Atom Indonesia

Journal homepage: http://atomindonesia.brin.go.id

# Brain Tumor Segmentation in MR Images Using Swin Transformer

**A. Nur, K. Nurhanafi, E. R. Putri***

*Department of Physics, Faculty of Sciences and Mathematics, Mulawarman University, Samarinda 75123, Indonesia*

**ARTICLE INFO**

*Article history:*
Received 11 December 2024
Received in revised form 12 January 2025
Accepted 14 February 2025

*Keywords:*

Artificial intelligence
Brain tumor
Magnetic Resonance Imaging (MRI)
Segmentation
Swin transformer

**ABSTRACT**

Brain tumors are abnormal tissue growths in the brain. These brain tumors can have a negative impact on human health, one of which can interfere with brain functions such as vision, balance, and so on. Therefore, early detection needs to be done, one of which is by using medical imaging modalities, i.e., MRI. However, analyzing MRI scans requires careful observation and a high level of proficiency. Thus, medical image segmentation is required. Segmentation is important in medical image analysis as it allows medical experts to distinguish between abnormal and normal tissues. This study aims to determine the ability of the swin transformer architecture in segmenting brain tumor MR images. The image data used was BraTS 2021 data with a total of 1,250 images. The data were divided into three, i.e., training set, validation set, and testing set with a ratio of 70:15:15. Swin Transformer provided two main concepts, i.e., hierarchical feature maps and attention window shifts. The Swin Transformer initially was divided the image into small patches, which were then converted into vector form. After that, it was passed through W-MSA for local area and SW-MSA for cross window area. Next, multiple patches were merged into one, so that the image resolution gradually decreased, and then restored back to the original resolution. Based on this, the segmentation results were evaluated using a confusion matrix using DSC, IoU, and sensitivity metrics. The results of brain tumors MR image segmentation with Swin Transformer obtained evaluation values, i.e., 0.97313 for DSC, 0.94767 for IoU, and 0.96450 for sensitivity. It can be concluded that the Swin Tranformer can effectively segment brain tumor MR images.



## INTRODUCTION

Brain tumors are abnormal tissue growths in the brain that originate in the brain or meninges [1-7]. Brain tumors can negatively affect human health and disrupt brain functions, such as speech, vision, balance, motor, cognitive, and health behavior. Therefore, it is important to be aware of the symptoms of a brain tumor, such as headaches, nausea, vomiting, seizures, visual disturbances, and impaired balance [8-14]. The survival rate of people with brain tumors is relatively low, but it can improve greatly if the tumor is detected at an early stage. According to WHO data, about 700,000 people in the world are affected by brain tumors every year. In 2019, about 86,000 people were diagnosed with brain tumors, and 16,830 people died from brain tumors. The average survival rate for people with brain tumors is about 35 % [15].

Early detection of a brain tumor is very important because it can increase the chances of successful treatment [16-18]. Early detection can be done through physical examination and supporting examinations, such as Computed Tomography (CT), Positron Emission Tomography (PET), and Magnetic Resonance Imaging (MRI) [19-23]. With early detection, treatment can be done earlier and more effectively, so the chances of successful treatment will be higher [24]. MRI is the first choice in brain tumor imaging as it offers more detailed images and better contrast compared to other medical imaging techniques [25-31]. The technique has undergone significant advancements in the last two to three decades, making it a very important diagnostic tool in the field of oncology [32]. However, analyzing MRI scans requires careful observation and a high level of proficiency [33-36]. This is not possible for a layperson, so only a trained radiologist or radiation oncologist can do it properly

*Corresponding author.
E-mail address: erlinda.putri@fmipa.unmul.ac.id
DOI: https://doi.org/10.55981/aij.2025.1580

[34, 37-39]. The main task in brain tumor MR image analysis is to identify and delineate abnormal regions in the brain before image segmentation [40].

Segmentation is important in medical image analysis as it allows medical experts to distinguish between adjacent tissues in different parts of the body [41,42]. Image segmentation is an important step towards medical image analysis [43-45]. Image segmentation plays a major role in digital image processing and is used in various fields of science, including medical images, for object identification and classification [44].

Before the advent of Deep Learning (DL) methods, various classical approaches were developed and widely used for medical image segmentation purposes. Mathematical models and low-level image processing form the basis for many of these techniques. Some of the methods that have made significant contributions in this field include thresholding, region growing, graph truncation, Bayesian approaches, probabilistic clustering, and atlas-based methods [46-48]. Although these methods have been widely used, previous studies still face limitations in segmenting complex medical images. For example, thresholding methods rely on a fixed threshold value, which is less effective in handling high-intensity variations in biological tissues. Region growing methods often have difficulty in dealing with vague object boundaries and are prone to segment leakage. Graph cut-based approaches can provide good segmentation results but require optimal parameter selection and can be computationally expensive. Similarly, Bayesian approaches and probabilistic clustering require data distribution assumptions that often do not match real conditions, resulting in less accurate segmentation. Atlas-based methods also have limitations in handling high inter-subject variation and often require complex and computationally expensive registration. In addition to these technical challenges, many of these classical methods are semi-automated and rely heavily on prior knowledge and manual intervention to achieve optimal segmentation results. This is a major obstacle in clinical applications that require high efficiency and consistent accuracy in medical image segmentation.

As an alternative to conventional methods, DL-based approaches have emerged as a more efficient and accurate solution. In contrast to classical methods that often require manual feature extraction and complex parameter adjustments, DL allows models to learn patterns and features directly from the data, reducing reliance on rule-based image processing techniques. In addition, DL models have the advantage of handling high inter-subject shape variation as well as tissue texture complexity better than conventional approaches. Advances in DL have brought significant changes in medical image segmentation approaches [20,46,49-51]. DL models, particularly Convolutional Neural Networks (CNN), have shown impressive performance in efficiently extracting complex features from image data [46,52-59]. The transition from systems that use handcrafted features of Machine Learning (ML) to systems that learn features directly from data (DL) has resulted in improved accuracy and efficiency in segmentation tasks [46]. However, although CNN has been the dominant model, recent developments in Artificial Intelligence (AI) architectures have introduced promising new approaches, such as Transformers and their variations [46,60].

A further development of the Transformer is the Swin Transformer. Swin Transformer is a very interesting AI architecture in the field of computer vision. This model was developed by Microsoft Research in 2021 [61]. Swin Transformer is built using the Transformer concept as its base. However, it introduces two key concepts, i.e., hierarchical feature maps and attention window shifting [61,62]. These two concepts enable Swin Transformer to handle large-scale image data efficiently, making it a great tool for complex computer vision tasks. The hierarchical feature map in Swin Transformer helps to effectively represent different levels of features in an image. This leads to a more comprehensive understanding of the context and improved comprehension of the input data. Meanwhile, the window attention shifting mechanism expands the interaction field of each block. Thus, this architecture can capture features with varying scales more effectively [61].

Medical image segmentation, especially in brain tumors, is crucial in diagnosis and treatment planning. MR images of brain tumors face challenges related to contrast, texture, and boundaries between healthy and unhealthy tissues. Therefore, an accurate and efficient segmentation method is necessary. Several recent studies have used various segmentation methods. Wang et al. [63] used Transformer to segment 2019 and 2020 BraTS data, with Dice Similarity Coefficient (DSC) results of 0.9000 and 0.9009. Aboussaleh et al. [64] used CNN for brain tumor segmentation, with a DSC result of 0.8235. Kumar [65] compared several methods, including CNN, DWT, K-Means Clustering, Level Set Method, Watershed Algorithm, and Otsu Thresholding, with the best results in CNN (accuracy 0.9139, recall 0.8695, precision 0.9523, and F-measure 0.9090). Hao et al. [66] performed Multi-Scale CNN (MSCNN), with an

accuracy of 0.8720 on a dataset of 100 and 0.9130 on a dataset of 400.

Based on the literature review, various methods have been used for brain tumor segmentation in MR images, including Transformer, CNN, ResNet, DenseUNet+, and GAN. However, these results are often limited by model complexity and incomplete preprocessing. Research by Wang et al. [63] showed the potential of the Transformer method in brain tumor segmentation, which prompted the authors to explore the Swin Transformer as a recent development of the Transformer. The Swin Transformer has the potential to overcome the limitations of previous models with an adaptive approach to variations in the scale and structure of medical images. This research introduces the Swin Transformer architecture with a more comprehensive preprocessing approach, which includes resizing, filtering, normalization, and augmentation. This approach is expected to provide superior results compared to previous methods, as well as offer a more efficient and accurate solution in brain tumor segmentation in MR images.

The purpose of this research is to evaluate the performance of the Swin Transformer in brain tumor MR image segmentation using the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and sensitivity metrics. The focus is on measuring the effectiveness of the model in distinguishing tumors from normal brain tissue, with the hope of improving segmentation accuracy and efficiency.

## METHODOLOGY

In this research, the tools and materials used are a laptop equipped with Python software (spyder) to process data and the BraTS 2021 (Multimodal Brain Tumor Segmentation Challenge 2021) brain tumor MR image dataset [67-69] obtained from www.kaggle.com. The dataset consists of 1,250 MRI image data, each of which is original brain tumor image data along with annotations or tumor masks. The BraTS 2021 data is available in NIfTI (.nii) format. All imaging datasets have been manually annotated by one to four raters following the same annotation protocol. These annotations were verified by experienced neuro-radiologists.

Data collection is the initial stage in this research. The dataset collected is an MR image of a brain tumor with a data mask. The dataset is divided into three sets, i.e., the training set (70 %), validation set (15 %), and testing set (15 %).

Before being used to train the model, brain tumor MR image data needs to be processed through preprocessing. Some of the preprocessing steps are resizing. This ensures that all images have the same dimensions, making processing easier. Next, image pixel value normalization is performed. Normalization is done to homogenize the range of pixel values from different images. Next, filtering is performed to remove noise and improve image quality. Data augmentation is then performed to increase the diversity of the dataset and prevent overfitting. Data augmentation is done with various techniques, i.e., elastic transform, horizontal flip, vertical flip, and both horizontal and vertical flip. Figure 1 shows an example of augmented data.

The brain tumor segmentation process with Swin Transformer architecture starts by dividing the MRI image into small parts called patches through patch partition. A patch partition will break down a large image into small pieces for easier processing.

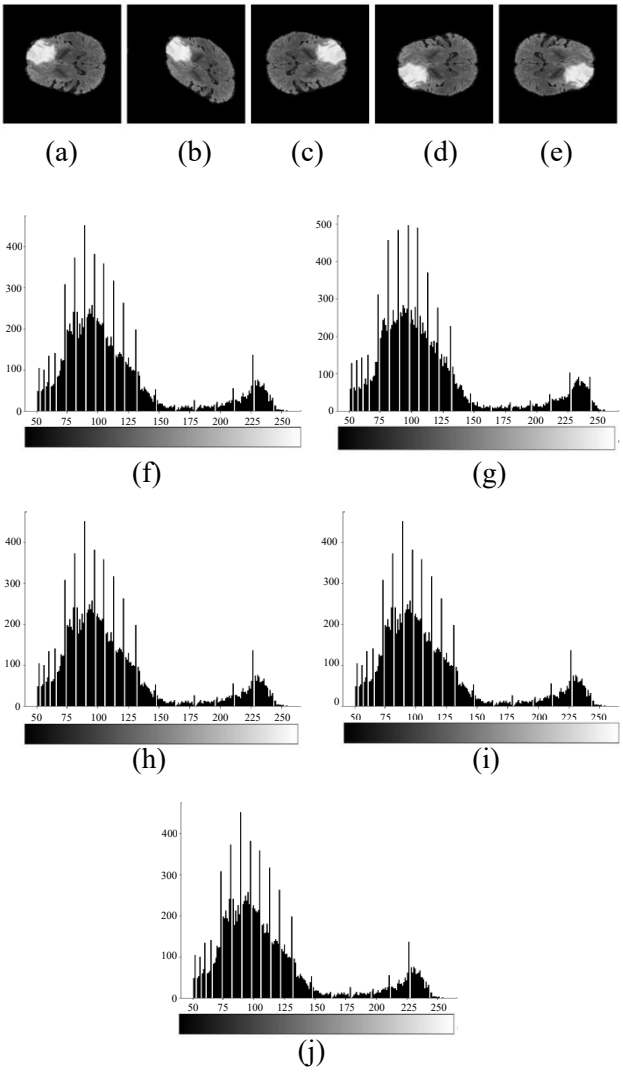


**Fig. 1.** Augmented images (a) original image, (b) elastic transform, (c) horizontal flip, (d) vertical flip, (e) both horizontal and vertical flip, (f) histogram of image a, (g) histogram of image b, (h) histogram of image c, (i) histogram of image d, and (j) histogram of image e.

After that, this collection of small patches is converted into vector form through a process called linear embedding so that each patch in the image has a more structured representation to be processed. Then, these image patches go through several stages of the Swin Transformer block, which has two main components, i.e., Window-based Multi-head Self-Attention (W-MSA) and Shifted Window-based Multi-head Self-Attention (SW-MSA). An illustration of the W-MSA process can be seen in Fig. 2.

In the first Fig. 2, it can be seen that self-attention is performed in small windows arranged in a grid. Each green square represents a part of the image that is input into the attention window. In the W-MSA stage, each window of a certain size processes the local information in it without regard to other windows. In other words, W-MSA groups these green squares into small areas where attention is only applied within that window. This allows effective processing of local information, especially for details within each window. In the illustration, the red box inside the window highlights the part that gets attention in this process. This W-MSA process can be written in the following Eqs. (1-2) [61,70,71].

$$\Omega(MSA) = 4hwC^2 + 2(hw)^2C \quad (1)$$

$$\Omega(W-MSA) = 4hwC^2 + 2M^2hwC \quad (2)$$

After W-MSA is applied, SW-MSA is tasked with expanding the range of attention by cyclically shifting the window, which can be seen in Fig. 3.

In SW-MSA, the attention window is shifted so that the new window position covers a different area compared to W-MSA. This shift allows the model to see information that was missed or hidden between the previous windows, extending the attention coverage to the entire image. This process combines information from multiple windows; and the range of attention is not limited to a single window. This cyclic shifting of windows gives the model the ability to integrate global information from different areas of the image. When the window is moved, the Swin Transformer can be written as the following Eqs. (3-6) [61,71-76].

$$\hat{z}^l = W-MSA\left(LN\left(z^{l-1}\right)\right) + z^{l-1} \quad (3)$$

$$z^l = MLP\left(LN\left(\hat{z}^l\right)\right) + \hat{z}^l \quad (4)$$

$$\hat{z}^{l+1} = SW-MSA\left(LN\left(z^l\right)\right) + z^l \quad (5)$$

$$z^{l+1} = MLP\left(LN\left(\hat{z}^l + 1\right)\right) + \hat{z}^l + 1 \quad (6)$$

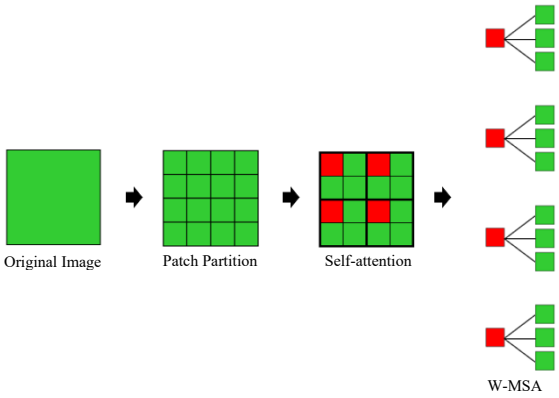


**Fig. 2.** Illustration of the W-MSA process.

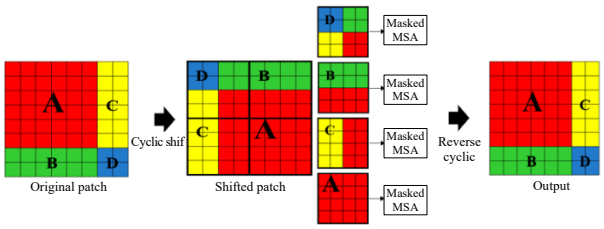


**Fig. 3.** Illustration of the SW-MSA process.

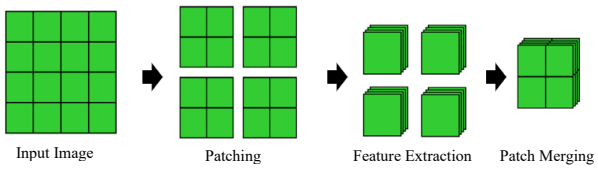


**Fig. 4.** Illustration of the patch merging process.

After each of these stages, patch merging is performed, which can be seen in Fig. 4.

In this section, multiple patches are merged into one so that the resolution of the image gradually decreases, but the information or features in it become more in-depth. This process continues until the image reaches the most compact stage in the bottleneck section, which serves as the connecting point between the encoder and decoder sections. Next, the image begins to be processed in the decoder section to restore the resolution to its original form. This process begins with patch expanding, which gradually increases the size of the image while still retaining, and the information from the encoder section connected through skip connections. These skip connections act as links that carry detailed information from the encoder to the decoder to ensure that each stage of image restoration retains important details. At the end of the process, the image goes through final patch expanding, which ensures the image resolution returns to full size so that each pixel has the correct class prediction. Figure 5 depicts the complete structure of the segmentation process.

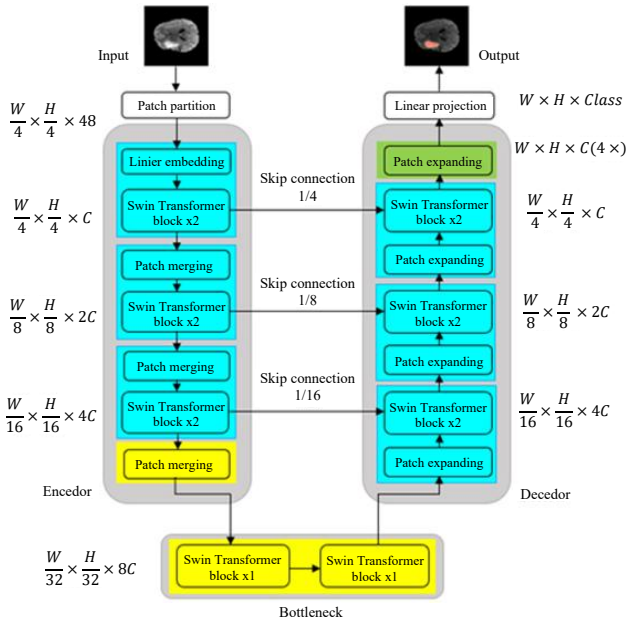


**Fig. 5.** Swin Transformer segmentation architecture.

The model was then trained using the Binary Cross Entropy with Logits Loss function (BCEWithLogitsLoss). A loss function was used to calculate how well the model makes predictions and was also used by the AdamW optimizer. During training, the model was evaluated on the validation set, and this process was repeated for several epochs until the validation loss no longer improved (early stopping). After training was completed, the model was tested on the testing set. The segmentation results were evaluated using the confusion matrix.

Segmentation results were evaluated for performance on the entire dataset using DSC [77-84], IoU [78-82,84-86], and sensitivity metrics [78,80-88] following Eqs. (7-9).

$$DSC = \frac{2\,TP}{2\,TP+FP+FN} \tag{7}$$

$$IoU = \frac{TP}{TP+FP+FN} \tag{8}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{9}$$

where True Positive (TP) is when the model detects the presence of a tumor, and there is in fact a tumor, meaning the model performed the detection correctly. True Negative (TN) occurs when the model detects no tumor, and in fact there is none, indicating that the model did not detect something that does not exist. False Positive (FP) is an error when the model detects a tumor when it does not exist. Conversely, a False Negative (FN) is when the model states there is no tumor when in fact there is one, which is a serious error because the tumor is not detected.

## RESULTS AND DISCUSSION

This study uses a dataset of brain tumor MR images from BraTS 2021 with a total of 1,250 data. Data testing is done by dividing the image into several sets, i.e., training set, validation set, and testing set, with a ratio of 70:15:15. The division is done by dividing into two sets first, i.e., the training set 70 % and testing set 30 %. Then, the testing set is divided into two for the validation set and the testing set. Thus, the portion for the testing set and validation set is 15 % each. Studies by Mahyoub et al. [89] and Febrianto et al. [90] also used a ratio of 70:15:15, and the results provided good accuracy in the case of brain tumors. According to Naceur et al. [91], the distribution of datasets with 70 % training set and 30 % testing set is the most optimal distribution.

In this research, before segmenting the brain tumor MR image, there is a preprocessing process. This stage consists of resizing, filtering, normalization, and augmentation. In this study, the MRI image was resized to a size of 224 × 224 pixels. This size was chosen because it is often used in medical image segmentation research as a standard input size. The resizing size used in this study uses references from the research of Alquran et al. [92] and Bianconi et al. [93]. This resizing aims to make all images used have the same dimensions as the tumor image but not so large that it burdens computational resources. By using this size, the model can be trained faster without sacrificing accuracy, maintaining a balance between image detail and training efficiency. This could potentially affect the segmentation results by speeding up the process without compromising the quality of the prediction. Furthermore, the filtering stage uses a median filter type. This filter is used to reduce noise in the image. According to Sheela and Suganthi [94], MR images usually consist of distortions and artifacts. The median filter is the most commonly used filter to filter out distortions and artifacts without losing important information in the image, such as image edges. According to Islam et al. [95], the median filter works well in protecting the smoothness of the image. Next, the normalization stage is performed to change the pixel values in the image to be in the range of 0 to 1. In this research, normalization is done by dividing each image pixel value by the highest (maximum) pixel value in the image. The image with the highest pixel value will

be 1, and the other pixels will be values between 0 and 1. In the last preprocessing stage, i.e., augmentation, this stage serves to increase the number of images without changing important information in each image. This process is used to prevent overfitting problems in the model. There are four augmentation techniques used in this research, i.e., elastic transform, horizontal flip, vertical flip, and both horizontal and vertical flip. This augmentation process was carried out on the training set so that initially, the data amounted to 875 images (70 % of the total image), so that in this process, the image data was multiplied 5 times from the initial amount, which resulted in 4,375 images. This happens because each original image undergoes augmentation from the four techniques used. Thus, the number of images generated is 875 original images plus 4 times the number of original images, which is 4,375 images in total.

This research uses the Swin Transformer architecture to segment brain tumors in MR images. In performing segmentation, there are several processes, which have previously been preprocessed on the image used. After that, of course, segmentation is done with the model itself. In this research, there are parameters used in the Swin Transformer architecture. The type of optimizer used in this research was AdamW. This selection is based on the slow convergence in training the Swin Transformer architecture, so the AdamW optimizer can accelerate convergence and reduce calculation losses during the training process [62]. The batch size used in this study is 32; this is a hyperparameter used to determine the number of images processed in one training model. Figure 6 demonstrates the segmentation result of the Swin Transformer.

This segmented image demonstrates the ability to detect and distinguish the tumor area from healthy brain tissue. The red color marks the tumor area, while the green line indicates the boundary between the tumor and the healthy tissue, ensuring that the detected area remains clearly separated.

The model is able to capture precise details of the tumor, especially in high-density sections. The accuracy in detecting complex tumor structures allows for sharper segmentation that is more faithful to the original shape. In addition, the model can adapt detection to variations in tumor size and shape, ensuring that not only the core is detected but also the surrounding areas that have similar characteristics.

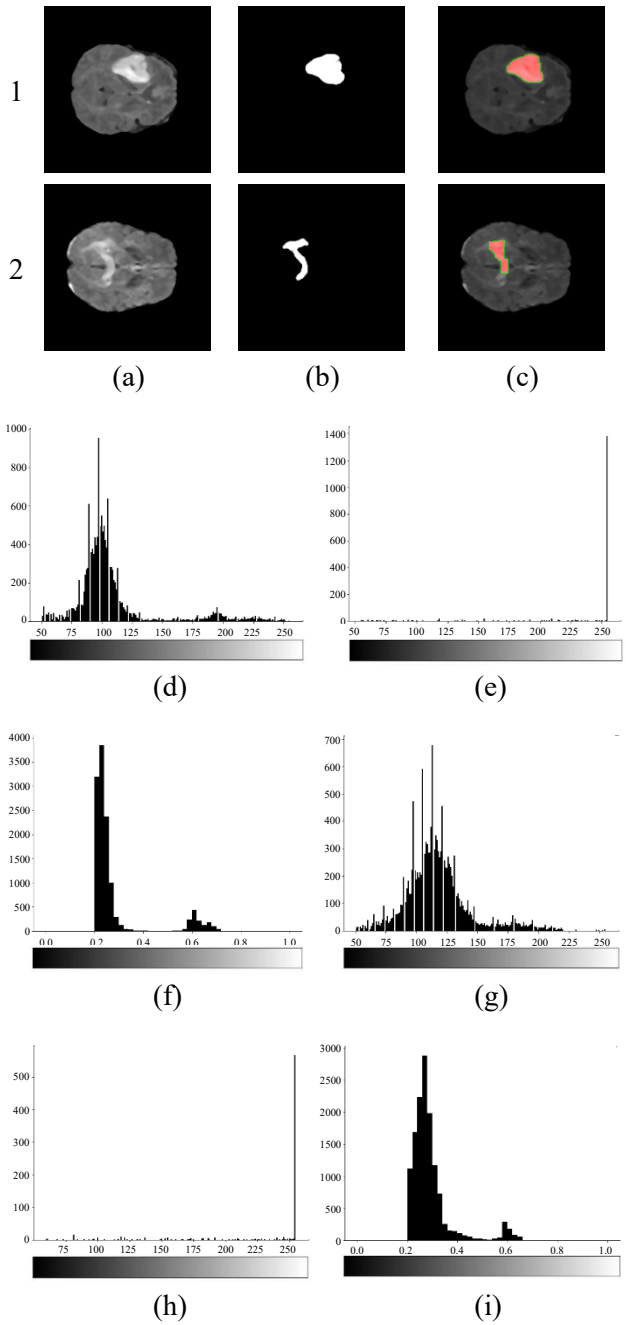


**Fig. 6.** Segmentation results of Swin Transformer
(a) original images, (b) mask, (c) segmentation result, (d) histogram of image 1a, (e) histogram of image 1b, (f) histogram of image 1c, (g) histogram of image 2a, (h) histogram of image 2b, and (i) histogram of image 2c.

Another noticeable advantage is the smooth and organized segmentation, where the boundary of the tumor can be well-identified without appearing blurred or merging with the surrounding tissue. This shows that the model has strong adaptability to variations in intensity and contrast in the image, providing results that are closer to reality.

The difference between the histogram of the original image and the segmented image shows that preprocessing processes such as resizing, filtering, normalization, and augmentation play a role in

filtering out irrelevant information and clarifying the structure of the tumor to be analyzed. The histogram of the original image shows a wider distribution of pixel intensities, while the histogram of the segmented image is more focused on a specific range, indicating the success of the segmentation in highlighting the target area.

Although some areas appear to be less optimally segmented, especially those with low intensity or fainter edges, the model still demonstrates the ability to detect the main part of the tumor. With precise segmentation results and clearer boundaries, this model has great potential in supporting more effective medical analysis and aiding in more accurate clinical decision-making.

The segmentation process was performed on all images in each dataset. All these segmentation results were evaluated with three metrics. The total number of images evaluated by the model was 4,750, which were then divided into three sets. The training set is 4,375 images, the validation set is 187 images, and the testing set is 188 images. Table 1 provides the acquisition of the confusion matrix.

Using Eqs. (7-9), the model performance obtained on the training set shows the following metrics: DSC with a value of 0.98579, IoU with a value of 0.97198, and sensitivity of 0.98618. In the validation set, DSC is obtained with a value of 0.97885, IoU with a value of 0.95858, and sensitivity of 0.97006. In the testing set, DSC is obtained with a value of 0.97313, IoU with a value of 0.94767, and a sensitivity of 0.96450.

From Table 1, it can be seen that the model has a high number of TPs, which indicates its ability to recognize the target area well in all sets. However, there are some FPs and FNs that need to be further analyzed to understand the potential causes of errors in segmentation. The higher FP values in the training set compared to the validation and testing sets suggest that the model may have experienced overfitting, where some non-tumor areas were classified as tumors. Meanwhile, the lower FP in the validation and testing sets indicates that the model has good generalization ability but still needs further improvement. One strategy to reduce FP is to apply more diverse augmentation data. In addition, the presence of FNs in all sets indicates that there are still areas of the tumor that are not detected by the model, which may result in a decrease in sensitivity. The persistence of FNs, although small in number, indicates that the model may still have difficulty in detecting some tumor areas with low contrast or irregular shape. To overcome this, it is possible to increase the input resolution to improve the sensitivity of the model to areas that are more difficult to recognize.

In the process of obtaining the performance evaluation results, there is certainly a process from the beginning of image set sharing to obtaining these values. During the training process, the model is trained to make increasingly accurate predictions for the segmentation results of the image data. Each time the model goes through the entire data set in the training process, this is referred to as an epoch. In this research, the maximum epoch is set as 100, the minimum is 25, and the early stopping is 5. This means that the model will be trained up to a maximum of 100 epochs, where every one epoch signifies that the model goes through the entire training image data once. However, training does not stop if it has not reached the minimum 25 epochs, even if early stopping has been met. Early stopping is a technique used to stop training early if the performance of the model has not improved. In this case, if after 5 consecutive epochs there is no improvement in performance, the training is automatically stopped, which avoids overfitting and speeds up the training process.

During the training process, the model stopped at the 31st epoch, indicating that early stopping was enabled. Although the maximum number of epochs had been set to 100, the model automatically stopped after 31 epochs because its performance no longer improved over the last 5 consecutive epochs, which corresponds to the early stopping setting. This indicates that at that point, the model had reached its optimal performance, and continuing the training longer would not result in further improvement. During the 31 epochs, it took 1574.388 minutes, which is equivalent to 26.2398 hours. Figure 7 illustrates the loss that occurred during the model training process.

**Table 1.** Confusion Matrix Acquisition of Segmentation Results.

| Confusion Matrix | TP | FP | FN | TN |
|---|---|---|---|---|
| Training Set | 3712 | 55 | 52 | 556 |
| Validation Set | 162 | 2 | 5 | 18 |
| Testing Set | 163 | 3 | 6 | 16 |

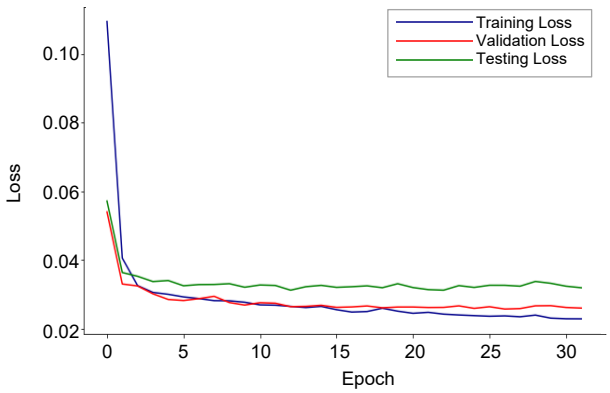


**Fig. 7.** Loss graph during training, validation, and testing of Swin Transformer.

In the graph, the x-axis is represented as the number of epochs, which is the number of times the entire dataset is processed during training, and the y-axis is represented as the loss value, which measures how far away the predicted results of the model are from the actual target. In the graph, there are three curves consisting of training loss in blue, validation loss in red, and testing loss in green.

Based on the graph above, when viewed from epoch 0 to 15 for training loss, it can be seen that the loss value has decreased sharply, from above 0.10 to drop quickly at the first epoch around 0.04. This shows that the model is able to learn the image data pattern in a short time. Validation loss also decreased, although not as fast as compared to training loss. This difference occurs because training loss is calculated based on image data that is learned directly by the model, so the model quickly adjusts its weights to minimize errors. In contrast, validation loss is calculated on image data that is not used during training, which aims to measure the generalization of the model. Testing loss also decreased but was higher than the validation loss. This indicates that the model is not fully optimized for the untrained image data. Then, from epochs 15 to 31, the training loss graph continues to decrease slowly, while the validation loss and testing loss tend to flatten. This shows that the model learns well from the training data. Although there is a small fluctuation in testing loss, the model can still maintain its performance.

Based on the performance evaluation, the segmentation results are helpful in determining the location and boundary of the tumor well. Segmentation plays an important role in the early stages of diagnosis to support therapy planning, such as the selection of appropriate treatment methods and calculation of radiation dose in cancer therapy. However, while it provides valuable information in the medical process, segmentation does not replace the final decision by the doctor. Further evaluation by the doctor is still required to ensure that the entire tumor area is well segmented and supports the overall clinical decision. If any part of the tumor is not detected or too little tissue is identified, this may indicate the need for model refinement, especially in the face of variations in size or more complex tumor morphologies.

To understand the effectiveness of the Swin Transformer architecture, it is important to compare it with other methods that have been used in previous studies. Various approaches have been taken for image segmentation, especially for brain tumors. Table 2 provides a comparison between some of the methods used in previous studies.

Some of these studies discuss various approaches in brain tumor segmentation using different models, preprocessing techniques, and datasets. Some of the models used include Swin Transformer, Transformer, CNN, ResNets, DenseUNet+, and GAN, each of which has its advantages, but their performance is highly dependent on preprocessing techniques and model complexity. A frequently used dataset is BraTS, with preprocessing techniques such as resizing, filtering, normalization, and augmentation to improve data quality.

The comparison results show that the Swin Transformer in this study provides the best results, especially in capturing global patterns in images and linking information from distant parts. The hierarchical design and window shift operations make the Swin Transformer more computationally efficient, although it is able to achieve high accuracy and is very reliable in detecting tumors, even on complex areas, it still requires a large amount of computing power and memory. As described by Pacal [61], the Swin Transformer approach of processing small blocks in turn aims to improve computational efficiency compared to the traditional Transformer model. This approach prevents too much computational overhead and saves resources. However, despite its efficiency, the Swin Transformer still requires sufficient hardware to achieve optimal performance.

**Table 2.** Comparison of Previous Research with Swin Transformer Architecture in Brain Tumor Segmentation.

| Research | Dataset | Preprocessing | Model Used | Performance |
|---|---|---|---|---|
| Wang et al. [63] | BraTs 2019 and 2020 | normalization, augmentation | Transformer | DSC: 0.900 and 0.901 |
| Aboussaleh et al. [64] | BraTs 2017 | normalization, augmentation | CNN | DSC: 0.823 |
| Shehab et al. [96] | BraTs 2015 | bias correction, normalization | ResNets | DSC: 0.860 |
| Çetiner & Metlek [77] | BraTs 2021 and FeTS 2021 | normalization, filtering, augmentation | DenseUNet+ | DSC: 0.950 and DSC: 0.870 |
| Ali et al. [97] | BraTs 2021 | resizing, normalization | Generative Adversarial Network (GAN) | DSC: 0.940; Sensitivity: 0.920 |
| Hatamizadeh et al. [98] | BraTs 2021 | normalization, augmentation | Swin Transformer | DSC: 0.927 |
| Zongren et al. [71] | BraTs 2021 | resizing, filtering, contrast processing | Swin Transformer | DSC: 0.932 |
| Present study | BraTs 2021 | resizing, filtering, normalization, augmentation | Swin Transformer | DSC: 0.973; IoU: 0.947; Sensitivity: 0.964 |

Meanwhile, models like CNN and ResNets are lighter but have lower performance. CNNs, for example, rely on complicated preprocessing and masks provided by experts, while ResNets require large GPU memory and long training time. DenseUNet+ and GAN also perform reasonably well, with advantages in handling variations in tumor size and shape, but DenseUNet+ requires long training times and is difficult to adapt to new data sets, while GAN requires large data sets and sometimes unstable training.

Overall, the Swin Transformer in this study stands out as the best choice for applications that require high accuracy and reliability in tumor detection, including in complex areas. The hierarchical design and window-shifting mechanism improve computational efficiency over traditional Transformer models while reducing the processing load. However, despite being more resource-efficient, this model still requires sufficient hardware for optimal performance. Therefore, other alternatives can be considered if resource constraints are a major factor or customization of the dataset and training process is required.

The use of Swin Transformer in brain tumor segmentation was also found in Hatamizadeh et al. [98] report, which has some differences compared to this study, especially in preprocessing and optimization techniques. Hatamizadeh et al. [98] applied preprocessing in the form of normalization (zero mean and unit standard deviation) and data augmentation such as intensity shift, intensity scale, and random axis mirror reversal. Meanwhile, this study uses resizing (224×224), filtering (median filter), and normalization (range 0-1), which are simpler but effective in improving training efficiency. The augmentation applied in this study is also more varied, including elastic transform, horizontal flip, vertical flip, as well as a combination of both, which helps to improve the model's robustness to variations in tumor shape. In terms of optimization, Hatamizadeh et al. [98] did not mention the optimizer used, while this study uses the AdamW optimizer, which is superior in handling the learning rate adaptively and reducing overfitting through the weight decay mechanism. In theory, a larger dataset generally gives better segmentation results, but this study with 1,250 images was able to produce a DSC of 0.973, which is higher than that of Hatamizadeh et al. [98], who used 1,470 images and obtained a DSC of 0.927. This shows that the right combination of preprocessing and optimization can improve model performance despite the smaller amount of data used.

Differences were also found in the study of Zongren et al. [71], especially in preprocessing techniques and the use of augmentation. Zongren et al. [71] applied resizing (128 × 128), filtering (Gaussian denoising), and contrast processing but did not include the use of data augmentation. In contrast to this study, which uses resizing to 224 × 224, allowing the model to capture more details in the image. The filtering used in this study is also superior, with the application of a median filter, which is more effective in removing impulsive noise without blurring the image edges. In addition, this study also applied data augmentation, which enriched the dataset and improved the model's robustness to variations in tumor structure, in accordance with the principle of data augmentation needed to improve the generalization ability of the model [99]. In terms of optimization, Zongren et al. [71] used the Adam optimizer, while this study used the AdamW optimizer, which is better at handling the learning rate adaptively and reducing overfitting through weight decay. Although the dataset used in this study is smaller (1,250 images compared to 2,000 images in Zongren et al. [71]), this study still produces higher segmentation performance (DSC 0.973 compared to 0.932), proving that better preprocessing and optimization strategies can outperform the advantage of data amount in improving segmentation accuracy.

## CONCLUSION

This study shows that the Swin Transformer architecture can effectively segment brain tumor MRI images based on the results of performance evaluation using DSC, IoU, and sensitivity metrics. These results indicate that the Swin Transformer has great potential in improving segmentation accuracy, which can support the diagnosis and treatment planning process.

However, this study still has some limitations. While the Swin Transformer offers high accuracy, it has greater computational complexity than conventional methods, requiring hardware with high processing power. This can be an obstacle in clinical implementation, especially in healthcare facilities with limited resources. In addition, the training and inference process of the Swin Transformer takes longer than some other lighter architectures, which may affect efficiency in real-time applications. Therefore, model optimization to reduce computational load and speed up inference is an important step in future research.

Future research can focus on improving the efficiency of the Swin Transformer without sacrificing accuracy, for example through model quantization or parameter compression techniques. In addition, further development can be done by

integrating this model into a clinical decision support system that can assist doctors in diagnosing and treating brain tumors more quickly and accurately. The application of this method to broader clinical scenarios also needs to be explored to ensure its effectiveness in real conditions, so that it can contribute more to the medical world.

## ACKNOWLEDGMENT

We are deeply grateful to the Department of Physics, Faculty of Science and Mathematics, Mulawarman University, Samarinda, Indonesia, for their unwavering support and resources, which have been fundamental to the success of this research endeavor.

## AUTHOR CONTRIBUTION

A. Nur contributed to preparation and acquisition data. A. Nur and E. R. Putri contributed to analysis and interpretation data. E. R. Putri contributed to the conception and design of the work. All authors were involved in drafting and commenting on the paper.

## REFERENCES

1. E. Ghafourian, F. Samadifam, H. Fadavian *et al*., Diagn. **13** (2023) 561.
2. H. A. Shah, F. Saeed, S. Yun *et al*., IEEE Access **10** (2022) 65426.
3. C. Srinivas, N.P. Nandini, M. Zakariah *et al*., J. Healthcare Eng. **8** (2022) 1.
4. R. Vankdothu and M. A. Hameed, Meas: Sens. **24** (2022) 100440.
5. A. Younis, L. Qiang, C. O. Nyatega *et al*., Appl. Sci. **12** (2022) 7282.
6. W. Zafar, G. Husnain, A. Iqbal *et al*., Results Eng. **24** (2024) 102994.
7. C. J. Tseng and C. Tang, Healthcare Anal. **4** (2023) 100217.
8. N. Ullah, J. A. Khan, M. S. Khan *et al*., Appl. Sci. **12** (2022) 5645.
9. S. Maqsood, R. Damaševičius and R. Maskeliūnas, Med. **58** (2022) 1090.
10. J. Walsh, A. Othmani, M. Jain *et al*., Healthcare Anal. **2** (2022) 100098.
11. M. Agarwal, G. Rani, A. Kumar *et al*., Results Eng. **22** (2024) 102117.
12. R. Zhou, J. Wang, G. Xia *et al*., Entropy **26** (2024) 385.
13. F. Ghandour, A. Squassina, R. Karaky *et al*., Brain Sci. **11** (2021) 1.
14. A. Zumel-Marne, M. Kundi, G. Castaño-Vinyals *et al*., J. Neuro-Oncol. **147** (2020) 427.
15. U. Zahid, I. Ashraf, M. A. Khan *et al*., Comput. Intell. Neurosci. **2022** (2022) 1465173.
16. S. Saeedi, S. Rezayi, H. Keshavarz *et al*., BMC Med. Inf. Decis. Making **23** (2023) 1.
17. A. Aleid, K. Alhussaini, R. Alanazi *et al*., Appl. Sci. **13** (2023) 3808.
18. A. Akter, N. Nosheen, S. Ahmed *et al*., Expert Syst. Appl. **238** (2024) 122347.
19. A. A. Asiri, A. Shaf, T. Ali *et al*., Diagn. **13** (2023) 2094
20. E. K Rutoh, Q. Zhi, N. Bahadar *et al.*, J. King Saud Univ. Comput Inf. Sci. **36** (2024) 102086.
21. Y. Yin, Z. Tang and H. Weng, Biomed. Eng. Online **23** (2024) 1.
22. S. Saladi, Y. Karuna, S. Koppu *et al*., Math. **11** (2023) 285.
23. P. Raut, G. Baldini, M. Schöneck *et al*., Front. Radiol. **3** (2021) 1.
24. H. Alsaif, R. Guesmi, B. M. Alshammari *et al*., Appl. Sci. **12** (2022) 3773.
25. S. Ahmad and P. K. Choudhury, IEEE Access. **10** (2022) 59099.
26. A. Alyami, N. Majrashi, L. Hazazi *et al*., J. Radiat. Res. Appl. Sci. **17** (2024) 100801.
27. T. C. Arnold, C. W. Freeman, B. Litt *et al*., J. Magn. Reson. Imaging **57** (2023) 25.
28. M. S. I. Khan, A. Rahman, T. Debnath *et al*., Comput. Struct. Biotechnol. J. **20** (2022) 4733.
29. T. B. Nguyen-Tat, T. Q. T. Nguyen, H. N. Nguyen *et al.*, Egypt Inf. J. **27** (2024) 1.
30. S. Rajput, R. Kapdi, M. Roy *et al*., Healthcare Anal. **5** (2024) 100307.
31. S. Xun, Y. Zhang, S. Duan *et al*., VRIH **6** (2024) 203.
32. B. S. Alemu, S. Feisso, E. A. Mohammed *et al*., Sci. Afr. 22 (2023) e01963.
33. A. Al-Fakih, A. Shazly, A. Mohammed *et al*., Alexandria Eng. J. **99** (2024) 108.
34. M. F. Almufareh, M. Imran, A. Khan *et al*., IEEE Access **12** (2024) 16189.
35. Y. Cao and Y. Song. Appl Sci. 14 (2024) 4919.

36. Z. Xiao, Y. Zhang, Z. Deng Z *et al*., Neuroimage **292** (2024) 120608.
37. W. Abbaoui, S. Retal, S. Ziti *et al*., J. Clin. Med. **13** (2024) 2323.
38. Y. A. L. Khalil, A. Ayaz, C. Lorenz *et al*., Comput. Med. Imaging Graphics **112** (2024) 1.
39. M. Mir, Z. S. Madhi, A. H. A Hamid *et al*., Sci. Rep. **14** (2024) 23341.
40. M. K. H. Khan, W. Guo, J. Liu *et al*., Exp. Biol. Med. **248** (2023) 1974.
41. N. S. Punn and S. Agarwal, Artif. Intell. Rev. **55** (2022) 5845.
42. S. Jardim, J. António and C. Mora *et al*., Procedia Comput. Sci. **219** (2023) 1485.
43. M. Chi, H. An, X. Jin *et al*., Entropy **26** (2024) 166.
44. P. Couto, T. Bento, H. Bustince *et al*., Appl. Sci. **12** (2022) 4865.
45. A. Sulaiman, V. Anand, S. Gupta *et al*., Sci. Rep. **14** (2024) 1345.
46. P. H. Conze, G. Andrade-Miranda, V. K. Singh *et al*., IEEE Trans. Radiat. Plasma. Med. Sci. **7** (2023) 545.
47. D. Karimi, H. Dou and A. Gholipour, IEEE Access **10** (2022) 29322.
48. G. Dong, Z. Wang, Y. Chen *et al.*, Sci. Rep. **14** (2024) 19425.
49. J. Egger, C. Gsaxner, A. Pepe *et al.*, Comput. Methods Programs Biomed. **221** (2022) 106874.
50. M. A. Ottom, H. A. Rahman, I. D. Dinov, IEEE J. Transl. Eng. Health Med. **10** (2022) 1800508.
51. W. Nhlapho, M. Atemkeng, Y. Brima *et al*., Inf. **15** (2024) 182.
52. M. Arabahmadi, R. Farahbakhsh and J Rezazadeh, Sens. **22** (2022) 1960.
53. G. Anari, G. G. de Oliveira, R. Ranjbarzadeh *et al*., Bioeng. **11** (2024) 945.
54. C. Bian, C. Hu and N. Cao, Bioeng. **11** (2024) 958.
55. R. S. Choi, J. Lee, M. Lee *et al*., IEEE Access **12** (2024) 84122.
56. S. Luo, L. Pan, Y. Jian *et al*., Alexandria Eng. J. **88** (2024) 133.
57. S. Mohammadi and M. Allali, Appl. Sci. **14** (2024) 3424.
58. P. Priyadarshini, P. Kanungo and T. Kar, e-Prime Adv. Electr. Eng Electron. Energy. **8** (2024) 100498.
59. R. A. Zeineldin, M. E. Karar, Z. Elshaer *et al*., Sci. Rep. **14** (2024) 3713.
60. R. Azad, M. T. Al-Antary, M. Heidari *et al*., IEEE Access **10** (2022) 108205.
61. I. Pacal, Int. J. Mach. Learn. Cybern. **15** (2024) 3579.
62. D. Feng, Z. Zhang and K. Yan, IEEE Access **10** (2022) 77432.
63. W. Wang, C. Chen and M. Ding *et al*., Lect. Notes Comput. Sci. **12901** (2021) 109.
64. I. Aboussaleh, J. Riffi, A. M. Mahraz *et al*., J. Imaging **7** (2021) 269.
65. A. Kumar, Multimedia Tools Appl. **82** (2023) 7117.
66. J. Hao, X. Li and Y. Hou, IEEE Access **8** (2020) 65758.
67. U. Baid, S. Ghodasara, S. Mohan *et al.*, Comput. Sci. (2021) 1.
68. S. Bakas, H. Akbari, A. Sotiras *et al*., Sci Data **4** (2017) 170117.
69. B. H. Menze, B. Jakab, S. Bauer *et al*., IEEE Trans. Med. Imaging **34** (2015) 1993.
70. L. Gao, H. Liu, M. Yang *et al*., IEEE J. Sel. Top. Appl. Earth. Obs. Remote Sens. **14** (2021) 10990.
71. L. Zongren, W. Silamu, W. Yuzhen *et al*., IEEE Access **11** (2023) 42895.
72. L. Cui, X. Jing, Y. Wang *et al*., IEEE J. Sel. Top. Appl. Earth. Obs. Remote Sens. **16** (2022) 369.
73. Y. Gu, Z. Piao and S. J. Yoo, Appl. Sci. **12** (2022) 468.
74. Z. Liao, N. Fan and K. Xu, Appl. Sci. **12** (2022) 4735.
75. C. Wei, S. Ren, K. Guo *et al*., Sens. **23** (2023) 3420.
76. U. Zidan, M. M. Gaber and M. M. Abdelsamea, Expert Syst. Appl. **216** (2023) 119452.
77. H. Çetiner and S. Metlek, J. King Saud Univ. Comput Inf. Sci. **35** (2023) 101663.
78. Z. Li, H. Zhang, Z. Li *et al*., Appl. Sci. **12** (2022) 7149.
79. M. Mubashar, H. Ali, C. Grönlund *et al*., Neural Comput. Appl. **34** (2022) 17723.
80. D. Müller, I. Soto-Rey and F. Kramer, BMC Res. Notes **15** (2022) 210.

81. S. Pan, X. Liu, N. Xie *et al*., BMC Bioinf. **24** (2023) 85.

82. P. Shi, M. Duan, L. Yang *et al*., Mater. **15** (2022) 4417.

83. Z. Shi, Y. Li, H. Zou *et al*., Sens. **23** (2023) 4897.

84. Y. Xu, S. Hou, X. Wang *et al*., Diagn. **13** (2023) 576.

85. G. Alfonso-Francia, J. C. Pedraza-Ortega, M. Badillo-Fernández *et al*., Diagn. **12** (2022) 3031.

86. C. Sheng, L. Wang, Z. Huang *et al*., J. Syst. Sci. Complexity **36** (2023) 257.

87. M. M. Al Rahhal, Y. Bazi, R. M. Jomaa *et al.*, J. Pers. Med. **12** (2022) 310.

88. T. T. Nguyen and T. V. Nguyen, IEEE Access **11** (2023) 95346.

89. M. Mahyoub, F. Natalia, S. Sudirman *et al*., *Brain Tumor Segmentation in Fluid-Attenuated Inversion Recovery Brain MRI using Residual Network Deep Learning Architectures,* in: Proceedings of the 15th International Conference on Developments in eSystems Engineering (DeSE), IEEE, New Jersey (2023) 486.

90. D. C. Febrianto, I. Soesanti and H. A. Nugroho, *Convolutional Neural Network for Brain Tumor Detection,* In: IOP Conference Series: Materials Science and Engineering, IOP Publishing Bristol (2020) 012031.

91. M. B. Naceur, M. Akil, R. Saouli *et al*., Med. Image Anal. **63** (2020) 101692.

92. H. Alquran, M. Alslatie, A. Rababah *et al*., Appl. Sci. **14** (2024) 6504.

93. A. Bianconi, L. F. Rossi, M. Bonada *et al*., Brain Inf. **10** (2023) 1.

94. C. J. J. Sheela and G. Suganthi, J. King Saud Univ. Comput Inf. Sci. **34** (2022) 557.

95. M. K. Islam, M. S. Ali, M. S. Miah *et al*., Mach Learn. Appl. **5** (2021) 100044.

96. L. H. Shehab, O. M. Fahmy, S. M. Gasser *et al.*, J. King Saud Univ. Eng. Sci. **33** (2021) 404.

97. A. Ali, M. Sharif, M. S. Faisal *et al*., IEEE Access **12** (2024) 183525.

98. A. Hatamizadeh, V. Nath, Y. Tang *et al*., Lect. Notes Comput. Sci. **12962** (2022) 272.

99. H. B. Cokrokusumo, I. Hariyati, L. E. Lubis *et al*., Atom Indones. **48** (2022) 171.