sample is 35%, but it has a large fluctuation in accuracy when selecting samples throughout the process. This indicates that the Random strategy is random and not suitable for practical tasks. The accuracy of the BALD and GAN + BALD models is also at an unstable level, but the accuracy of the GAN + BALD model is always at a higher value. In most cases, the accuracy of the GAN + BALD model outperforms most of the other methods. The accuracy of the AIFT model is always increasing its state, but its value is only higher than that of the Random strategy. The accuracy of the Entropy strategy, BALD strategy, and Confidence strategy is relatively high when the training sample is less than 35%. During the increase of the labeled training sample from 35% to 60%, the Entropy, Confidence, and BALD strategies maintain a slowly increasing trend and are in a steady state. This also implies that the small increase of the dataset does not improve the model performance significantly. Our method consistently maintains the highest accuracy after reaching 35% of the training samples. In particular, when the training samples reach 50, the accuracy is close to the best value of the initial training, 0.95. In addition, at this time, the accuracy of our method is 7.7% higher than the GAN + BALD algorithm. Furthermore, it is evident that our proposed strategy is

the first to achieve the accuracy of the base model training. This shows that it requires only a small amount of labeled data to achieve the best performance of the model.

The relationship between the recall rate of each method and the samples is shown in Figure 9. When the labeled data are only 10%, the recall of our method is the highest, followed by the GAN + BALD model. Both recall percentages are higher than 0.75. For the remaining models, their recall rates vary less and are between 0.7 and 0.75. Overall, the recall rate of each model gradually improves as the labeled sample increases. When the training samples reach 55%, the recall of our proposed method approaches 0.95 (“convergence” value). At this point, the O-MedAL model, BALD, and GAN + BALD strategies also have a recall of 0.9. The Confidence strategy has the lowest recall of less than 0.85. Then, the Random strategy, Entropy strategy, and AIFT model are in order. In comparison, the performance of the Entropy strategy, confidence strategy, and Random strategy is poor, and the recall rate is only about 0.85. Moreover, the recall rates of these three algorithms all fluctuate greatly with the increase of training data. Despite being only 35% of the training samples, the recall rate of the entropy and confidence strategies is above 0.85. However, they were always in an unstable state, suggesting they were less effective.

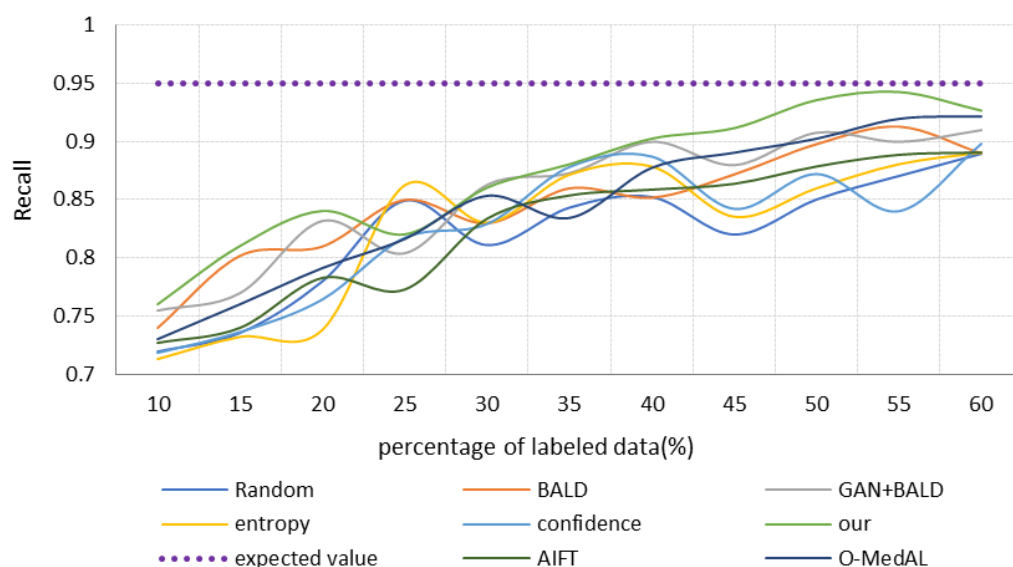


Figure 9. Recall with the increase of iterations.

The strategy that takes into account uncertainty and diversity converges faster than the strategy that only relies on uncertainty. It is worth pointing out that in the early stages of iterative training, the quality of GAN reconstruction samples is not high. This is because the number of labeled samples is small, and the global information of the data distribution captured by GAN is less. Therefore, strategies combined with GAN perform worse than the pure uncertainty strategy. However, in the later stage, as the quality of GAN-reconstructed samples improves, strategies combined with GAN gradually gain the upper hand, especially our proposed method. It first reaches the “convergence” value and then maintains a stable and efficient state.

Figure 10 compares the performance and the number of training samples for the fully trained and proposed frameworks. When different methods achieve and fully train the same performance, there is a large difference in the amount of labeled data used. As can be seen from the results presented in Figure 8, our method requires only 55.64% of the labeled data for training when it achieves the same performance as training with all labels. This is very valuable when assisting in the analysis of pathological images of osteosarcoma in actual hospitals. The reduction in the need for labeled data means a significant reduction in the cost of annotation.

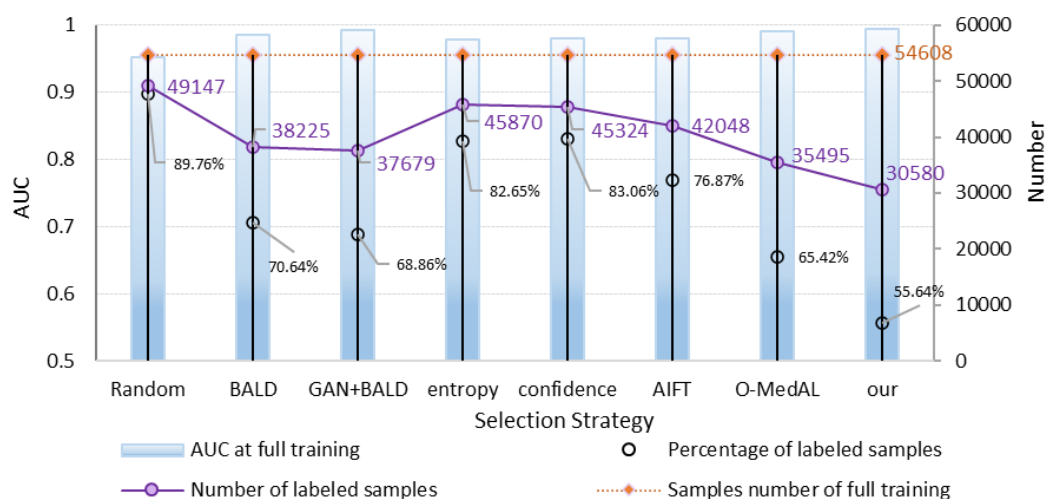


Figure 10. Performance and training sample number comparison of the full-training and proposed framework.

(4) Analysis of the results of physician verification

From the prediction results of each stage of our model, the doctors randomly selected 100 images for validation to measure the model’s accuracy from a medical standpoint. As shown in Table 3, the predicted values are the prediction results of the model when trained with different proportions of labeled samples. The actual values are the results of the randomly selected samples evaluated by a physician after the examination. When using a smaller number of labeled samples (10% of the labeled data), the system predicts less accurate results. For example, the model output determined 45 images to be positive, but after the doctor’s examination, only 21 images were actually positive. The model predicts 55 images as negative, but only 25 images are actually negative. As more labeled training samples are added, the model’s performance gradually becomes better. When the trained labeled data reached 50%, the prediction accuracy improved significantly to 0.74. After all the data had been fully trained, the model had an accuracy and recall of 0.89 and 0.88, respectively. This basically agrees with the findings of the analysis mentioned above. It shows that our proposed method has good performance. In a practical setting, it can be used as an auxiliary basis for doctors’ clinical diagnosis to improve the accuracy and efficiency of diagnosis and reduce the workload of doctors. Especially in the analysis of medical images, it only needs less annotation cost to achieve high performance, which effectively reduces the cost of building an intelligent medical assistance system.

Table 3. Analysis of model prediction results and physician examination results.

Percentage of Labeled Samples	True		
	Predicted	Positive	Negative
10%	Positive	21	24
	Negative	30	25
50%	Positive	41	15
	Negative	11	33
100%	Positive	46	5
	Negative	6	43

4.4. Discussion

The performance of the model classification is not guaranteed when using the Random selection strategy because the chosen training samples have a high degree of randomness. Although the model’s overall performance improves as the number of training samples rises, there is little correlation between the performance improvement and the

selection method. This is primarily because growing the training sample size improves the algorithm's performance. Uncertainty-based strategies [18,60] (including Entropy and Confidence) are an effective improvement on the Random strategy. Such methods not only end up with higher classification accuracy, but they are able to achieve better accuracy with relatively less labeled data. When model iterations increase, a single BALD strategy may cause overfitting of local samples, which limits the performance of BALD strategies. GAN learns knowledge from the overall data distribution and adds it to the reconstructed samples to increase the richness of the data. Therefore, it can greatly improve the classification performance of the model. However, in the early stage of model training, when the number of labeled samples is small, the performance of BALD is better than that of the GAN + BALD model. This is mainly because GAN acquires less global information of the data, and the quality of the reconstructed samples is not high. As the training increases and the number of labeled samples increases, the quality of GAN-generated samples improves, and the performance of the GAN + BALD strategy has more substantial improvement.

Both the AIFT model [64] and the O-MedAL model [65] are based on active learning, but both have improved performance compared to strategies based purely on active learning or randomly selected labeled samples. The AIFT model is an approach that concentrates active learning and transfer learning into a single framework. Starting with a pre-trained CNN, AIFT looks for "valuable" annotated samples and progressively enhances the CNN by continuous fine-tuning. Because AIFT actively selects the most informative and representative samples, it also outperforms most experimental methods. The O-MedAL model samples unlabeled images by maximum uncertainty and distance from labeled samples. As the model improves the classification accuracy of the samples, the recognition of unlabeled samples is also increasing. It uses relatively few samples and achieves the same performance as when fully trained.

The results show that our OHICsA method has the best performance. Not only does it have the highest classification accuracy when fully trained, but it requires only 60% of the training samples to achieve similar performance as when trained with full samples with labels. It significantly reduces the cost of building such a medical system. This is mainly since our approach combines active learning and generative models. The active-learning strategy is designed to combine sample class diversity and uncertainty to select "more valuable" samples, and the GAN model with strong learning data distribution capability generates pseudo-samples to extend the original dataset. This approach results in a faster performance improvement of the model. However, since our selection strategy focuses on selecting "valuable" or "informative" images, in the early stage of model training, there may be anomalous cases of selected images, and the quality of CACEGAN reconstruction samples is low. Therefore, the performance of the model is not outstanding at this stage. With the iteration of the model and the increase of training data, the performance of our method improves more substantially. In general, our method is the first to reach the "convergence" value, and thereafter it ensures a relatively stable and efficient level. It effectively overcomes the problem of low medical labeled data and achieves better classification performance with fewer labeled samples.

5. Conclusions

The study proposes an active-learning-based method for histopathological images of osteosarcoma for assisted classification. The system exploits the powerful fitting ability of deep convolutional networks to improve classification performance and uses a combination of active learning and generative adversarial networks to obtain faster performance gains. On the one hand, GAN generates more realistic pseudo-samples that augment the original dataset. On the other hand, the strategy of combining diversity and proactivity is used to select samples with a large amount of information. The classification results can be used as an aid for physicians to identify pathological images of osteosarcoma, reducing the workload and misdiagnosis rate in osteosarcoma diagnosis. More importantly, the method requires only a small amount of labeled data to achieve similar performance when trained

on all labeled data. It is designed mainly to reduce the cost of labeled data and building the system.

In the future, as computing power grows and deep-learning methods advance, we will pay more attention to how well model training works in its initial stages. Our research will concentrate on finding solutions to the issues of low sample quality produced by GAN and anomalies of the images chosen by the diversity strategy. Additionally, enhancing model training effectiveness will be a key concern of our strategy.

Author Contributions: Writing—original draft, F.G. and J.Z.; writing—review and editing, J.L. and J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, 12 months after the publication of this article, will be considered by the corresponding author.

Acknowledgments: Availability of data and materials: all data analyzed during the current study are included in the submission.

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

References

1. Wu, J.; Guo, Y.; Dai, Z. A medical assistant segmentation method for MRI images of osteosarcoma based on DecoupleSegNet. *Int. J. Intell. Syst.* **2022**, *37*, 8436–8461. [\[CrossRef\]](#)
2. Eaton, B.R.; Schwarz, R.; Vatner, R.; Yeh, B.; Claude, L.; Indelicato, D.J.; Laack, N. Osteosarcoma. *Pediatr. Blood Cancer* **2021**, *68* (Suppl. S2), e28352. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Liu, F.; Gou, F.; Wu, J. An Attention-Preserving Network-Based Method for Assisted Segmentation of Osteosarcoma MRI Images. *Mathematics* **2022**, *10*, 1665. [\[CrossRef\]](#)
4. Zhou, L.; Tan, Y. A Residual Fusion Network for Osteosarcoma MRI Image Segmentation in Developing Countries. *Comput. Intell. Neurosci.* **2022**, *2022*, 7285600. [\[CrossRef\]](#)
5. Rathore, R.; van Tine, B.A. Pathogenesis and Current Treatment of Osteosarcoma: Perspectives for Future Therapies. *J. Clin. Med.* **2021**, *10*, 1182. [\[CrossRef\]](#)
6. Wang, L.; Yu, L.; Zhu, J.; Tang, H. Auxiliary Segmentation Method of Osteosarcoma in MRI Images Based on Denoising and Local Enhancement. *Healthcare* **2022**, *10*, 1468. [\[CrossRef\]](#)
7. Fei, F.; Harada, S.; Wei, S.; Siegal, G.P. Chapter 40—Molecular pathology of osteosarcoma. In *Bone Sarcomas and Bone Metastases—From Bench to Bedside*, 3rd ed.; Heymann, D., Ed.; Elsevier: Amsterdam, The Netherlands, 2022; pp. 579–590.
8. Chang, L.; Moustafa, N.; Bashir, A.K.; Yu, K. AI-Driven Synthetic Biology for Non-Small Cell Lung Cancer Drug Effectiveness-Cost Analysis in Intelligent Assisted Medical Systems. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 5055–5066. [\[CrossRef\]](#)
9. Roessner, A.; Lohmann, C.; Jechorek, D. Translational cell biology of highly malignant osteosarcoma. *Pathol. Int.* **2021**, *71*, 291–303. [\[CrossRef\]](#)
10. Yu, G.; Chen, Z.; Tan, Y. Medical decision support system for cancer treatment in precision medicine in developing countries. *Expert Syst. Appl.* **2021**, *186*, 115725. [\[CrossRef\]](#)
11. Wu, J.; Gou, F.; Tian, X. Disease Control and Prevention in Rare Plants Based on the Dominant Population Selection Method in Opportunistic Social Networks. *Comput. Intell. Neurosci.* **2022**, *2022*, 1489988. [\[CrossRef\]](#)
12. Yang, W.; Luo, J.; Wu, J. Application of Information Transmission Control Strategy Based on Incremental Community Division in IoT Platform. *IEEE Sens. J.* **2021**, *21*, 21968–21978. [\[CrossRef\]](#)
13. Xiao, P.; Huang, H.; Zhou, Z.; Dai, Z. An artificial intelligence multiprocessing scheme for the diagnosis of osteosarcoma MRI images. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 4656–4667. [\[CrossRef\]](#)
14. Roh, Y.; Heo, G.; Whang, S.E. A Survey on Data Collection for Machine Learning: A Big Data—AI Integration Perspective. *IEEE Trans. Knowl. Data Eng.* **2021**, *33*, 1328–1347. [\[CrossRef\]](#)
15. Zhan, X.; Long, H.; Duan, X.; Kong, G. A Convolutional Neural Network-Based Intelligent Medical System with Sensors for Assistive Diagnosis and Decision-Making in Non-Small Cell Lung Cancer. *Sensors* **2021**, *21*, 7996. [\[CrossRef\]](#)
16. Cho, J.W.; Kim, D.-J.; Jung, Y.; Kweon, I.S. MCDAL: Maximum Classifier Discrepancy for Active Learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 1–11. [\[CrossRef\]](#)
17. Lv, B.; Liu, F.; Gou, F.; Wu, J. Multi-Scale Tumor Localization Based on Priori Guidance-Based Segmentation Method for Osteosarcoma MRI Images. *Mathematics* **2022**, *10*, 2099. [\[CrossRef\]](#)

18. Li, G.J.; Porter, M.A. A Bounded-Confidence Model of Opinion Dynamics with Heterogeneous Node-Activity Levels. *arXiv* **2022**, arXiv:10.31235/osf.io/r6asm. Available online: <https://osf.io/preprints/socarxiv/r6asm> (accessed on 23 September 2022).
19. Xiong, W.; Zhou, X. A Reputation Value-Based Task-Sharing Strategy in Opportunistic Complex Social Networks. *Complexity* **2021**, *2021*, 8554351. [[CrossRef](#)]
20. Wu, J.; Yang, S.; Gou, F.; Zhou, Z.; Xie, P.; Xu, N.; Dai, Z. Intelligent Segmentation Medical Assistance System for MRI Images of Osteosarcoma in Developing Countries. *Comput. Math. Methods Med.* **2022**, *2022*, 7703583. [[CrossRef](#)]
21. He, T.; Jin, X.; Ding, G.; Yi, L.; Yan, C. Towards Better Uncertainty Sampling: Active Learning with Multiple Views for Deep Convolutional Neural Network. In Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 8–12 July 2019; pp. 1360–1365.
22. Nanda, S.K.; Ghai, D.; Ingole, P.; Pande, S. Soft Computing Techniques-based Digital Video Forensics for Fraud Medical Anomaly Detection. *Comput. Assist. Methods Eng. Sci.* **2022**.
23. Dharmale, S.G.; Gomase, S.A.; Pande, S. Comparative Analysis on Machine Learning Methodologies for the Effective Usage of Medical WSNs. In Proceedings of Data Analytics and Management, Proceedings of the International Conference on Data Analytics and Management, Polkowice, Poland, 26 June 2021; Springer: Singapore, 2022; pp. 441–457.
24. Yadav, N.; Alfayeed, S.M.; Khamparia, A.; Pandey, B.; Thanh, D.N.; Pande, S. HSV model-based segmentation driven facial acne detection using deep learning. *Expert Syst.* **2022**, *39*, e12760. [[CrossRef](#)]
25. Zhou, Z.; Tan, Y. A cascaded multi-stage framework for automatic detection and segmentation of pulmonary nodules in developing countries. *IEEE J. Biomed. Health Inform.* **2022**, 1–12. [[CrossRef](#)] [[PubMed](#)]
26. Zhu, J.-J.; Bento, J. Generative adversarial active learning. *arXiv* **2017**, arXiv:1702.07956. Available online: <https://arxiv.org/abs/1702.07956> (accessed on 24 August 2022).
27. Tran, T.; Do, T.-T.; Reid, I.; Carneiro, G. Bayesian generative active deep learning. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6295–6304.
28. Xue, Y.; Ye, J.; Zhou, Q.; Long, R.; Antani, S.; Xue, Z.; Cornwell, C.; Zaino, R.; Cheng, K.; Huang, X. Selective synthetic augmentation with HistoGAN for improved histopathology image classification. *Med. Image Anal.* **2021**, *67*, 101816. [[CrossRef](#)] [[PubMed](#)]
29. George, K.; Faziludeen, S.; Sankaran, P.; Paul, J.K. Deep learned nucleus features for breast cancer histopathological image analysis based on belief theoretical classifier fusion. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON), Proceedings of the IEEE Region 10 International Conference TENCON, Kochi, India, 17–20 October 2019*; IEEE: New York, NY, USA, 2019; pp. 344–349.
30. Gupta, V.; Bhavsar, A. Partially-Independent Framework for Breast Cancer Histopathological Image Classification. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019; pp. 1123–1130. [[CrossRef](#)]
31. Wang, P.; Li, P.; Li, Y.; Wang, J.; Xu, J. Histopathological image classification based on cross-domain deep transferred feature fusion. *Biomed. Signal Process. Control* **2021**, *68*, 102705. [[CrossRef](#)]
32. Nave, O. Adding features from the mathematical model of breast cancer to predict the tumour size. *Int. J. Comput. Math. Comput. Syst. Theory* **2020**, *5*, 159–174. [[CrossRef](#)]
33. Sekhar, A.; Biswas, S.; Hazra, R.; Sunaniya, A.K.; Mukherjee, A.; Yang, L. Brain Tumor Classification Using Fine-Tuned GoogLeNet Features and Machine Learning Algorithms: IoMT Enabled CAD System. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 983–991. [[CrossRef](#)]
34. Wu, W.; Gao, L.; Duan, H.; Huang, G.; Ye, X.; Nie, S. Segmentation of pulmonary nodules in CT images based on 3D-UNET combined with three-dimensional conditional random field optimization. *Medical Physics* **2020**, *47*, 4054–4063. [[CrossRef](#)]
35. Gur, S.; Wolf, L.; Golgher, L.; Blinder, P. Unsupervised microvascular image segmentation using an active contours mimicking neural network. In Proceedings of the IEEE/CVF international conference on computer vision, Seoul, Korea, 27 October–2 November 2019; pp. 10722–10731.
36. Ragland, B.D.; Bell, W.C.; Lopez, R.R.; Siegal, G.P. Cytogenetics and Molecular Biology of Osteosarcoma. *Lab. Investig.* **2002**, *82*, 365–373. [[CrossRef](#)]
37. Mishra, R.; Daescu, O.; Leavey, P.; Rakheja, D.; Sengupta, A. Convolutional Neural Network for Histopathological Analysis of Osteosarcoma. *J. Comput. Biol.* **2017**, *25*, 313–325. [[CrossRef](#)]
38. Anisuzzaman, D.M.; Barzekar, H.; Tong, L.; Luo, J.; Yu, Z. A deep learning study on osteosarcoma detection from histological images. *Biomed. Signal Process. Control* **2021**, *69*, 102931. [[CrossRef](#)]
39. Shen, Y.; Dai, Z. Osteosarcoma MRI Image-Assisted Segmentation System Base on Guided Aggregated Bilateral Network. *Mathematics* **2022**, *10*, 1090. [[CrossRef](#)]
40. Fu, Y.; Xue, P.; Ji, H.; Cui, W.; Dong, E. Deep Model with Siamese Network for Viability and Necrosis Tumor Assessment in Osteosarcoma. *Med. Phys.* **2019**, *47*, 4895–4905. [[CrossRef](#)] [[PubMed](#)]
41. Barzekar, H.; Yu, Z. C-Net: A Reliable Convolutional Neural Network for Biomedical Image Classification. *Expert Syst. Appl.* **2020**, *187*, 116003. [[CrossRef](#)]
42. D’Acunto, M.; Martinelli, M.; Moroni, D. From human mesenchymal stromal cells to osteosarcoma cells classification by deep learning. *J. Intell. Fuzzy Syst.* **2019**, *37*, 7199–7206. [[CrossRef](#)]
43. Wu, J.; Gou, F.; Tan, Y. A Staging Auxiliary Diagnosis Model for Non-small Cell Lung Cancer Based on the Intelligent Medical System. *Comput. Math. Methods Med.* **2021**, *2021*, 6654946. [[CrossRef](#)]

44. Ouyang, T.; Yang, S.; Dai, Z. Rethinking U-Net from an Attention Perspective with Transformers for Osteosarcoma MRI Image Segmentation. *Comput. Intell. Neurosci.* **2022**, *2022*, 7973404. [[CrossRef](#)]
45. Tian, X.; Yan, L.; Jiang, L.; Xiang, G.; Li, G.; Zhu, L.; Wu, J. Comparative transcriptome analysis of leaf, stem, and root tissues of *Semiliquidambar cathayensis* reveals candidate genes involved in terpenoid biosynthesis. *Mol. Biol. Rep.* **2022**, *49*, 5585–5593. [[CrossRef](#)]
46. Cui, R.; Chen, Z.; Tan, Y.; Yu, G. A Multiprocessing Scheme for PET Image Pre-Screening, Noise Reduction, Segmentation and Lesion Partitioning. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1699–1711. [[CrossRef](#)]
47. Gou, F.; Wu, J. Triad link prediction method based on the evolutionary analysis with IoT in opportunistic social networks. *Comput. Commun.* **2022**, *181*, 143–155. [[CrossRef](#)]
48. Wu, J.; Yu, L.; Gou, F. Data transmission scheme based on node model training and time division multiple access with IoT in opportunistic social networks. *Peer-to-Peer Netw. Appl.* **2022**. [[CrossRef](#)]
49. Long, H.; He, K. Effective Data Optimization and Evaluation Based on Social Communication with AI-Assisted in Opportunistic Social Networks. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 4879557. [[CrossRef](#)]
50. Qin, Y.; Li, X.; Yu, K. A management method of chronic diseases in the elderly based on IoT security environment. *Comput. Electr. Eng.* **2022**, *102*, 108188. [[CrossRef](#)]
51. Ling, Z.; Yang, S.; Gou, F.; Dai, Z.; Wu, J. Intelligent Assistant Diagnosis System of Osteosarcoma MRI Image Based on Transformer and Convolution in Developing Countries. *IEEE J. Biomed. Health Inform.* **2022**. [[CrossRef](#)] [[PubMed](#)]
52. Li, L.; Gou, G.; Wu, J. Modified Data Delivery Strategy Based on Stochastic Block Model and Community Detection in Opportunistic Social Networks. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 5067849. [[CrossRef](#)]
53. Gou, F.; Wu, J. Data Transmission Strategy Based on Node Motion Prediction IoT System in Opportunistic Social Networks. *Wirel. Pers. Commun.* **2022**, *126*, 1751–1768. [[CrossRef](#)]
54. Wu, J.; Xia, J.; Gou, F. Information transmission mode and IoT community reconstruction based on user influence in opportunistic social networks. *Peer-to-Peer Netw. Appl.* **2022**, *15*, 1398–1416. [[CrossRef](#)]
55. Deng, Y.; Gou, F.; Wu, J. Hybrid data transmission scheme based on source node centrality and community reconstruction in opportunistic social networks. *Peer-to-Peer Netw. Appl.* **2021**, *14*, 3460–3472. [[CrossRef](#)]
56. Jiao, Y.; Qi, H. Capsule network assisted electrocardiogram classification model for smart healthcare. *Biocybern. Biomed. Eng.* **2022**, *42*, 543–555. [[CrossRef](#)]
57. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016.
58. Zhou, Z.; Shin, J.; Zhang, L.; Gurudu, S.; Gotway, M.; Liang, J. Fine-Tuning Convolutional Neural Networks for Biomedical Image Analysis: Actively and Incrementally. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4761–4772.
59. Shen, Y.; Gou, F.; Wu, J. Node Screening Method Based on Federated Learning with IoT in Opportunistic Social Networks. *Mathematics* **2022**, *10*, 1669. [[CrossRef](#)]
60. Gou, F.; Wu, J. Message Transmission Strategy Based on Recurrent Neural Network and Attention Mechanism in Iot System. *J. Circuits Syst. Comput.* **2022**, *31*, 2250126. [[CrossRef](#)]
61. Tian, X.; Wu, J. Optimal matching method based on rare plants in opportunistic social networks. *J. Comput. Sci.* **2022**, *64*, 101875. [[CrossRef](#)]
62. Wu, J.; Liu, Z.; Gou, F.; Zhu, J.; Tang, H.; Zhou, X.; Xiong, W. BA-GCA Net: Boundary-Aware Grid Contextual Attention Net in Osteosarcoma MRI Image Segmentation. *Comput. Intell. Neurosci.* **2022**, *2022*, 3881833. [[CrossRef](#)] [[PubMed](#)]
63. Fu, X.; Wang, Y.; Belkacem, A.N.; Zhang, Q.; Xie, C.; Cao, Y.; Cheng, H.; Chen, S. Integrating Optimized Multiscale Entropy Model with Machine Learning for the Localization of Epileptogenic Hemisphere in Temporal Lobe Epilepsy Using Resting-State fMRI. *J. Healthc. Eng.* **2021**, *2021*, 1834123. [[CrossRef](#)] [[PubMed](#)]
64. Zhou, Z.; Shin, J.Y.; Gurudu, S.R.; Gotway, M.B.; Liang, J. Active, continual fine tuning of convolutional neural networks for reducing annotation efforts. *Med. Image Anal.* **2021**, *71*, 101997. [[CrossRef](#)] [[PubMed](#)]
65. Smalagic, A.; Costa, P.; Gaudio, A.; Khandelwal, K.; Mirshekari, M.; Fagert, J.; Walawalkar, D.; Xu, S.; Galdran, A.; Zhang, P.; et al. O-MedAL: Online active deep learning for medical image analysis. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1353. [[CrossRef](#)]