Proceedings of the 7th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2025.25644

Cerebral Edema Image Segmentation Based on Improved U-Net Modeling

Jiayang Zhuo

University of Amsterdam, Amsterdam, Netherlands
jiayang.zhuo@gmail.com

Abstract. In this study, for the clinical needs of accurate segmentation of brain edema region, an improved U-net model based on deep learning is proposed, which achieves a significant improvement in the performance of medical image segmentation by integrating three core modules, namely, multi-scale perception, lightweight convolution and attention mechanism. First, the Min-ASPP module is introduced in the encoder part, which solves the feature dilution problem of traditional ASPP in tiny lesion recognition by effectively capturing multi-scale contextual information through an improved void space pyramid structure; second, the decoding path is reconstructed by using the GSConv module, whose hybrid architecture of depth-separable convolution and standard convolution maintains the feature expression capability while keeping the model parameter number is reduced; finally, the AG attention mechanism is embedded at the jump junction, which strengthens the feature differentiation between edematous regions and normal tissues through adaptive feature calibration. After the histogram equalization and elastic deformation enhancement in the data preprocessing stage, the model shows excellent performance indexes on the test set: the Dice coefficient reaches 94.32%, and the IoU value reaches 91.93%. The clinical application value of this achievement lies in the fact that the high-precision and high-efficiency automatic segmentation algorithm can assist clinicians to quickly locate the scope of edema, quantitatively assess the progress of the disease, and provide a reliable basis for surgical planning and efficacy assessment, which is of great practical significance in promoting the intelligent diagnosis and treatment of stroke, craniocerebral injury and other diseases.

Keywords: Brain edema segmentation, U-net, Min-ASPP module, GSConv module, AG attention mechanism.

1. Introduction

Cerebral edema is a pathological state caused by the abnormal accumulation of water in brain tissue, which is commonly seen in stroke, traumatic brain injury, tumors and infections, and can lead to increased intracranial pressure, hypoxia of brain tissue and even brain hernia, seriously threatening patients' lives [1]. Accurate identification and segmentation of brain edema region is of great significance for clinical diagnosis, treatment plan development and prognosis assessment. Traditional medical image segmentation methods rely on manual feature extraction, which is

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.25644

difficult to cope with the complex characteristics of cerebral edema, such as variable morphology, blurred boundaries, and low contrast with normal tissues, and is susceptible to noise interference, with limited generalization ability [2]. With the popularization of medical imaging technology, massive data provides the basis for automated segmentation, but how to efficiently process high-dimensional and multimodal images and realize accurate segmentation has become a key challenge in clinical and scientific research.

Deep learning significantly improves the accuracy and efficiency of brain edema segmentation by constructing a multilayer neural network to automatically learn deep image features [3]. The encoder-decoder architecture represented by U-Net, combined with jump connections to preserve multi-scale spatial information, has become a benchmark model for medical segmentation. In this paper, we optimize the U-net model based on Min-ASPP module, ASPP module and AG attention mechanism for segmentation of brain edema images.

2. Related work

Brain edema segmentation is a key task in medical image analysis, and traditional methods rely on manual feature extraction or shallow neural networks, which are difficult to cope with complex lesion morphology and fuzzy boundaries [4]. In recent years, U-net and its variants have improved the performance of medical image segmentation through jump-joining and encoder-decoder architectures, such as incorporating the ASPP (Atrous Spatial Pyramid Pooling) module to enhance multi-scale feature representation, or introducing attention mechanisms (e.g., SE, CBAM) to optimize feature fusion [5]. However, existing methods still face challenges: the standard ASPP module has a large number of parameters and is prone to introduce redundant computations, traditional convolution operations have difficulty in balancing lightweighting and feature discriminative properties, and shallow attention mechanisms have limited ability to model long-range dependencies [6]. Some studies have attempted to combine grouped convolution (e.g., MobileNet), but more efficient multi-scale modeling and feature optimization mechanisms are still needed for the variable scale and low contrast of lesions in brain edema segmentation.

Cerebral edema lesions are highly heterogeneous in images, and existing U-net variants often have limited segmentation accuracy due to insufficient multiscale feature fusion or computational redundancy. Standard ASPP extracts multiscale features by parallel convolution with different null rates, but complex structures are prone to overfitting in small datasets; traditional convolutional layers are computationally expensive, while lightweight designs may lose spatial detail information [7]. In addition, the low contrast between brain edema regions and normal tissues requires stronger context-awareness and feature refinement. The Min-ASPP module proposed in this paper streamlines the ASPP structure and enhances local-global feature interactions, and the GSConv module balances efficiency and accuracy through a hybrid design of grouped convolution and standard convolution, and suppresses irrelevant background interference by combining with the AG attention mechanism, which can be targeted to solve the above problems, and provide a more reliable lesion localization and quantification tool for the clinic [8].

The main contributions of this paper:

(1) Designing a lightweight Min-ASPP module to reduce the number of parameters through hierarchical null convolution and feature compression strategies, while fusing local details and global contextual information to improve the characterization of multi-scale features of brain edema;

(2) Introducing the GSConv module to replace part of the standard convolution, and utilizing grouped convolution and channel blending to enhance the feature multiplexing efficiency, and retaining the discriminative features while reducing the computational cost;

DOI: 10.54254/2755-2721/2025.25644

(3) Embedding the AG attention mechanism in the jump connection to dynamically calibrate the encoder and decoder features to suppress the background noise and strengthen the lesion region response.

3. Dataset source and data preprocessing

The dataset used in this paper is from the open source dataset at (https://www.kaggle.com/datasets/nikhilroxtomar/brain-tumor-segmentation/data). The dataset contains 3064 pairs of MRI brain images and their respective binary masks representing the tumor.

Selected portions of the dataset are shown in Figure 1. The images shown in the first row are the original images and the images shown in the second row are the corresponding masks.

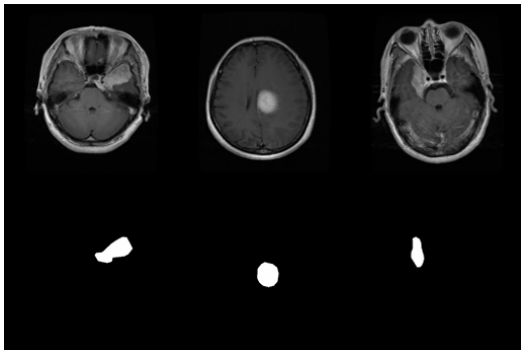

Figure 1. Selected portions of the dataset

We statistically analyzed the size of the images and learned that each image is composed of 512 x 512 pixels, which matches the format that the model is able to handle without the need for cropping and rotation.

We output the histogram of the image for subsequent image enhancement studies. From the histogram, it is clear that the dataset used in this experiment has potential for further image enhancement. The image histogram is shown in Figure 2.

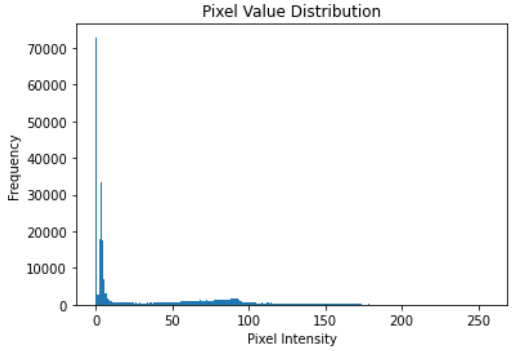

Figure 2. The image histogram

We use histogram equalization for image enhancement with the aim of enhancing the contrast of the lesion region.

Histogram equalization is a classical image enhancement technique that is mainly used to improve the global contrast of an image. Its core principle is to extend the gray level distribution, which is originally concentrated in a narrow interval, to a wider range by redistributing the pixel gray level values. The method first counts the grayscale histogram of the original image, calculates

DOI: 10.54254/2755-2721/2025.25644

the cumulative distribution function (CDF) of each grayscale level, and then maps the original grayscale values to the new uniformly distributed interval according to the CDF. This nonlinear transformation can effectively enhance dark or bright details, which is especially suitable for medical images, low exposure photos and other scenes with insufficient contrast. The denoised image is shown in Fig. 3. Finally, we perform noise detection on the image and after checking there is no noise and no need to denoise the image.

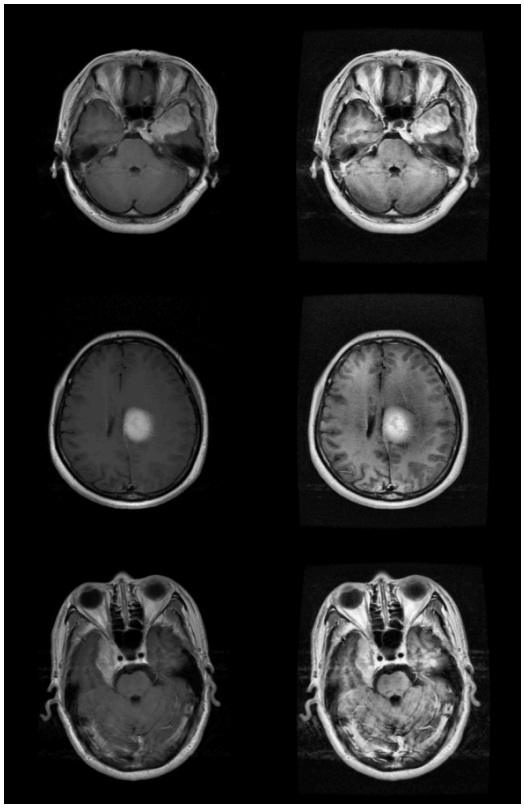

Figure 3. The denoised image

4. Method

4.1. U-net

U-net is a classical convolutional neural network model, and its core design is based on an encoder-decoder structure, which realizes the balance between feature extraction and image reconstruction through a symmetric U-shaped network. The encoder part consists of multiple downsampling modules, each of which contains a convolutional layer and a pooling layer to gradually extract the deep semantic features of the image and reduce the spatial dimensions; the decoder part gradually recovers the image details and spatial resolution through upsampling operations, and ultimately outputs the segmentation results matching the input dimensions [9]. The network structure of the Unet model is shown in Fig. 4.

DOI: 10.54254/2755-2721/2025.25644

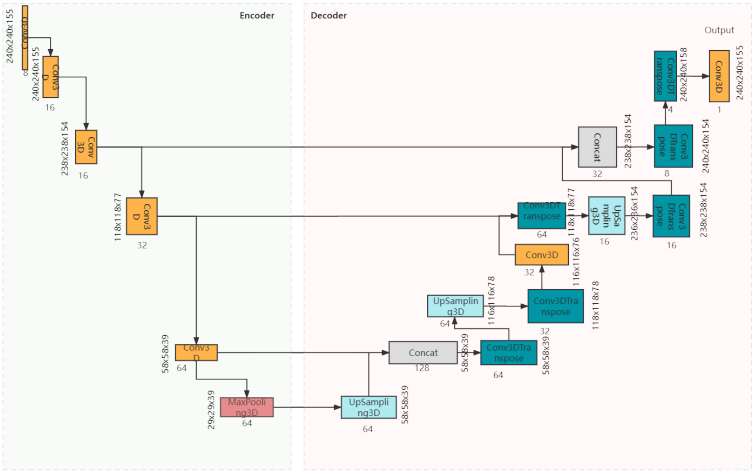

Figure 4. The network structure of the Unet model

The core innovation of UNet lies in its jump connection mechanism and symmetric network design. The feature maps of each stage of the encoder are passed directly to the corresponding layer of the decoder through jump connections, enabling the decoder to utilize both shallow high-resolution details and deep abstract semantic information in the up-sampling process. This multi-scale feature fusion approach is particularly important in medical images because lesion regions often have blurred boundaries and low contrast with normal tissues. In addition, UNet employs a fully convolutional structure to avoid parameter redundancy in the fully connected layers and mitigates the challenge of insufficient sample size of medical images through data enhancement techniques. Its loss function usually adopts cross-entropy or Dice coefficient loss, the latter is especially suitable for scenarios with category imbalance.

4.2. AG Attention Gate

AG Attention Gate is a dynamic feature enhancement method, which realizes accurate focusing on key information by autonomously calculating the importance weights of each region of the feature map.The network structure of AG Attention Gate is shown in Fig. 5 [10].

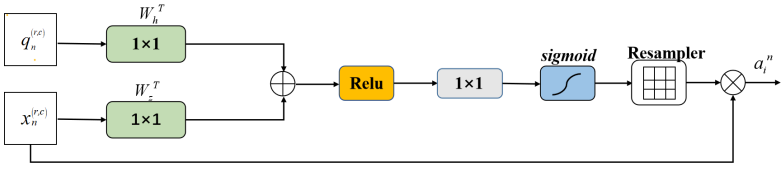

Figure 5. The network structure of AG Attention Gate

Its core principle lies in constructing the gating structure and utilizing the interaction between input features and gating signals to generate the attention coefficients. Specifically, AG is realized through the following steps: firstly, the input feature map is fused with the gating signal for feature fusion, usually by element-by-element summation or splicing; subsequently, the spatial attention map is generated by convolutional layer and activation function, which normalizes the weights to the interval of [0,1] by the Sigmoid function; finally, the attention map is multiplied with the original feature map to highlight the effective region and suppress the noise. This mechanism essentially

DOI: 10.54254/2755-2721/2025.25644

establishes the semantic association of cross-layer features, and guides the bottom-layer feature selection through the high-level semantics, forming an adaptive feature screening mechanism.

GSConv is a lightweight convolutional module for efficient inference, and its core design idea is to significantly reduce the computational complexity while maintaining the model accuracy by fusing standard convolution, depth-separable convolution, and channel-mixing operation. The module mainly targets the demand for real-time performance of mobile/edge devices, and strikes a balance between computational efficiency and feature expression capability through structural innovation, and is widely used in computer vision tasks such as target detection.The network structure of GSConv is shown in Fig. 6.

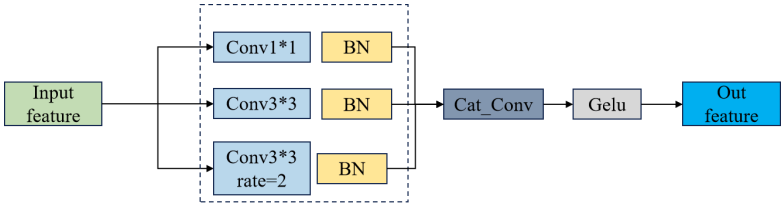

Figure 6. The network structure of GSConv

GSConv consists of three key components: a standard convolutional layer, a depth-separable convolutional layer, and a channel-mixing layer. The module first performs two branches of processing in parallel: the main branch uses standard convolution for feature extraction, and the auxiliary branch uses depth-separable convolution (depth convolution + point-by-point convolution). The output feature maps of the two branches are spliced in the channel dimension, and then interacted with each other through the channel shuffling operation. This hybrid architecture not only retains the strong characterization ability of standard convolution, but also significantly reduces the number of parameters through depth-separable convolution.

4.3. Min-ASPP

Min-ASPP (Minimal Atrous Spatial Pyramid Pooling) is a lightweight multi-scale feature fusion module that aims to streamline the traditional ASPP structure while maintaining the ability to efficiently capture multi-scale contextual information. Traditional ASPP extracts multiscale features by stacking multiple null convolution and global pooling layers with different expansion rates in parallel, but the computational cost is high and the redundant design of expansion rates may introduce noise.The core design idea of Min-ASPP is to minimize the number of branches and optimize the combination of expansion rates, so that the key scale features can be retained while reducing the computational complexity. Specifically, it discards the very high expansion rate convolutional layer in ASPP, retains only the expansion rate that is sensitive to small- and medium-scale information, and combines 1×1 convolution with a global feature layer to significantly reduce the number of parameters. In addition, Min-ASPP introduces depth-separable convolution instead of standard convolution to further compress the computational overhead, and dynamically adjusts the importance of each branch feature through the adaptive weight fusion mechanism to suppress invalid context interference and enhance the complementarity of cross-scale features.The network structure of Min-ASPP is shown in Fig. 7.

DOI: 10.54254/2755-2721/2025.25644

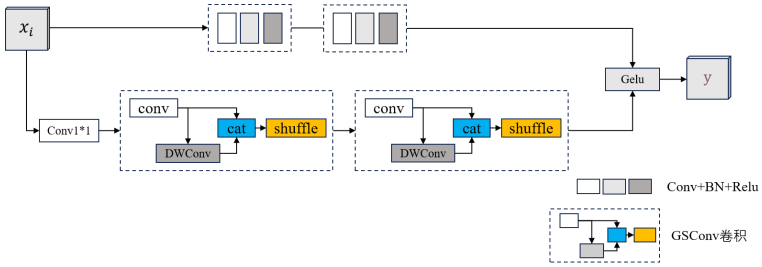

Figure 7. The network structure of Min-ASPP

At the implementation level, Min-ASPP typically contains three parallel branches: a 1×1 convolution for local feature extraction, two depth-separable null convolutions to capture contexts with medium sensory fields, and a global average pooling layer to provide image-level semantic prior. After the outputs of all branches are unified in resolution by bilinear interpolation, a channel attention mechanism is used to automatically learn the fusion weights of each channel for efficient feature aggregation.

4.4. U-net improvement strategy

The U-net model optimized based on Min-ASPP module, GSConv module and AG attention mechanism mainly improves the segmentation performance through multi-scale feature extraction, lightweight design and attention-guided feature fusion. The specific network structure is shown in Fig. 8.

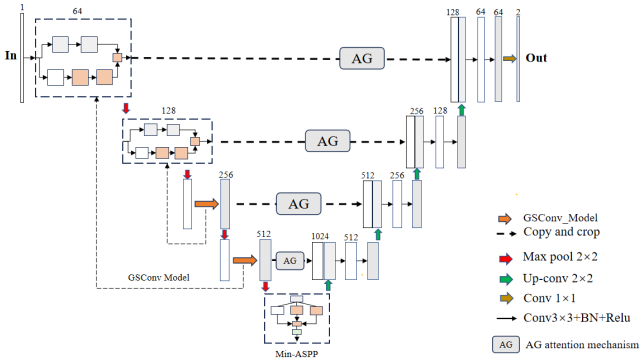

Figure 8. The specific network structure

In the encoding stage, the Min-ASPP module captures multi-scale context information with different expansion rates through a refined and simplified null convolution pyramid structure, which enhances the model's ability to sense large and small targets while reducing parameter redundancy; combined with the GSConv module (which fuses Ghost technology with depth-separable convolution), it breaks down the standard convolution into a low-cost linear operation with lightweight convolution, significantly reduces the computational complexity of encoding and decoding paths, and realizes model lightweighting under the premise of guaranteeing feature expression capability, which is especially suitable for mobile or real-time scenes.

In the feature fusion stage, the AG attention mechanism is embedded in the jump connection, which automatically assigns attention weights, suppresses background noise and strengthens the key details of the target region by dynamically learning the spatial correlation between low-level features and high-level semantic features. This mechanism optimizes the fusion effect of cross-level

DOI: 10.54254/2755-2721/2025.25644

features, enabling the decoder to recover target edges and fine structures more accurately, especially in complex occlusion or low-contrast scenarios, effectively improving segmentation accuracy and robustness. The three work in tandem to balance model efficiency, multi-scale perception and detail reconstruction capability.

5. Result

In terms of software, python 3.12 is used in this experiment, in terms of hardware configuration, 32G of RAM, 4090 graphics card is used, epoch is set to 100, batch size is set to 16, and training, validation and test sets are divided according to 40%, 40% and 30%. The segmentation effect of the model is evaluated using Dice and IoU.

Firstly, the loss changes of the training set and validation set during the training process of the model proposed in this paper are output, as shown in Fig. 9.As can be seen from the loss change of the training set, the loss value eventually decreases from the initial 0.971 to 0.062, indicating that the image segmentation effect on the training set achieves good results; as can be seen from the loss change of the validation set, the loss value gradually converges from the initial 0.914 to 0.317, indicating that the model has been generalized very well, and it also exhibits very good results on the validation set .

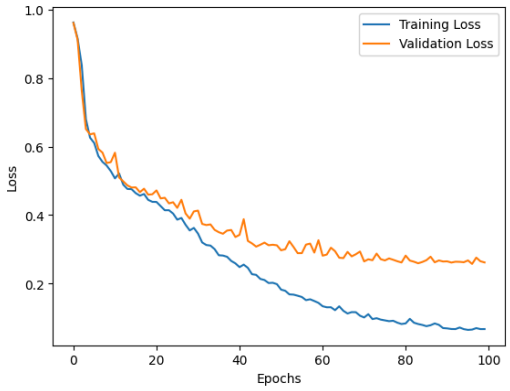

Figure 9. The loss changes of the training set

The image segmentation results of the model are evaluated using the test set and the output image segmentation results are shown in Fig. 10. From the segmentation results on the test set, it can be seen that the model proposed in this paper achieves a segmentation performance of 94.32% for the Dice coefficient and 91.93% for the IoU metrics, and is able to segment the image of the brain edema portion very well.

DOI: 10.54254/2755-2721/2025.25644

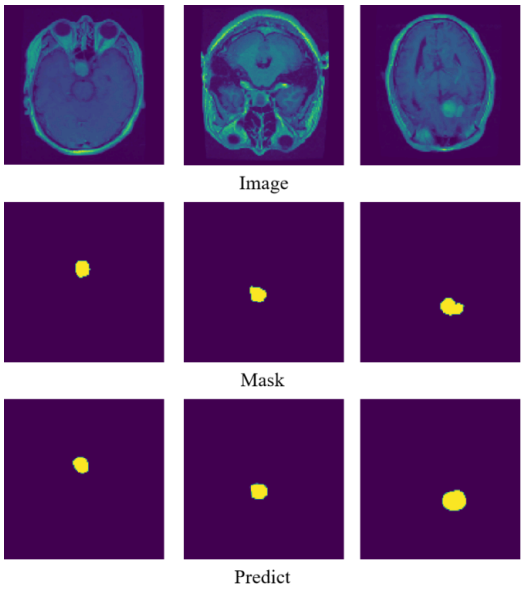

Figure 10. The output image segmentation results

6. Conclusion

In this study, an improved U-net segmentation model based on deep learning is proposed for the clinical needs of accurate segmentation of brain edema lesions, which effectively improves the performance of medical image analysis through multi-dimensional technological innovation. First, a triple optimization strategy of Min-ASPP module, GSConv module and AG-attention mechanism is innovatively integrated in the model architecture: the Min-ASPP module enhances the feature capture ability of lesions of different sizes while reducing the number of parameters through the cascade design of multi-scale cavity convolution; the GSConv module employs a hybrid structure of depth-separable and standard convolution to ensure the feature expression ability; and the GSConv module adopts a hybrid structure of depth-separable and standard convolution to ensure the feature expression ability of lesions while ensuring the feature expression ability of lesions. which significantly reduces the computational complexity while ensuring the feature expression capability; and the AG attention mechanism strengthens the focusing capability on the key regions of lesions by establishing the joint channel-space attention weights across dimensions. Pre-processing means such as data normalization, random rotation and elastic deformation are used during the experiment to effectively expand the diversity of training samples. Comparison tests on the clinical brain edema MRI dataset show that the present model achieves a breakthrough performance of 94.32% for the Dice coefficient and 91.93% for the IoU metric, and its segmentation results demonstrate significant advantages in terms of lesion edge clarity and small lesion detection rate. The innovative results of this study not only provide a reliable technical means for the quantitative assessment of brain edema, but also the proposed modularization improvement idea provides a relocatable solution for other medical image segmentation tasks, which is of great value in promoting the clinical application of intelligent assisted diagnosis systems. This technological breakthrough marks a key step in the clinical applicability and algorithmic efficiency of medical image analysis based on deep learning, and lays a solid foundation for building an intelligent diagnosis and treatment system in the era of precision medicine.

DOI: 10.54254/2755-2721/2025.25644

References

[1] Chao, Pei-Ju, et al. "Using deep learning models to analyze the cerebral edema complication caused by radiotherapy in patients with intracranial tumor." Scientific Reports 12.1 (2022): 1555.
[2] Cui, Junzhao, et al. "Automatic segmentation of hemispheric CSF on MRI using deep learning: Quantifying cerebral edema following large hemispheric infarction." Heliyon 10.5 (2024).
[3] de Almeida, Pedro Aurélio Coelho, and Díbio Leandro Borges. "A deep unsupervised saliency model for lung segmentation in chest X-ray images." Biomedical Signal Processing and Control 86 (2023): 105334.
[4] Arvind, S., et al. "Improvised light weight deep CNN based U-Net for the semantic segmentation of lungs from chest X-rays." Results in Engineering 17 (2023): 100929.
[5] Kalam, Rehna, Ciza Thomas, and M. Abdul Rahiman. "Brain tumor detection in MRI images using adaptive-ANFIS classifier with segmentation of tumor and edema." Soft Computing 27.5 (2023): 2279-2297.
[6] Ghimire, Samip, and Santosh Subedi. "Estimating Lung Volume Capacity from X-ray Images Using Deep Learning." Quantum Beam Science 8.2 (2024): 11.
[7] Gaggion, Nicolás, et al. "CheXmask: a large-scale dataset of anatomical segmentation masks for multi-center chest x-ray images." Scientific Data 11.1 (2024): 511.
[8] Jiang, Liang, et al. "MRI radiomics features from infarction and cerebrospinal fluid for prediction of cerebral edema after acute ischemic stroke." Frontiers in Aging Neuroscience 14 (2022): 782036.
[9] Chen, Lingdong, et al. "Development of lung segmentation method in x-ray images of children based on TransResUNet." Frontiers in radiology 3 (2023): 1190745.
[10] Foroushani, Hossein Mohammadian, et al. "Accelerating prediction of malignant cerebral edema after ischemic stroke with automated image analysis and explainable neural networks." Neurocritical care 36.2 (2022): 471-482.