Real-time Brain tumor segmentation on MRI images using yolov9

Pham Thanh Lam and Le Nguyen Hoang Lam[†]

[1*]University of Information Technology, Vietnam National University, Ho Chi Minh City,Vietnam.

Contributing authors: 21520055@gm.uit.edu.vn; 21522274@gm.uit.edu.vn;

[†]These authors contributed equally to this work.

Abstract

Brain tumor segmentation plays a crucial role in the diagnosis and treatment of brain tumors. In this paper, we propose a real-time brain tumor segmentation method using the YOLOv9 architecture on MRI images. The YOLOv9 model, originally designed for object detection, is adapted to perform pixel-level segmentation of brain tumors. We used preoperative imaging and genomic data of 110 patients from 5 institutions with lower-grade gliomas from The Cancer Genome Atlas. The proposed method utilizes the powerful feature extraction capabilities of the YOLOv9 network to accurately identify tumor regions in MRI scans. We train the model on a large dataset of annotated MRI images to learn the distinctive features of brain tumors. Experimental results on a diverse set of MRI datasets demonstrate the effectiveness and efficiency of our approach. The availability of real-time segmentation can aid radiologists and clinicians in making timely and accurate diagnoses, facilitating improved patient care and treatment planning.

Keywords: segmentation, yolo, brain tumor

1 Introduction

The Introduction section, of referenced text Deep learning-based models have demonstrated far better performance than past artificial intelligence systems in various fields, such as computer vision, language processing, and speech recognition. In recent years, researchers in the field of deep learning have mainly focused on how to develop more powerful system architectures and learning methods, or more general objective functions, such as loss function, label assignment and auxiliary supervision. These methods methods focus on how to design the most appropriate objective functions so that the prediction results of the model can be closest to the ground truth. Meanwhile, an appropriate architecture that can facilitate acquisition of enough information for prediction has to be designed. Existing methods ignore a fact that when input data undergoes layer-by-layer feature extraction and spatial transformation, large amount of information will be lost. Most past approaches have ignored that input data may have a non-negligible amount of information loss during the feedforward process. This loss of information can lead to biased gradient flows, which are subsequently used to update the model. The above problems can result in deep networks to establish incorrect associations between targets and inputs, causing the trained model to produce incorrect predictions.

The YOLOv9 architecture aim to address this issue. The concept programmable gradient information (PGI) was proposed to solve the information bottleneck problem and the problem that the deep supervision mechanism is not suitable for lightweight neural networks. GELAN, a highly efficient and lightweight neural network, was also introduced. In terms of object detection, GELANhasstrong and stable performance at different computational blocks and depth settings. It can indeed be widely expanded into a model suitable for various inference devices. The introduction of PGI allows both lightweight models and deep models to achieve significant improvements in accuracy. The YOLOv9, designed by combining PGI and GELAN, has shown strong competitiveness. Its excellent design allows the deep model to reduce the number of parameters by 49% and the amount of calculations by 43% compared with YOLOv8, but it still has a 0.6% AP improvement on MS COCO dataset.

YOLOv9 capabilities can be leveraged and applied in the detecting and segmenting brain tumors task. Lower-grade gliomas (LGG) are a group of WHO grade II and grade III brain tumors. As opposed to grade I which are often urable by surgical resection, grade II and III are in ltrative and tend to recur and evolve to higher-grade lesion. Imaging can provide important information before surgery or in cases when resection is not possible. The first step when extracting tumor features was the segmentation of MRI. Very recent studies in this area have discovered an association of tumor shape features extracted from MRI with its genomic subtypes [5, 6].

2 Related Works

2.1 The YOLO series

The most widely used real-time object detector at present is the YOLO series. The YOLO (You Only Look Once) series is a family of object detection models that have gained significant popularity and achieved state-of-the-art performance in computer vision tasks. YOLO revolutionized object detection by introducing a real-time, end-to-end approach that detects and localizes objects in images or videos in a single pass. YOLO models use a single neural network to simultaneously predict bounding boxes and class probabilities for multiple objects within an image. This makes them highly efficient and capable of processing images in real-time. YOLO models employ a grid-based approach to divide the input image into grid cells and predict bounding boxes and class probabilities within each cell.

The YOLO architecture incorporates convolutional neural networks (CNNs) for feature extraction and subsequent fully connected layers for object detection. It employs anchor boxes to handle objects of different sizes and aspect ratios and utilizes techniques like feature pyramid networks and skip connections to capture objects at different scales within the network. The YOLO series has demonstrated impressive performance on various object detection benchmarks, showcasing its effectiveness in tasks such as object localization, instance segmentation, and real-time video analysis. These models have found applications in diverse domains, including autonomous driving, surveillance systems, and medical imaging.

With each iteration, the YOLO series has pushed the boundaries of real-time object detection, constantly refining and improving the accuracy and speed of object localization. Researchers and developers continue to build upon the YOLO framework, exploring new innovations and adaptations to address the evolving challenges in computer vision applications.

YOLOv9, which is the model used in this paper, is a direct improvement from YOLOv7. GELAN was used to improve the architecture and the training process with the proposed PGI using YOLOv7 as base, the approach makes YOLOv9 the top real-time object detector of the new generation.

2.2 Association of genomic subtypes of lower-grade gliomas

Lower-grade gliomas (LGG) are a group of WHO grade II and grade III brain tumors. As opposed to grade I which are often curable by surgical resection, grade II and III are infiltrative and tend to recur and evolve to higher-grade lesion. Patients with tumors from different molecular groups substantially differ in terms of typical course of the disease and overall survival. Imaging can provide important information before surgery or in cases when resection is not possible. The first step when extracting tumor features was the segmentation of MRI. Very recent studies in this area have discovered an association of tumor shape features extracted from MRI with its genomic subtypes [] [5, 6].

Deep learning is a new field of machine learning that is recently revolutionizing the automated analysis of images. There are many examples of successful applications of deep learning in medical imaging and more specifically in brain MRI segmentation. In recent years, progress in deep learning for automatic brain segmentation matured to a level that achieves performance of a skilled radiologist [] [16]. Development of models that yield high quality segmentation of LGG in brain MRI would potentially allow for automatization of the process of tumor genomic subtype identification through imaging that is fast, inexpensive, and free of inter-reader variability. []Et combine the field of deep learning and radiogenomic and propose a fully automatic algorithm for quantification of tumor shape and test whether the assessed shape features are prognostic of tumor molecular subtypes, as well as a brain tumors segmentation dataset, which is the dataset used in this paper to evaluate the performance of the YOLOv9 architecture.

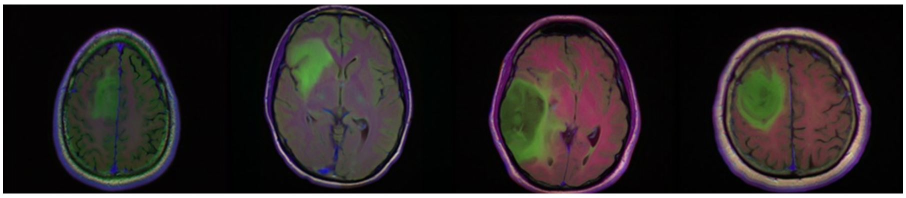


Fig. 1 MRI images from The Cancer Imaging Archive (TCIA)

3 Dataset

3.1 Patient population

The data used in this study was obtained from The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA). We identified 120 patients from TCGA lower-grade glioma collection who had preoperative imaging data available, containing at least a fluid-attenuated inversion recovery (FLAIR) sequence. Ten patients had to be excluded since they did not have genomic cluster information available. The final group of 110 patients was from the following 5 institutions: Thomas Jefferson University (TCGA-CS, 16 patients), Henry Ford Hospital (TCGA-DU, 45 patients), UNC (TCGA-EZ, 1 patient), Case Western (TCGA-FG, 14 patients), Case Western – St. Joseph's (TCGA-HT, 34 patients) from TCGA LGG collection. The entire images set of 110 patients was split into 2 non overlapping subsets of 100000000 images and 275 images. This was done for evaluation.

3.2 Imaging data

Imaging data was obtained from The Cancer Imaging Archive which contains the images corresponding to the TCGA patients and is sponsored by the National Cancer Institute. There were 101 patients with all sequences available, 9 patients with missing post-contrast sequence, and 6 with missing pre-contrast sequence. The number of slices varied among patients from 20 to 88. In order to capture the original pattern of tumor growth, we only analyzed preoperative data. The assessment of tumor shape was based on FLAIR abnormality since enhancing tumor in LGG is rare.

A medical school graduate with experience in neuroradiology imaging, manually annotated FLAIR images by drawing an outline of the FLAIR abnormality on each slice to form training data for the automatic segmentation algorithm. A board eligible radiologist verified all annotations and modified those that were identified as incorrect. Dataset of registered images together with manual segmentation masks for each case used in our study is released and made publicly available at the following link: https://kaggle.com/mateuszbuda/lgg-mri-segmentation.

4 Method

In deep networks, the phenomenon of input data losing information during the feed-forward process is commonly known as information bottleneck [59], and its schematic diagram is as shown in Figure 2.

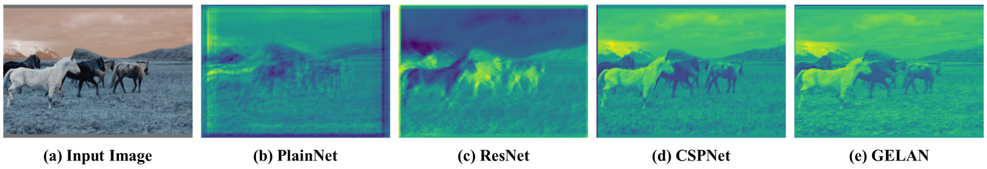
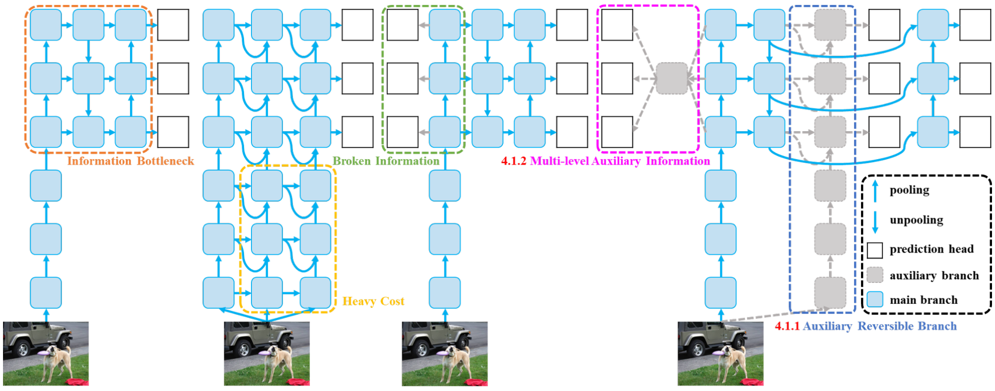


Fig. 2 Visualization results of random initial weight output feature maps for different network architectures: (a) input image, (b) PlainNet, (c) ResNet, (d) CSPNet, and (e) proposed GELAN. From the figure, we can see that in different architectures, the information provided to the objective function to calculate the loss is lost to varying degrees, and our architecture can retain the most complete information and provide the most reliable gradient information for calculating the objective function.

At present, the main methods that can alleviate this phenomenon are as follows: The use of reversible architectures : this method mainly uses repeated input data and maintains the information of the input data in an explicit way; The use of masked modeling: it mainly uses reconstruction loss and adopts an implicit way to maximize the extracted features and retain the input information; and Introduction of the deep supervision concept: it uses shallow features that have not lost too much important information to pre-establish a mapping from features to targets to ensure that important information can be transferred to deeper layers. However, the above methods have different drawbacks in the training process and inference process. For example, a reversible architecture requires additional layers to combine repeatedly fed input data, which will significantly increase the inference cost. In addition, since the input data layer to the output layer cannot have a too deep path, this limitation will make it difficult to model high-order semantic information during the training process. As for masked modeling, its reconstruction loss sometimes conflicts with the target loss. In addition, most mask mechanisms also produce incorrect associations with data. For the deep supervision mechanism, it will produce error accumulation, and if the shallow supervision loses information during the training process, the subsequent layers will not be able to retrieve the required information. The above phenomenon will be more significant on difficult tasks and small models.

To address the above-mentioned issues, YOLOv9 propose a new concept, which is programmable gradient information (PGI). The concept is to generate reliable gradients through auxiliary reversible branch, so that the deep features can still maintain key characteristics for executing target task. The design of auxiliary reversible branch can avoid the semantic loss that may be caused by a traditional deep supervision process that integrates multi-path features. In other words, YOLOv9 program gradient information propagation at different semantic levels, and thereby achieve the best training results. The reversible architecture of PGI is built on auxiliary branch, so there is no additional cost. Since PGI can freely select loss function suitable for the target task, it also overcomes the problems encountered by mask modeling. The proposed PGI mechanism can be applied to deep neural networks of various sizes and is more general than the deep supervision mechanism, which is only suitable for very deep neural networks.



(a) PAN [37] (b) RevCol [3] (c) Deep Supervision (d) Programmable Gradient Information

Fig. 3 PGI and related network architectures and methods. (a) Path Aggregation Network (PAN)) [], (b) Reversible Columns (RevCol) [], (c) conventional deep supervision, and (d) our proposed Programmable Gradient Information (PGI). PGI is mainly composed of three components: (1) main branch: architecture used for inference, (2) auxiliary reversible branch: generate reliable gradients to supply main branch for backward transmission, and (3) multi-level auxiliary information: control main branch learning plannable multi-level of semantic information

4.1 Auxiliary Reversible Branch

In YOLOv9's PGI, auxiliary reversible branch was proposed to generate reliable gradients and update network parameters. By providing information that maps from data to targets, the loss function can provide guidance and avoid the possibility of finding false correlations from incomplete feedforward features that are less relevant to the target. The maintenance of complete information by introducing reversible architecture, but adding main branch to re versible architecture will consume a lot of inference costs. Auxiliary reversible branch can be removed during the inference phase, the inference capabilities of the original network can be retained. Any reversible architectures in PGI can be chosen to play the role of auxiliary reversible branch.

4.2 Multi-level Auxiliary Information

For object detection, different feature pyramids can be used to perform different tasks, for example together they can detect objects of different sizes. The deep supervision architecture including multiple prediction branch is shown in Figure 3 (c). After connecting to the deep supervision branch, the shallow features will be guided to learn the features required for small object detection, and at this time the system will regard the positions of objects of other sizes as the background. However, the above deed will cause the deep feature pyramids to lose a lot of information needed to predict the target object. Therefore, each feature pyramid needs to receive information about all target objects so that subsequent main branch can retain complete information to learn predictions for various targets.

The concept of multi-level auxiliary information is to insert an integration network between the feature pyramid hierarchy layers of auxiliary supervision and the main branch, and then uses it to combine returned gradients from different prediction heads, as shown in Figure 3 (d). Multi-level auxiliary information is then to aggregate the gradient information containing all target objects, and pass it to the main branch and then update parameters. At this time, the characteristics of the main branch's feature pyramid hierarchy will not be dominated by some specific object's information. As a result, this method can alleviate the broken information problem in deep supervision. In addition, any integrated network can be used in multi-level auxiliary information. Therefore, the required semantic levels can be planned to guide the learning of network architectures of different sizes.

5 Result

The team performed experiments training the prepared data using the yolov9-seg model with the task of image segmentation. The training process is performed for 50 epochs with 2 different batches of 32 and 64. From the experiment, the team took the results of the training model with a batch of 32 to make comments and evaluations. The performance of the YOLOv9 model on the segmentation task was evaluated using a range of loss and performance metrics. To evaluate the accuracy of the model with the test data set, the team evaluated 4 indicators: Precision, Recall, mAP50, mAP50-95.

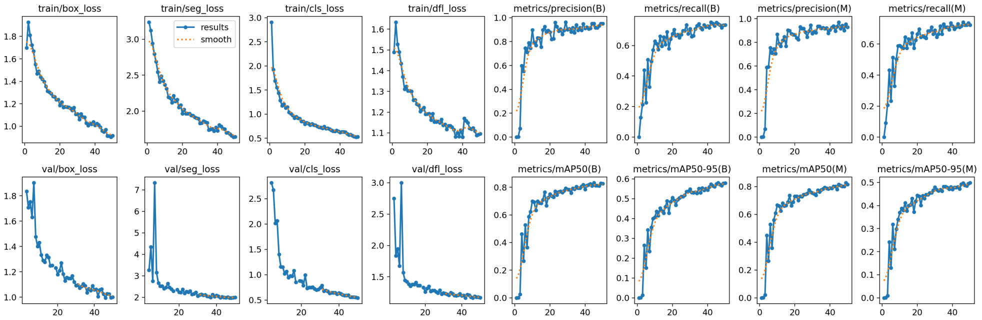
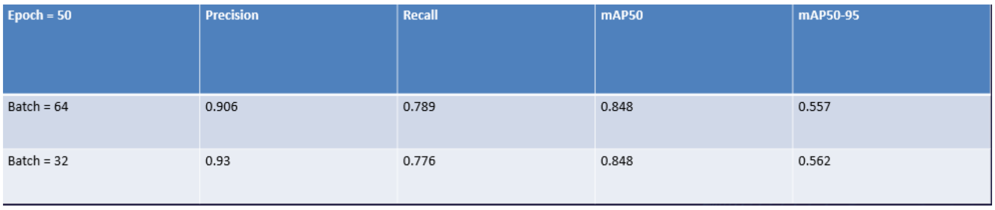


Fig. 4 The diagram shows the training and testing process on the training and validation data sets

The model's training losses, including train/box_loss, train/seg_loss, train/cls_loss, and train/dfl_loss, consistently decreased over the training epochs. The train/box_loss dropped from approximately 1.9 to below 0.9, indicating improved accuracy in bounding box predictions, the train/seg_loss reduced from around 3.2 to below 1.6, reflecting enhanced segmentation capabilities, suggesting overall improvements in handling deformable parts of objects. Similarly, the validation losses followed a downward trend: the val/box_loss decreased from around 1.9 to approximately 1.0, the val/seg_loss dropped from about 4.3 to around 2.0. This consistent reduction in both training and validation losses indicates effective model training and good generalization to unseen data, with minimal signs of overfitting.

The model's performance metrics showed substantial improvements and stability over the training period. Precision metrics for bounding boxes (metrics/precision(B)) and masks (metrics/precision(M)) both increased and stabilized around 0.9, demonstrating high accuracy in predictions. Recall metrics for bounding boxes (metrics/recall(B)) and masks (metrics/recall(M)) increased and stabilized around 0.7, indicating robust detection capabilities. The mean Average Precision at IoU threshold of 0.5 (metrics/mAP50(B), metrics/mAP50(M)) reached approximately 0.8, reflecting high performance in segmentation and detection tasks. The mean Average Precision across IoU thresholds from 0.5 to 0.95 (metrics/mAP50-95(B), metrics/mAP50-95(M)) increased and stabilized around 0.5, showcasing the model's ability to maintain high precision and recall across varying IoU levels.



Epoch = 50	Precision	Recall	mAP50	mAP50-95
Batch = 64	0.906	0.789	0.848	0.557
Batch = 32	0.93	0.776	0.848	0.562

Fig. 5 The results table evaluates the model's accuracy on the test data set

Precision (0.93): This index is very high, showing that among the points predicted by the model as positive, 93% are correct. This means that the model is less likely to produce positive results. false positives. Recall (0.776): This index is relatively lower than precision, indicating that the model can only detect about 77.6% of true positives in the test set. This suggests that the model can be ignored. Missing a significant number of false negatives. mAP@50 (0.848): Mean Average Precision at IoU threshold 0.5 is 84.8%, which is a good index, showing that the model is able to detect and segment objects quite accurately when using IoU threshold 0.5. mAP@50-95 (0.562): Mean Average Precision over IoU thresholds from 0.5 to 0.95 is 56.2%. This is a more comprehensive index than mAP@50 and shows that as IoU thresholds increase, performance improves. of the model decreases. This suggests that the model may have difficulty requiring a higher level of accuracy in the location and size of objects.

From the results table above, we can see that the model is still able to predict accurately above a relative IOU threshold of 0.5. When we increase the threshold to higher levels, the model's effectiveness will decrease to only 56.2% on another data set to which the model has not been exposed.

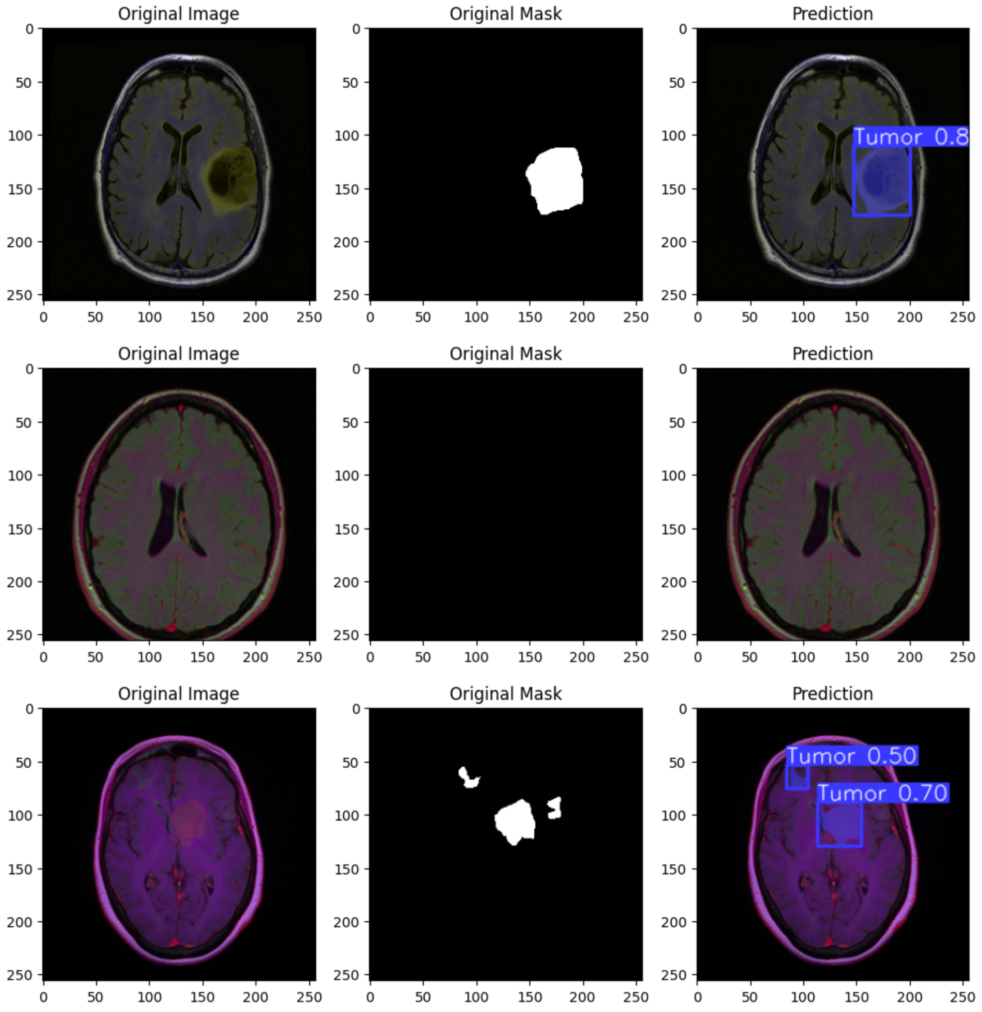


Fig. 6 Some images include input images, mask images and prediction results in the test data set

6 Discussion

In this project, we train a deep learning model to learn fully automatically evaluated image features of neuroendocrine tumors and perform image segmentation of tumor-containing locations. u. The results of the problem are at a moderate level. Deep learning algorithms are used to divide tumors into two types: Non-tumor and Tumor, then segment images containing tumors and return the result as the location of the modeled tumor area. forecast. Using imaging to predict tumor appearance is of great importance, and if accurate models are developed, it could be integrated into current treatment paradigms in a variety of ways. In the simplest scenario, if the model is highly accurate, it could replace humans in predicting whether a patient's MRI image shows a head tumor or not. However, there are many other ways in which even moderately performing models can contribute valuable information. Imaging data are available early in the process and therefore a preliminary assessment of tumor biology before surgery may still be useful in guiding next steps. In the absence of in-depth tissue analysis, approximate visual classification can be of great value as it helps to predict early whether a patient has the disease so that timely treatment can be initiated. Finally, even if the overall accuracy is not perfect, it is still possible to perform at high positive or high predictive values and still use alternative image-based models. To triage patients for more intensive treatment, even if it is a minority of patients. Automated tumor segmentation such as the method presented in this study has many advantages. The algorithm is deterministic, meaning that given the same image, the algorithm will always make the same evaluation. Applying computer algorithms is inexpensive and results are obtained quickly.

7 Limitation

Our work has limitations, this is only a first step towards image-based image segmentation replacement. We only built a tumor region segmentation model on a rather limited data sample size. The mechanism was used in the study (110 patients) because data containing comprehensive genomic tests along with imaging are still rare. Regarding segmentation algorithms, there are many methods to perform automatic segmentation of brain tumors that can be considered to compare and further improve our results. Various algorithms have been proposed such as ResNet, Inception and DenseNet are integrated. Combining pre-trained segmentation models also gives very good results. We have not handled the data imbalance in the training data set, leading to many cases where the model will miss, leading to bad results. Evaluation of the Recall index is not highly accurate.

8 Conclusion

Brain tumor segmentation using MRI images using deep learning models is a powerful and effective diagnostic imaging tool that plays an important role in cancer assessment and treatment. However, it is important to note that it is still not possible to guarantee that algorithms can predict completely accurately and further research and development is still needed. At the same time, there are many limitations in terms of cost, data and resources to build a machine learning model that can be used to support humans in this field.

References

[1] Mateusz Buda, M.A.M. Ashirbani Saha: Association of genomic subtypes of lower-grade gliomas with shapefeatures automatically extracted by a deep learning algorithm (2019). https://www.researchgate.net/publication/333679533_Association_of_genomic_subtypes_of_lower-grade_gliomas_with_shape_features_automatically_extracted_by_a_deep_learning_algorithm

[2] Buda, M.: Brain mri segmentation. https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation

[3] Ultralytics: Yolov9: A leap forward in object detection technology. https://docs.ultralytics.com/models/yolov9/

[1] [2] [3]