Feng J W. (2024) COMPUTATIONAL INTELLIGENCE FOR INJURY DIAGNOSIS AND TREATMENT DECISION SUPPORT FOR ATHLETES IN MEDICAL IMAGE PROCESSING. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 24 (98) pp. 41-57.
DOI: https://doi.org/10.15366/rimcafd2024.98.004

ORIGINAL

COMPUTATIONAL INTELLIGENCE FOR INJURY DIAGNOSIS AND TREATMENT DECISION SUPPORT FOR ATHLETES IN MEDICAL IMAGE PROCESSING

Jun Wei Feng

School of Graduate Studies, Jose Rizal University, Manila 0900, Philippines.
E-mail: f18865279056@163.com

Recibido 03 de enero de 2024 **Received** January 03, 2024
Aceptado 03 de septiembre de 2024 **Accepted** September 03, 2024

ABSTRACT

With the rapid development of national fitness and professional sports, the problem of sports trauma has become more and more prominent, which has caused serious impact on the training and competition of athletes. It is urgent to improve the level of diagnosis and treatment. The analysis technique of identifying and diagnosing cells and tissues for athletes in medical images with the help of deep learning algorithms has gradually become a popular research direction in the field of medical image diagnosis. Convolutional neural network (CNN), as an efficient deep learning algorithm, is widely used in the field of medical image diagnosis. However, since CNN models need to initialise the parameters before training, various problems may arise when the initial parameters are not properly selected. Firstly, for the initial weights of the CNN model, the traditional method is to use random initialisation, which leads to problems such as slow training speed and low diagnostic accuracy of the model. Secondly, for the selection of the hyperparameter of the model, the traditional method is to use manual adjustment or grid search, which not only consumes a lot of time and computational resources, but also usually fails to select the most suitable hyperparameter, which leads to the problems of lower diagnostic accuracy of the model. In order to solve the above problems, this paper firstly proposes a new self-supervised medical image segmentation architecture. By designing an agent task for pre-training, the model is better able to extract and process the visual information of medical images, and then fine-tuned on the segmentation task, as a way to solve the difficulty of the lack of large-scale labelled data for medical images. The effectiveness of the proposed algorithm for athletes in the medical image segmentation task is verified through a large

number of experiments conducted on two mainstream datasets. And the comparison with other mainstream models shows that the model performs well in most scenarios.

KEYWORDS: Computational Intelligence; Athletes Injury Diagnosis; Treatment Decision Support; Medical Image Processing; Deep Learning

1. INTRODUCTION

The prevalence and complexity of sports injuries pose challenges to traditional diagnosis and treatment. The use of medical image processing and machine learning and other technologies can achieve more accurate injury identification and diagnosis, and provide intelligent treatment decision support based on big data analysis, which helps shorten the recovery time of athletes and improve competitive performance. At the same time, through continuous data accumulation and analysis, a sound knowledge base for the diagnosis and treatment of sports injuries can be established, providing a solid foundation for future research and application in this field. Therefore, this study has important practical significance and development prospects, and has important value for improving the service level of sports medicine. Medical image segmentation is a very important aspect of medical image processing as it helps doctors to better understand the athlete's condition (Isensee et al., 2019; Siddique et al., 2021).

In medical images, different tissues, organs or lesion areas have different density, morphology, texture and other features, which are very crucial for doctors to make correct diagnosis and treatment. Traditional medical image segmentation methods (Hu et al., 2021; Ma et al., 2021) usually require the manual design of features and the use of rule-based algorithms for segmentation, which consumes a lot of time and money. This method consumes a lot of manpower and time, and the segmentation results are usually not accurate in the professional field. This method consumes a lot of manpower and time, and the segmentation results are usually not accurate enough in professional fields. However, with the recent advances in deep learning and other technologies, deep learning-based medical image segmentation (Liu et al., 2021; Wang et al., 2022) has become a mainstream method. techniques based on deep learning have become mainstream methods. These methods can automatically learn the features of medical images and use models such as convolutional neural networks to obtain more efficient and accurate segmentation results. networks to obtain more efficient and accurate segmentation results (Claudino et al., 2019; Xiao et al., 2023).

In the era of athlete data, there are a large number of different types of data in the athlete biomedical field. Common biomedical data include: omics data, drug substance data, disease data, electronic medical record data,

imaging data, wearable device data, and athlete report data. wait. According to statistics, as of August 25, 2021, the GEO database in the United States has included more than 100,000 sets of data, coming from 18,847 platforms, containing more than 2.62 million samples, involving 4,290 species. Among them, there are 42,628 sets of human data, coming from 5,325 platforms and containing nearly 1.41 million samples. The European Array Express database contains 71,250 sets of data, including more than 2.29 million samples, with a data volume of 46.83TB. These data usually have characteristics such as large data volume, high dimensionality, small samples, high noise, and uneven data distribution. Because deep learning algorithms enable computer systems to improve from experience and data through continuous self-learning, large-scale biomedical data is considered a prerequisite for the success of many deep learning algorithms. Therefore, in the biomedical field, a large amount of real and effective medical data makes deep learning algorithms feasible in intelligent medical applications (Conze et al., 2023; Qureshi et al., 2023). In the field of intelligent medical applications (Arabahmadi et al., 2022), medical imaging data stored in image form holds more than 90% of medical information and is the most important information source in disease diagnosis and treatment. The analysis of these medical images is also the most important technology to assist in the clinical diagnosis and treatment of diseases. means, so the comprehensive intelligence of medical imaging data analysis and diagnosis plays a crucial role in the construction of intelligent medical application fields. However, due to the technical limitations of traditional medical testing instruments and equipment, and the differences in the technology of medical staff to detect disease conditions, even the best medical staff with the highest level of technical skills in diagnosing and detecting diseases cannot be compared with each other. If you don't cooperate, you will also face the risk of missed diagnosis or incorrect diagnosis.

Therefore, in order to reduce various risks in disease diagnosis, in this article, we study two types of medical imaging data through deep learning algorithms (Zhan, 2024). The traditional clinical diagnosis method that relies on the subjective experience of doctors has some limitations, and the diagnosis time is long, which cannot meet the needs of athletes' timely recovery. At the same time, medical imaging technologies such as CT, MRI and ultrasound continue to develop, and image quality and resolution have been greatly improved, providing a good foundation for intelligent diagnosis using computer vision and machine learning. In addition, the accumulation of massive sports injury data and the wide application of AI technologies such as deep learning in the medical field have created favourable conditions for the establishment of intelligent diagnostic models. To sum up, the physical and mental health of athletes is crucial to the development of sports career, and improving the diagnosis and treatment level of sports injuries has become an urgent problem to be solved in sports medicine. The research on athlete intelligent injury diagnosis in medical image processing came into being, which provides new

technical support. Convolutional Neural Network (CNN) is a widely used deep learning algorithm. In medical imaging diagnosis, CNNs are mainly used in the recognition of medical images such as X-Ray, Magnetic Resonance Images (MRI), Histopathological images, Fundus images, and Computed Tomography (CT).

The CNN is mainly used to back-propagate the error of the forward propagation result by means of continuous iteration to update the parameters of the model, i.e., Weights and Biases, to finally achieve the learning objective (Wu et al., 2022). In this way, the selection of initial parameters plays a crucial role in the training effect of CNN. The initial parameters of CNN are mainly divided into two kinds, the initial weights and the Hyperparameter of the model. Generally speaking, the initial weights of CNN models are randomly initialised by various methods, which can lead to problems such as model training failure in the process of training the model, while the Hyperparameter of the CNN model is usually selected according to the performance of the model in the training process, which can cause a large amount of resource consumption. For these two problems of CNN model training, we hope to find an effective solution. Meta-heuristic algorithms are very suitable for the optimisation of complex neural network models due to their simplicity and high efficiency. Meta-heuristic algorithms are based on intuitive or empirical constructions, and provide a solution to the problem to be solved with acceptable resource consumption. CNN is a multilayered artificial neural network constructed to mimic the biological vision system, and in recent years researchers have been committed to combining the advantages of meta-heuristic algorithms and CNNs to achieve the goal of using both the meta-heuristics and the CNNs.

In recent years, researchers have endeavoured to combine the advantages of meta-heuristic algorithms and CNNs to achieve the goal of efficiently solving problems using neural networks while overcoming the shortcomings of neural networks (Khan et al., 2022). Furthermore, the convolutional neural network approach inherently suffers from a serious limitation: at the initial stage of the input image, since the convolutional kernel of a CNN is usually not very large, the model can only make use of local information in order to understand the input image, and this local perspective may limit the distinguishability of the features extracted by the encoder (Farshad et al., 2022).

This is a drawback that cannot be completely avoided using CNNs (Song & Montenegro-Marin, 2021). Of course, some plug-and-play modules based on self-attention mechanisms inserted between the encoder and the decoder can provide global context, allowing the model to better understand the image and improve features from a global perspective. However, if the model initially obtains the wrong features due to this local perspective, can it be corrected at

a subsequent stage using the global context, which creates new problems. In conclusion, self-supervised learning has a long history of research in the field of medical imaging and has received increasing attention and applications in recent years. Different researchers have proposed different self-supervised learning methods which provide more efficient and accurate solutions for medical image analysis and diagnosis (Ramkumar et al., 2022). The main contributions are as follows:

(1) This paper proposes a new approach to medical image segmentation for athlete using Transformer models. This approach builds on the success of Transformer models in other computer vision tasks and demonstrates their potential for medical image segmentation. Experimental results on two benchmark datasets show that the proposed Transformer-based model outperforms current mainstream models in terms of segmentation accuracy and has good generalisation capabilities.

(2) In this paper, a self-supervised pre-trained agent task is designed for learning feature representations that are more suitable for medical image segmentation. The task captures important image features for feature representation for segmentation with respect to human anatomy in medical images. The method differs from traditional supervised learning methods by avoiding the difficulty of acquiring large labelled datasets and provides a promising alternative for feature learning in medical image segmentation.

2. Methodology

In this paper, we propose a self-supervised pre-trained model designed for agent tasks targeting athlete medical images to pave the way for the proposal of Transformer-based self-supervised medical image segmentation, and specify the details of each agent task. Then each important module of the model and the whole is described in detail, including the attention mechanism and the extraction of multi-scale semantic information.

2.1 Transformer Model

Transformer is a deep learning model based on the attention mechanism, which was initially applied to natural language processing tasks, such as machine translation and text generation. Later, Transformer was also introduced into the field of computer vision for processing visual tasks, such as image classification, object detection, semantic segmentation, etc. In computer vision, Transformer is mainly applied to process sequence data, such as pixels, feature maps in images, etc. It captures the relationship between different elements in a sequence through the mechanism of self-attention and extracts the key information in the sequence for subsequent classification, detection, segmentation and other tasks.

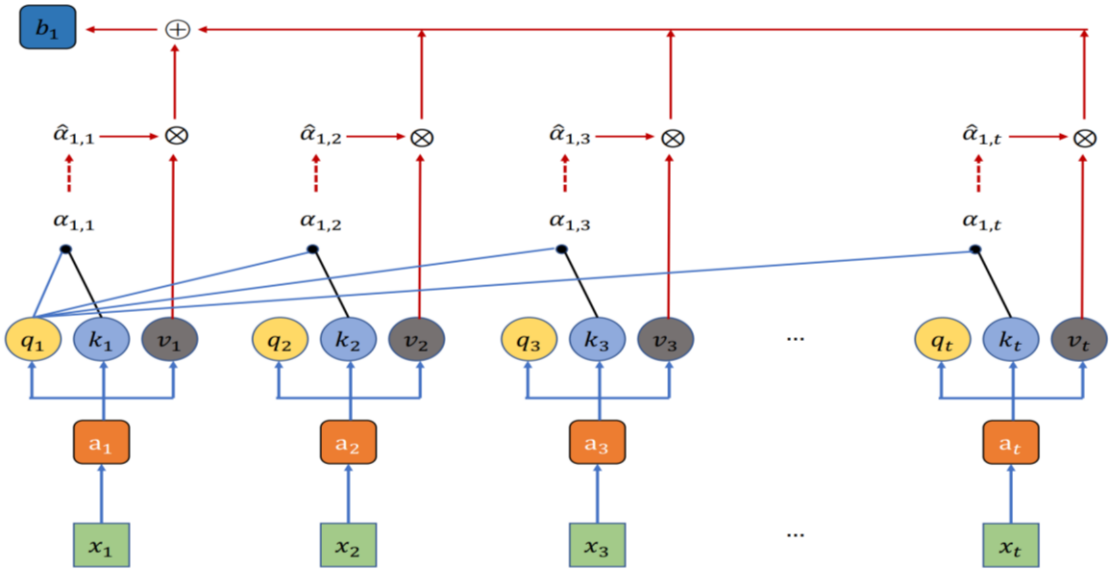


Figure 1: Schematic diagram of the self-attention mechanism calculation.

Compared to traditional convolutional neural networks, Transformer does not require a fixed window size or convolutional kernel size when processing sequence data, thus providing better scalability and adaptability. Transformer has achieved good performance in some tasks that require global information about an image, such as semantic segmentation and image generation.

2.1.1 Self-Attention Mechanism

Speaking of attention, it is used in graphic conversion in computers. Given a picture and a simple description of the picture, in order for the computer to learn the important and relevant information about both, we react to the weights of the different information by means of attention. Because a picture is composed of several pixel dots, it contains rich semantic information, but the vast majority of this information is redundant or even unrelated to the picture description. Similar to the perception of the world observed by people's eyes in real life, the focusing of our eyes is also a mechanism of attention, through which the observed object can be presented more clearly, and the blurring out of other secondary information will not affect the perception of the observed subject. Similarly, attention in this way gives a higher weight to the region corresponding to the graphic, which is more conducive to learning deeper features. Therefore, the interpretability aspect of attention in AI is more in line with our intuitive perception, as shown in Figure 1. The success of the Transformer family of models is largely attributed to the Self-Attention (SA) mechanism, as SA has the ability to establish dependencies over long distances. The key idea behind the SA mechanism is to learn self-alignment, i.e., to determine the relative importance of individual markers (patch embeddings) with respect to all other markers in the sequence. For a 2D image,

the image $X_p \in R^{H\times W\times C}$ is first reshaped into a sequence of 2D vectors (called Flattened Patch) $X_p \in R^{N\times(P^2C)}$, where H and W denote the height and width of the original image, respectively, C denotes the number of channels, $P \times P$ denotes the pixel size of each image block, and $N = \frac{HW}{P^2}$ denotes the number of image blocks. These 2D vectors are projected into D dimensional space through a trainable linear projection layer and can be represented in matrix form as $X_p \in R^{N\times D}$. The goal of SA is to capture the intrinsic connection between all. this N embedding, which is done by defining three learnable weight matrices to transform the input X into query, key-value pairs:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{D_q}}\right)VW^{V_i} \quad (1)$$

$$Z = \text{Attention}(X) = AV \quad (2)$$

2.1.2 Multiple Self-Attention Mechanism

The MHSA is a model of an attention mechanism consisting of multiple SA blocks (also called headers). As shown in Figure 2, these blocks are connected according to channels to model the complex dependencies between different elements in the input sequence. Each block has its own learnable weight matrix, denoted by W^{Q_i} ,W^{K_i}, W^{V_i}, which is used to calculate the weight of each element in the input sequence. By using multiple blocks, MHSA can better capture the interactions between different features in the sequence and improve the performance of the model.

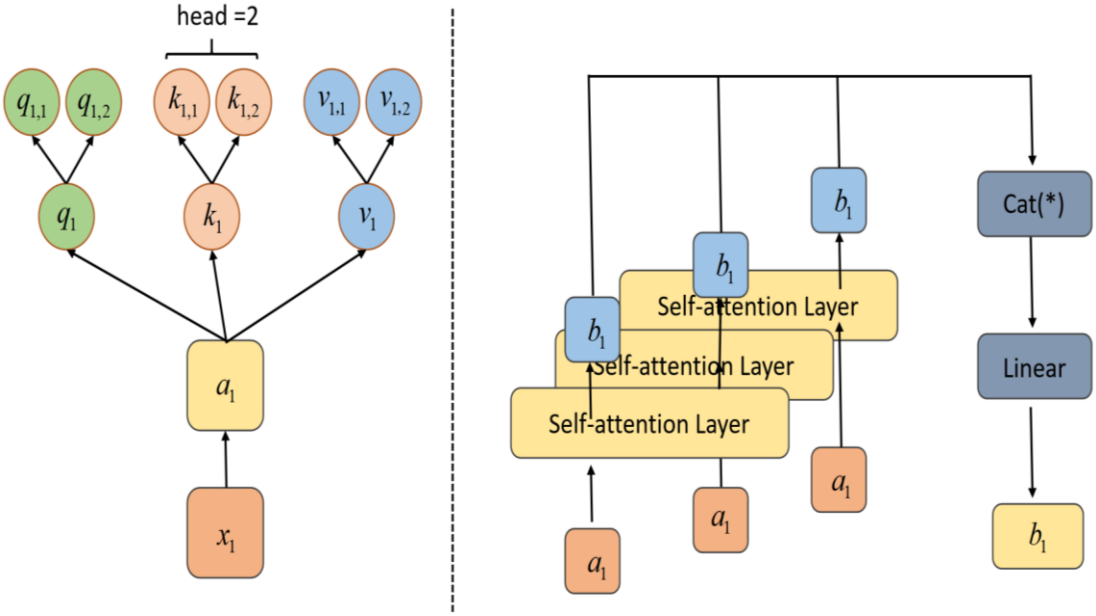


Figure 2: Schematic diagram of multi-head self-attention mechanism calculation.

$$MHSA(Q,K,V) = [Z_0, Z_1, \cdots, , Z_{h-1}]W^O \quad (3)$$

$$Z_i = \text{softmax}\left(\frac{QW^{Q_i}(KW^{K_i})^T}{\sqrt{\frac{D_q}{h}}}\right)VW^{V_i} \quad (4)$$

Since the complexity of computing the soft max of the SA block is quadratic in the length of the input sequence, the SA block has a high computational overhead for processing long sequences such as high-resolution medical images, which limits its applicability. Recently, many efforts have been made to reduce the complexity, including sparse attention, linearised attention, low-rank attention, memory compression-based methods and improved MHSA.

2.2 Proposed algorithm

2.2.1 Sliding Window Transformer Module

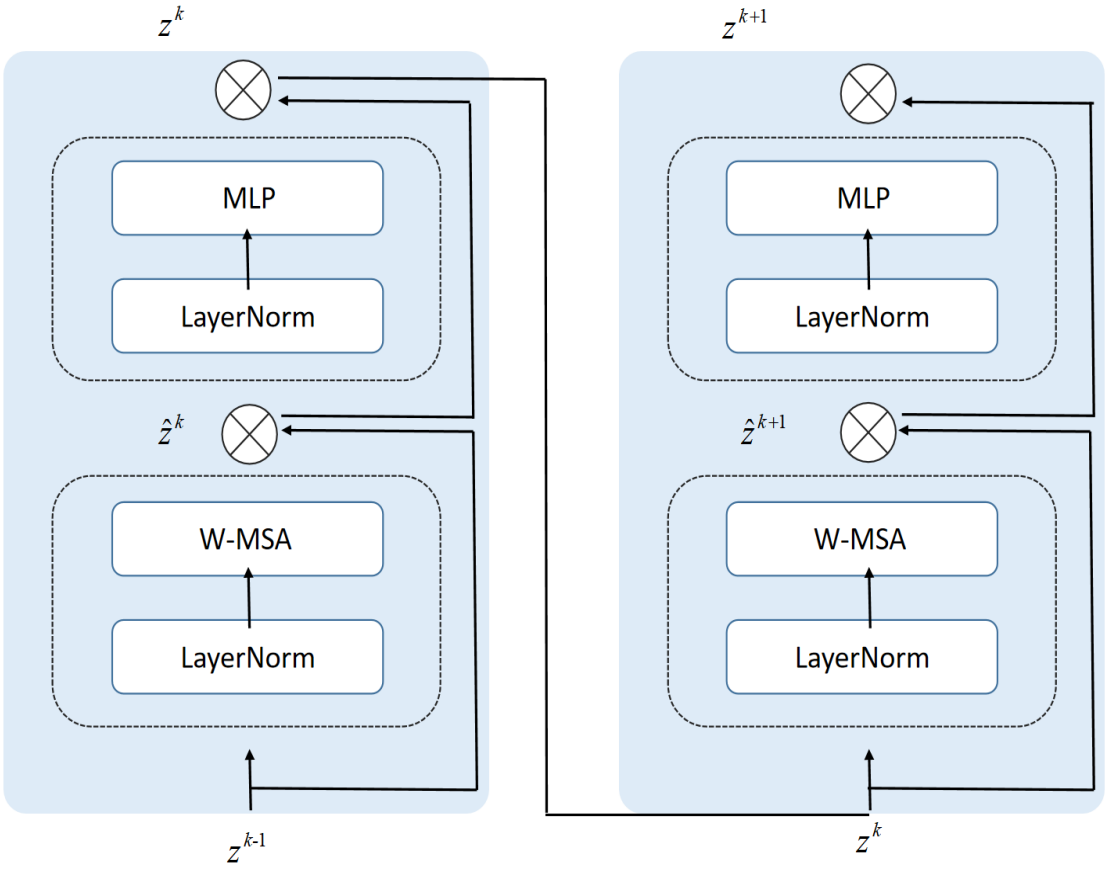


Figure 3: Schematic diagram of multi-head self-attention mechanism calculation.

The Sliding Window Transformer module consists of two consecutive sub-modules, each Sliding Window Transformer sub-module consists of LayerNorm regularisation, Multihead Self-Attention Module, Residual Connection, and Multi-Layer Perceptron (MLP) with Nonlinear Activation

Function, as shown in Figure 3 below. The Window-based Multihead Self-Attention Module (W-MSA) and Sliding Window-based Multihead Self-Attention Module (SW-MSA) are located in two consecutive sub-modules, where a LayerNorm layer is added before each MLP module and each MSA module, and they are connected after each module using a residual approach. It can be expressed by the following equation:

$$\hat{z}^k = W - MSA\left(LN(z^{k-1})\right) + z^{k-1} \tag{5}$$

$$z^k = MLP\left(LN(\hat{z}^k)\right) + \hat{z}^k \tag{6}$$

$$\hat{z}^{k+!} = SW - MSA\left(LN(z^k)\right) + z^k \tag{7}$$

$$z^{k+1} = MLP\left(LN(\hat{z}^{k+1})\right) + \hat{z}^{k+1} \tag{8}$$

The difference between SW-MSA and W-MSA is that the input features of the latter are divided into non-overlapping windows, each of which generally contains $M \times M$ blocks in the 2-dimensional plane and $M \times M \times M$ blocks in the 3-dimensional space. They will only complete self-attention within a local window, and the lack of effective information interaction between windows is very unfavourable for long-range feature extraction.

To solve this in order to solve the problem without adding extra computation, an MSA with offset windows, i.e., SW-MSA, is added. by a kind of cyclic shifting efficient batch processing, some of the windows in the partitioned batch consist of multiple non-adjacent sub-windows in the feature map while maintaining the same number of windows as the regular partitioning. Due to the presence of the shifts, the computation of a set of two consecutive sub-Transformer modules is completed each time, and the information between different windows can be transferred efficiently, which significantly enhances the modelling capability.

2.2.2 Model Structure

Previously, related work based on Transformer networks has been applied to medical image segmentation tasks. In these works, the Transformer module they set up was either used as an auxiliary module after the convolutional layer operation, or the Transformer module was simply used as a feature encoder at the deepest layer, resulting in a significant reduction in the advantages of the Transformer module for spatial contextual feature extraction. In contrast to previous work that speaks of Transformer as an auxiliary encoder, this section, inspired by sliding windows, proposes an improved TF-UNet

network model that directly applies Transformer to encode voxel images and positional embeddings in a more direct way, and the model structure is shown in Figure 4.

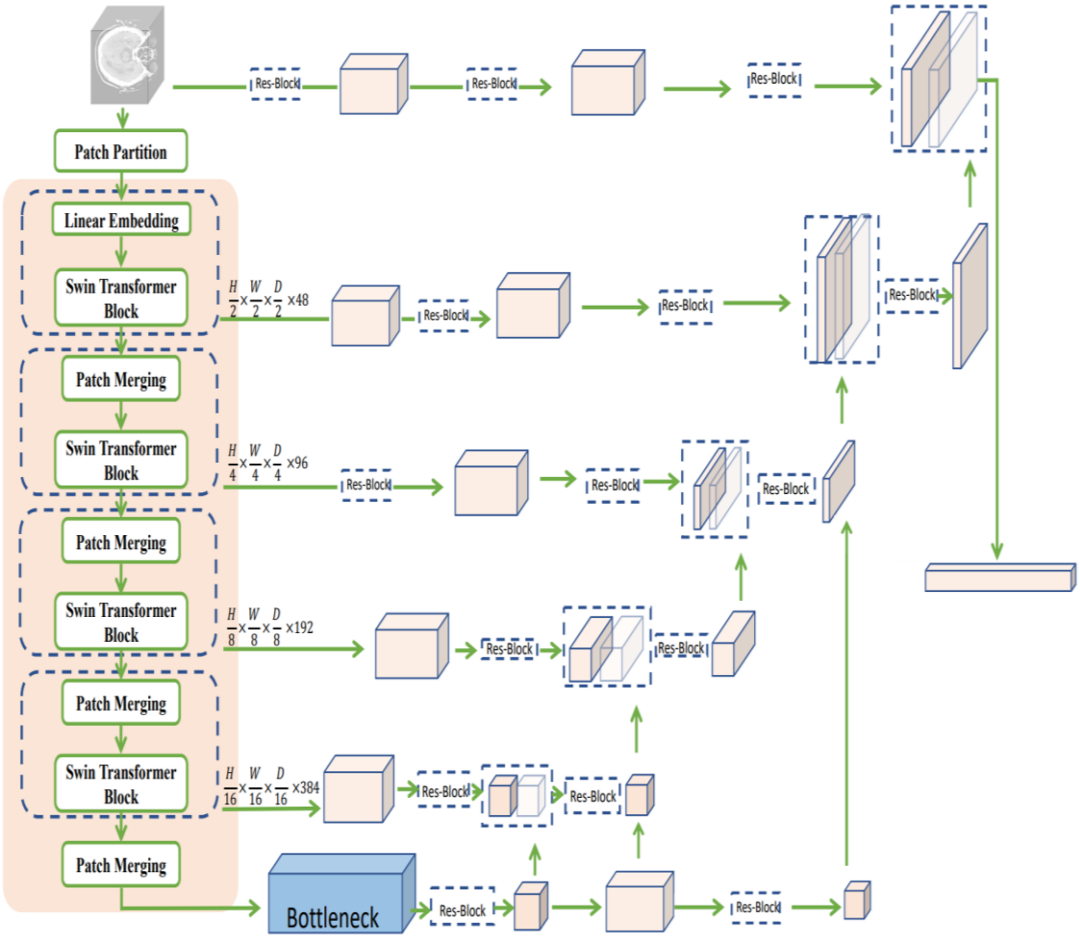


Figure 4: Schematic representation of the proposed model structure.

The encoder divides the input 3D medical image into non-overlapping voxel blocks by means of a Patch Partition module, where each voxel block is treated as a token, assuming a 3D voxel size of (H',W',D'). The decoder feeds the feature representation extracted from the Bottleneck into a Res-Block consisting of two 3 × 3 × 3 convolutional layers with Instance normalisation.

The output is up-sampled using the inverse convolutional layer and then spliced with the encoder's input of features in the same dimension as the residual block. The stacked features are used as the current layer's features and are sent to the upper layers, again going through the above steps, until they are upsampled to the model input voxel resolution. In addition, down sampling causes loss of information between context spaces, to circumvent this problem, this paper stacks and fuses multiple shallow features from the Transformer encoder at different scales with up sampled deep features from the CNN decoder.

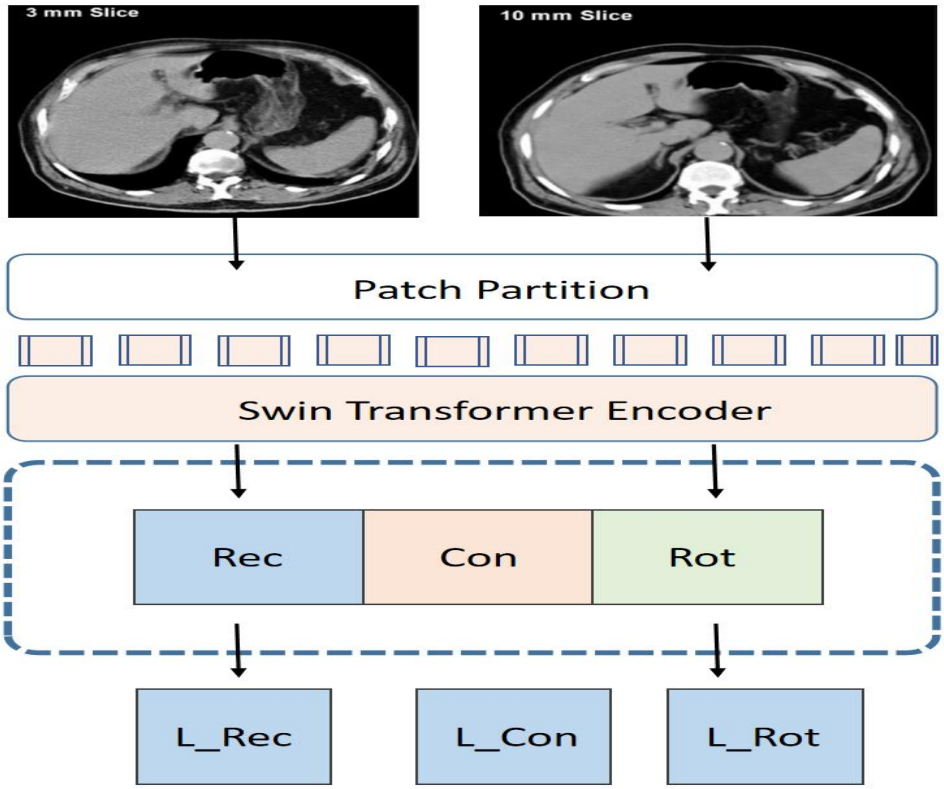


Figure 5: Schematic representation of a self-supervised pre-training agent task.

2.3 Self-Supervised Pre-Training Agent

Depending on the photographer's shooting time or shooting place, the objects and scenery in natural scene images will present a variety of different spatial layouts. Unlike human medical images, the anatomical structures of human organs in images imaged by CT and other means present natural and consistent contextual information, i.e., the spatial distribution of different tissues or foci satisfies the basic laws of human anatomy. Therefore, with this feature, by designing suitable agent tasks, the underlying patterns of human anatomy can be learnt, enabling the model to learn deeper semantic and contextual information for medical images. For 3D medical images, it has more rich semantic information across planes between different slices than 2D images. To address this characteristic, the article designs three agent tasks of contrast coding, voxel rotation and mask reconstruction to achieve better image representation in terms of image similarity, geometric space and contextual information respectively, as shown in Figure 5.

2.3.1 Comparative Encoding

In visual representation learning, self-supervised contrast coding performs well in migration downstream tasks. Under normal circumstances, often multiple pairs of samples from the same input image after augmentation and generalisation transformations are regarded as positive samples, while samples generated from different input images are regarded as negative

samples. The task of contrast coding is to make the features extracted from these samples in the same feature space, with the positive sample pairs as close as possible, while the negative sample pairs are as far away from each other as possible, so as to make the feature map cover the semantic information of the input image more fully. This is done by taking the encoder of the proposed model separately and attaching a self-supervised comparison header containing a linear layer to the tail, and mapping the augmented input to the feature space denoted as z. Using cosine similarity as a distance metric for the encoded representation, a small batch of samples of the number N is randomly sampled, and the image is passed through two different augmentations to generate $2N$ pairs of samples. Positive sample pairs z_i and z_j, and the rest of $2(N-1)$ are encoded as contrasts between negative samples with respect to them, and the loss function is defined by the following formula.

$$L_{\mathrm{Con}} = -\log \frac{\exp\left(\frac{\mathrm{sim}\left(z_i, z_j\right)}{\tau}\right)}{\sum_{k=1}^{2N} \lambda_{|k \neq i|} \exp\left(\frac{\mathrm{sim}\left(z_i, z_j\right)}{\tau}\right)} \tag{9}$$

where $\lambda_{|k \neq i|} \in \{0,1\}$ is used to distinguish positive and negative sample pairs, which is equal to 1 if and only if $k \neq i$, otherwise it is 0. τ is the normalized temperature hyperparameter to facilitate gradient backpropagation. sim represents the l_2 norm.

2.3.2 Loss Function

The encoder of the proposed model is trained to minimize the total loss function by using multiple pre-training objectives such as image rotation, mask reconstruction and contrastive encoding, as follows:

$$L_{total} = \lambda_1 L_{Re\,c} + \lambda_2 L_{\mathrm{Con}} + \lambda_3 L_{\mathrm{Rot}} \tag{10}$$

where $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

3. Experiment and Results

3.1 Datasets

(1) MSD: The MSD consists of 10 segmentation tasks from different organs and image modalities, as shown in Figures 3-5, 6 sets of CTs and 4 sets of MRIs. These tasks are designed to address difficulties across medical images, including issues such as small training sets, small targets, unbalanced samples, and multimodal data. Therefore, the MSD challenge can be used as a comprehensive benchmark for assessing the generality of medical image segmentation methods. Some examples of the MSD dataset are shown in Figure 6.

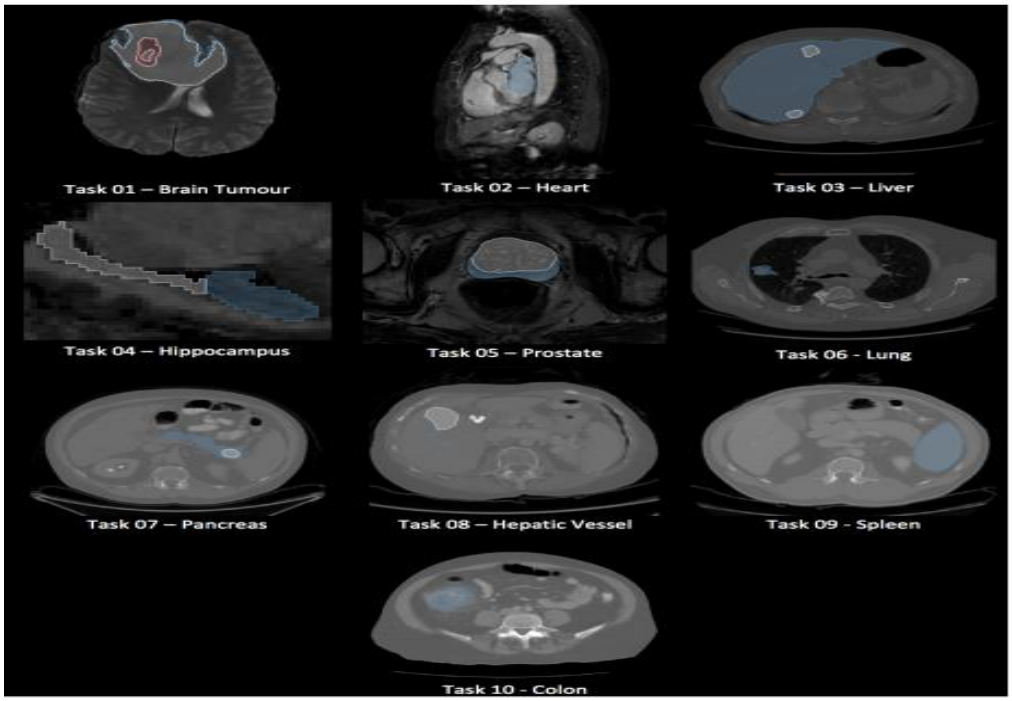


Figure 6: Examples of MSD dataset.

(2) BTCV: The BTCV contains CT scans of 30 abdominal subjects, of which 13 organs were labelled under the supervision of a radiologist at Vanderbilt University Medical Centre. All CT scans were contrast-enhanced at the portal vein and consisted of 80 to 225 512 × 512-pixel slices with slice thicknesses ranging from 1 to 6 mm, some examples of the BTCV dataset are shown in Figure 7.

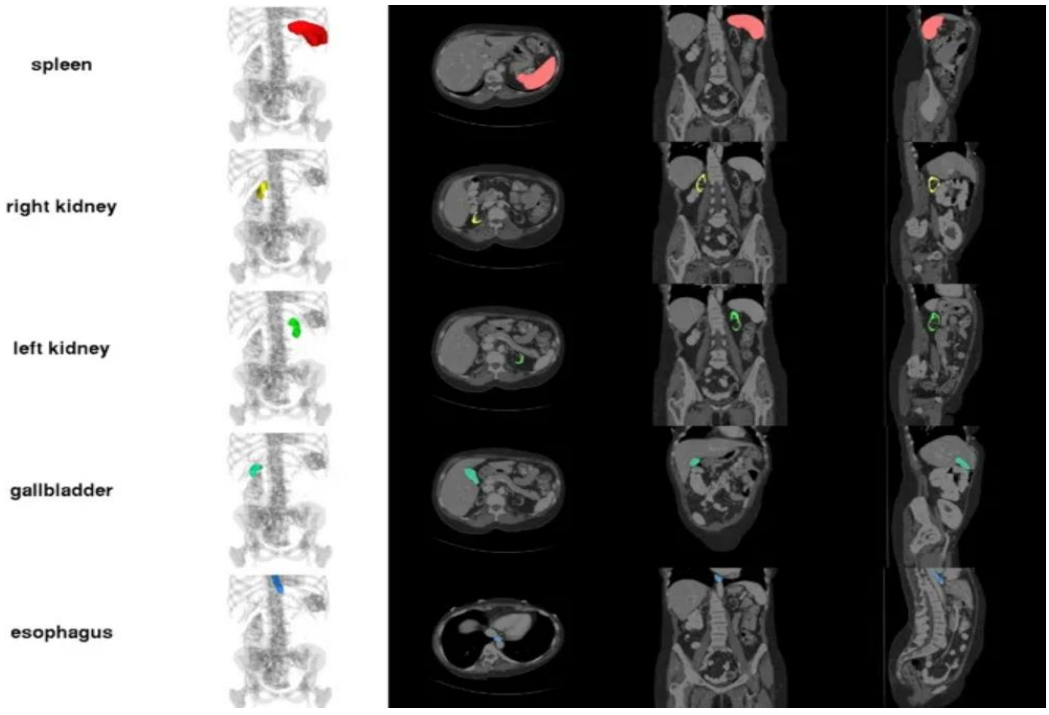


Figure 7: Examples of MSD dataset.

3.2 Experimental Setup

The experimental setup consists of two parts, the experimental environment setup used during the experiment and the hyperparameter setup of the model used, as shown in Tables 1 and 2.

Table 1: Experimental environment setup.

TYPE	PARAMETERS
OS	Ubuntu 16.04
GPU	Nvidia RTX 4090
RAM	24G
PYTHON	3.6.5
TRANSFORMERS	3.0.2
PYTORCH	1.4.0

Table 2: Hyperparameter settings.

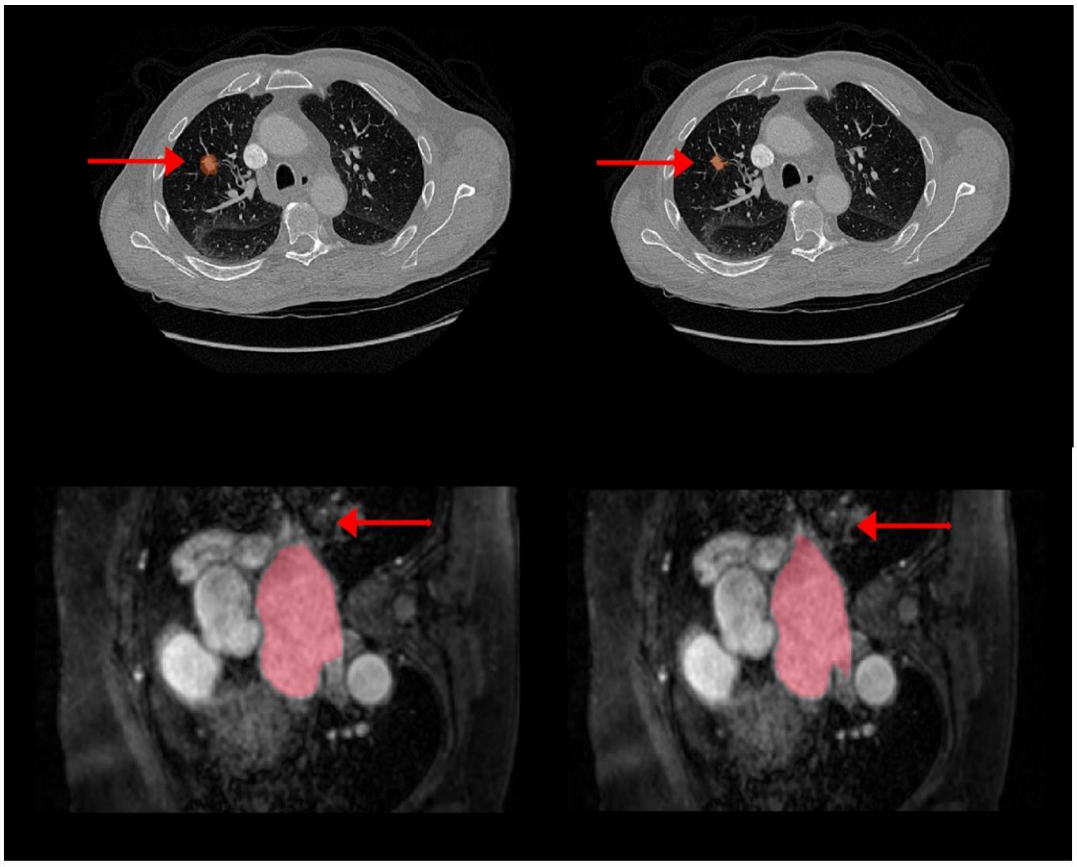
HYPERPARAMETER	VALUE
BATCH_SIZE	32
DROPOUT	0.4
OPTIMIZER	Adam
LEARNING RATE	3e-5
EPOCH_NUM	2000

3.3 Experimental Results and Analysis

The experiments were conducted for detailed experimental investigation of 10 kinds of task data on the MSD Challenge data. On the CT dataset, only the parameters of the pre-training weights were fine-tuned, including lungs, liver, colon, pancreas, and spleen; on the MRI dataset, such as the heart, brain tumour, prostate, and hippocampus, the experiments were trained from the beginning to be fair due to the domain differences between the CT images and the MRI images. The experimental results for each MSD task are shown in Table 3 and Figure 8. The experiments selected several current mainstream models and TF-UNet for comparison, including nn-UNet, DiNTs, TransUNet and TransBTS, etc. The experiments used U-Net as the benchmark model to make the experimental results more intuitive. In addition to the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD95), the Normalised Surface Distance (NSD) for voxels is also used to evaluate the results of the MSD experiments, and the NSD measures the degree of overlap between the predicted and the true values of the voxel surfaces, which is a better representation of the degree of consistency between two structures. express the consistency metric between the two structures.

Table 3: Results of the comparison of the comprehensive performance of the proposed method with other methods on the MSD dataset.

METHOD	AVERAGE ACCURACY		
	DSC (%)	NSD (%)	HD95 (mm)
U-NET	60.31	71.69	37.23
NN-UNET	68.17	79.36	27.12
DINTS	73.26	82.57	11.28
TRANSBTS	72.69	83.02	10.98
OURS	75.11	84.21	10.23



(a) DiTNS (b) Ours

Figure 8: Visualisation of results of comparative experiments.

From Figure 8, it is easy to see that both TransBTS and TF-UNet are still able to segment clearly in the core tumour area, but outward to the enhanced tumour area and the whole tumour area, the TransBTS segmentation contour starts to tear, and the segmentation result is not as good as that of TF-UNet at several marked points in the figure. In addition, from Table 3, it can be seen that the proposed method achieved the best performance with an average DICE of 75.11%, DSC of 84.21% and HD of 10.23mm to achieve the best performance. In addition, the results of the comparison experiments on the BTCV dataset are given in Table 4, again demonstrating the superiority of the proposed method.

Table 4: Results of the comparison of the comprehensive performance of the proposed method with other methods on the BTCV dataset.

METHOD	AVERAGE ACCURACY		
	DSC (%)	NSD (%)	HD95 (mm)
U-NET	80.65	85.36	23.69
NN-UNET	81.25	86.95	22.17
DINTS	82.23	88.64	21.25
TRANSBTS	83.67	89.27	12.69
OURS	85.16	90.02	11.36

4. Conclusion

In this paper, a new architecture for self-supervised medical image segmentation for athletes is proposed. The difficulty of the lack of large-scale labelled data for athlete's medical images is addressed by designing an agent task for pre-training so that the model can better extract and process the visual information of medical images, and then fine-tuning it on the segmentation task. The effectiveness of the proposed algorithm in the medical image segmentation task is verified through a large number of experiments conducted on two mainstream datasets. And the comparison with other mainstream models shows that the model performs well in most scenarios. In future work we will explore more efficient and reliable self-supervised pre-training agent tasks for athletes to further improve the robustness and applicability of the model, and further research on how to improve the computational efficiency of the Transformer model for athletes.

Reference

Arabahmadi, M., Farahbakhsh, R., & Rezazadeh, J. (2022). Deep learning for smart Healthcare—A survey on brain tumor detection from medical imaging. *Sensors*, *22*(5), 1960.

Claudino, J. G., Capanema, D. d. O., de Souza, T. V., Serrão, J. C., Machado Pereira, A. C., & Nassis, G. P. (2019). Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sports Medicine-Open*, *5*, 1-12.

Conze, P.-H., Andrade-Miranda, G., Singh, V. K., Jaouen, V., & Visvikis, D. (2023). Current and emerging trends in medical image segmentation with deep learning. *IEEE Transactions on Radiation and Plasma Medical Sciences*, *7*(6), 545-569.

Farshad, A., Yeganeh, Y., Gehlbach, P., & Navab, N. (2022). Y-net: A spatiospectral dual-encoder network for medical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention,

Hu, M., Zhong, Y., Xie, S., Lv, H., & Lv, Z. (2021). Fuzzy system based medical

image processing for brain disease prediction. *Frontiers in Neuroscience*, *15*, 714318.

Isensee, F., Jäger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2019). Automated design of deep learning methods for biomedical image segmentation. *arXiv preprint arXiv:1904.08128*.

Khan, T. M., Robles-Kelly, A., & Naqvi, S. S. (2022). T-Net: A resource-constrained tiny convolutional neural network for medical image segmentation. Proceedings of the IEEE/CVF winter conference on applications of computer vision,

Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. *Sustainability*, *13*(3), 1224.

Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., & Martel, A. L. (2021). Loss odyssey in medical image segmentation. *Medical Image Analysis*, *71*, 102035.

Qureshi, I., Yan, J., Abbas, Q., Shaheed, K., Riaz, A. B., Wahid, A., Khan, M. W. J., & Szczuko, P. (2023). Medical image segmentation using deep semantic-based methods: A review of techniques, applications and emerging trends. *Information Fusion*, *90*, 316-352.

Ramkumar, P. N., Luu, B. C., Haeberle, H. S., Karnuta, J. M., Nwachukwu, B. U., & Williams, R. J. (2022). Sports medicine and artificial intelligence: a primer. *The American journal of sports medicine*, *50*(4), 1166-1174.

Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE access*, *9*, 82031-82057.

Song, H., & Montenegro-Marin, C. E. (2021). Secure prediction and assessment of sports injuries using deep learning based convolutional neural network. *Journal of Ambient Intelligence and Humanized Computing*, *12*(3), 3399-3410.

Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., & Nandi, A. K. (2022). Medical image segmentation using deep learning: A survey. *IET image processing*, *16*(5), 1243-1267.

Wu, X., Zhou, J., Zheng, M., Chen, S., Wang, D., Anajemba, J., Zhang, G., Abdelhaq, M., Alsaqour, R., & Uddin, M. (2022). Cloud-based deep learning-assisted system for diagnosis of sports injuries. *Journal of Cloud Computing*, *11*(1), 82.

Xiao, J., Liu, H., Wei, D., & Xiong, L. (2023). APPLICATION PROSPECT OF BONE SUBSTITUTE MATERIALS IN ORTHOPEDIC TRAUMA REPAIR UNDER CT IMAGES FOR IMPROVED RECOVERY OF ATHLETE PLAYERS. *rimcafd*, *23*(89).

Zhan, C. (2024). Application of artificial intelligence in the development of personalized sports injury rehabilitation plan. *Molecular & Cellular Biomechanics*, *21*(1), 326-326.