JBE Journal of Biostatistics and Epidemiology
J Biostat Epidemiol. 2025;11(1): 15-35

ORIGINAL ARTICLE
DOI: *https://doi.org/10.18502/jbe.v11i1.19315*
*Received **October** 08, 2024; Accepted **December** 17, 2024;*
*Published online **March** 15, 2025.*

A Quantitative Comparison between Focal Loss and Binary Cross-Entropy Loss in Brain Tumor Auto-Segmentation Using U-Net

Mahdi **Shafiei Neyestanak**[1,2], Hamid **Jahani**[3], Mohsen **Khodarahmi**[4], Javad **Zahiri**[5], Mostafa **Hosseini**[2], Amirali **Fatoorchi**[6], Mir Saeed **Yekaninejad**[1*]

[1]Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran.
[2]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, San Diego, CA, USA.
[3]Department of Data Science, Faculty of Interdisciplinary Science and Technology, Tarbiat Modares University, Tehran, Iran.
[4]Bahar Medical Imaging Center, Karaj, Iran.
[5]Department of Neuroscience, University of California San Diego, La Jolla, San Diego, CA, USA.
[6]Department of Industrial engineering, IT engineering Group, P.hD. student, K. N.Toosi University of Technology, Tehran, Iran.

ABSTRACT

Introduction: Brain tumors are among the most fatal cancers and cause the deaths of many people annually. Early diagnosis of a brain tumor can help save the patient's life.

Methods: We have collected a dataset consisting of 314 brain MRI images in all planes taken after administering a contrast medium with the dimension of 800*512, which offers the highest resolution. First, skull stripping has been implemented to separate the brain from other parts in the images. Next, we have annotated the tumors in the images under the supervision of experienced radiologists to create ground truth. To determine the most effective model versions for all three loss functions, hyperparameter tuning was performed. Following the comparison, the study further evaluates the effectiveness of two loss functions, Binary Cross-Entropy (BCE) and Focal loss, specifically in handling tumor regions within the dataset.

Results: The two proposed loss functions were evaluated using 5-fold cross-validation, and the average precision, recall, and F1-score were 76.16%, 71.9%, and 74.52 for BCE loss and 82.92%, 79.32%, and 81% for the Focal loss on the test data, respectively. Moreover, the accuracy for BCE loss was 99.03% and 99.44% for the Focal loss.

Conclusion: We recommend using BCE loss cautiously in classification tasks without data imbalance and emphasize the adoption of Focal loss for more accurate and reliable results in brain tumor segmentation.

Key words: Brain tumor segmentation; Deep learning; Convolutional neural network; U-net- architecture; Diagnosis; MRI image

*Corresponding Author:
yekaninejad@yahoo.com

INTRODUCTION

Brain tumors are among the most lethal cancers, causing the deaths of a significant number of people annually. According to the 2018 Cancer Registry results, over 18 million registered cases of cancer in both sexes, of which approximately 300,000 were related to brain cancers. Additionally, there were more than 9 million deaths attributable to global cancers in 2018, of which more than 240,000 (2.71%) were due to brain cancers.[1] An estimated 16,830 deaths were attributed to primary malignant brain tumors in the US in 2018. According to a study from 2011 to 2015, in the US, the 5-year and 10-year relative survival rates of patients with malignant tumors were 35.0% and 29.3%, respectively. Based on this study, brain and other central nervous system tumors (both malignant and non-malignant) had an average annual age-adjusted incidence of 11.20 per 100,000 people aged 15–39 years. Also, these tumors were the second most common cancer in males in this age group and the third most common cancer in females in this age group.[2]

Magnetic resonance imaging (MRI) is the primary scanning tool for detecting and diagnosing tumors, as it offers excellent contrast for soft tissues. In general, MRI serves three main functions in the workup of intra-axial tumors: tumor diagnosis and classification, treatment planning, and post-treatment management.[3,4] MRI images, based on this, play a crucial role in discovering tumors and potentially saving the lives of patients. It is worth mentioning that before the scan, a contrast medium, known as a dye, is administered to enhance the image's precision.

Image segmentation entails dividing a digital image into distinct segments, each representing a specific visual characteristic. More precisely, it involves labeling individual pixels in the image, where pixels sharing the same label possess similar visual attributes. This technique is widely used for object and edge detection within images[5] Brain tumor image segmentation is a crucial process performed by experts to distinguish distinct regions within a brain image, enabling the identification and differentiation of tumors from normal brain structures. This precise segmentation plays a vital role in treatment planning within the medical field. Locating tumors is an arduous task, heavily reliant on the experience, skills, and meticulous decisions of physicians, who analyze each slice of the image. Timely detection of cancerous tumors significantly enhances a patient's chances of recovery and survival following treatment. Nevertheless, this intricate task demands a substantial time investment from qualified experts to accurately segment brain tumors.[6] Furthermore, the precise location of tumors often involves some level of uncertainty. Gliomas, which are primary brain tumors, arise from glial cells that support nerve cells. Due to the extensive spatial distribution of glial cells, both High-Grade Gliomas (HGG) and Low-Grade Gliomas (LGG) can manifest at various locations throughout the brain.[7]

In addition, this process is prone to be influenced by individual opinions and if the person delineating the region of tumor is not a well-trained technologist, it will usually yield poor segmentation results.[6, 7]

Historically, brain tumor segmentation has evolved from manual methods to more sophisticated computational approaches. Early methods in the late 1990s and early 2000s relied on traditional machine learning algorithms with hand-crafted features. These included expert systems using multi-

spectral histograms, segmentation using templates, graphical models with intensity histograms, and tumor boundary detection from latent atlases.[8,9] These early techniques primarily focused on segmenting the whole tumor region, often under strong and unrealistic assumptions. Additionally, the manually designed feature engineering was limited by prior knowledge and could not be generalized effectively. These early methods also struggled with challenges such as appearance uncertainty and data imbalance.

With the advent of deep learning, researchers shifted towards using deep neural networks (DNNs) to address the challenges in brain tumor segmentation.[10,11] Pioneering work involved customizing deep convolutional neural networks (DCNNs) for accurate tumor segmentation. Breakthroughs like the Fully Convolutional Network (FCN)[12] and U-Net[13] led to innovations focusing on building fully convolutional encoder-decoder networks, enabling end-to-end tumor segmentation.

With automatic learning, capturing global context, scalability, and adaptability through fine-tuning Convolutional Neural Networks (CNNs) achieve superior performance in various segmentation tasks due to their ability to learn robust features and generalize well on diverse datasets, making them a powerful and widely used approach for image segmentation in modern computer vision applications.[14,15] U-Net is a highly effective CNN architecture known for its skip connections, U-shaped design captures both local and global information, enabling strong performance in biomedical segmentation. its efficiency in segmenting medical images stems from its encoder-decoder structure, which allows for precise localization and context understanding. The model has become a cornerstone in the field, leading to numerous extensions such as UNet++ and Attention-UNet, which build upon its foundation to improve performance. In diffusion models, U-Nets play a critical role by adapting across various time steps to iteratively reconstruct original data from noise-introduced training data, enhancing the generative capabilities of these models.[13,16,17] Despite significant advancements in improving the performance of segmentation tasks, the optimal setup of hyperparameters for training the best model for this task remains an open question. Additionally, few studies have focused on identifying the most effective loss function for brain tumor segmentation. Isensee et al.[18] demonstrated that enhancing segmentation performance involves more than just adjusting the architecture; factors such as the choice of loss function, training strategy, and post-processing also play critical roles.

Contribution of the study

We collected a high-resolution dataset of brain MRI images with contrast medium, enhancing clarity and ensuring diverse tumor variations for improved model training. In collaboration with radiology experts, we selected high-quality images with appropriate dimensions in all planes. Skull stripping was performed on each image for increased accuracy.

Our study highlights the importance of evaluation metrics like the Dice coefficient and Hausdorff distance, which enhance segmentation accuracy and tumor localization. We conducted a thorough comparison of these metrics to identify the most effective one for improving segmentation performance in brain tumor segmentation task which less attention paid into.

Additionally, we explored various loss functions (e.g., Dice loss, Cross-Entropy loss, Tversky loss) to guide model training and improving tumor boundary identification. Our analysis provides insights into selecting the most suitable loss function and contributes to the advancement of medical image analysis. Ultimately, our work sets a new benchmark for brain tumor detection using U-Net and informs future research in the field.

METHODS

In this study, we conducted a comprehensive literature review on key loss functions in the segmentation task, including Tversky Loss, Dice Loss, and Focal Loss, which are commonly utilized in image segmentation with U-Net. Our focus was on comparing the efficacy of the Binary Cross-Entropy (BCE) loss function and Focal Loss, particularly when dealing with the tumor region of interest (ROI) in our dataset. To determine the most effective model versions for all three loss functions, we performed hyperparameter tuning using Ray Tune.[19] During this process, we also examined how the number of epochs influenced the maximum IoU score of our models. Our aim was to identify which loss function yielded the best results for this specific task. Our findings indicated that BCE loss and Focal loss outperformed the other loss functions, further emphasizing their importance in image segmentation tasks, particularly when dealing with the tumor region.

For our hyperparameter tuning process, we focused on optimizing the number of epochs and the alpha values associated with loss functions. This process is crucial in machine learning and deep learning models, as it involves selecting the best set of hyperparameters for a given algorithm to achieve optimal performance on a specific task. The alpha parameter in loss functions controls the balance between false positives and false negatives. Additionally, the number of epochs represents the number of times the model passes through the entire training dataset. It can significantly influence the training process, affecting both underfitting and overfitting. By adjusting the number of epochs, we can find the optimal training duration that balances model performance.

During hyperparameter tuning for BCE loss, we focused on adjusting the learning rate, weight decay, and batch size, as BCE loss does not have an alpha parameter like Tversky Loss or Focal Loss.

Table 1 provides an overview of the hyperparameter values tested for each loss function. The hyperparameter tuning process was conducted separately for each of the three loss functions. This systematic approach allowed us to identify the optimal hyperparameters for each loss function, leading to more effective segmentation models.

After hyperparameter tuning, we utilized the optimal hyperparameters to retrain the models from the ground up. For Focal Loss, the best combination, with an alpha value of 0.6, was trained for 20 epochs using the validation data, resulting in a promising Dice coefficient of 81%. On the other hand, the top-performing BCE loss model was also trained for 20 epochs, yielding an impressive F1 score of 75%. Accordingly, we used these different loss functions in the standard U-Net model to assess their overall performance in the semantic segmentation of brain tumors.

Table 1. hyperparameter values explored for each loss function

Loss function	Hyperparameters	Measures
Dice loss	#Epochs	5, 10,15,20,25,30
Tversky loss	#Epochs alpha	5, 10,15,20,25,30 0.5,0.6,0.7,0.8,0.9
Focal loss	#Epochs alpha	5,10,15,20,25,30 0.5,0.6,0.7,0.8,0.9

Binary Cross-Entropy loss function (BCE)

Cross-entropy is a measure of the discrepancy between two probability distributions. The performance of a semantic segmentation model is evaluated using cross-entropy loss. The probability output value of the cross-entropy loss function ranges between 0 and 1. This value increases when the predicted probability of a pixel in the image belonging to the actual class is higher. This process can be viewed as a classification algorithm. Binary cross-entropy is a specific form of cross-entropy, where the target of the prediction is either 1 or 0.

Cross-entropy is the default loss function used for image segmentation, as it is mathematically related to accuracy.[20] In the equation below, we have provided the Log loss formula. The term p In this formula is the probability of class 1, and (1- p) is the probability of class 0. When a pixel belongs to class 1, the first piece of the formula becomes involved, and if the pixel belongs to class 0, the second part of the equation becomes active. Through this process, we can calculate the Binary cross-entropy.[21]

$$-p \cdot log(\hat{p}) + (1 - p) \cdot log(1 - \hat{p})$$

Focal Loss

The Focal Loss function is designed to address the class imbalance problem that often occurs in binary classification tasks.[22] In medical image segmentation, dealing with imbalanced data is a common challenge. This occurs when one class, typically the "negative" class, dominates the dataset, leading to biased training and suboptimal performance. To address this issue, the Focal Loss function is used, selectively reducing the loss for well-classified examples while prioritizing hard-to-classify and misclassified examples. By focusing on these difficult samples, Focal Loss helps improve the overall performance of the model.

To address the class imbalance problem, we can incorporate weights into the cross-entropy loss formula. These weights are assigned to hard-to-classify examples, specifically false negatives, while correctly classified examples are referred to as true negatives. By introducing these weight adjustments, the model can better handle imbalanced data, allowing it to focus on challenging cases and improving its performance in medical image segmentation tasks.

$$Focall\ Loss(p, \hat{p}) = -\big(\alpha(1-\hat{p})^{\gamma} p log(\hat{p}) + (1-\alpha)\{\hat{p}\}^{\gamma}\ (1-p) log(1-\hat{p})\big)$$

Dataset

The training data is a set of 314 MRI images (800 x 512 pixels) from the brains of 108 patients diagnosed by expert with a brain tumor and went to the imaging center between 10 Jan 2021 to 13 Feb 2022. These MRI images were taken with and without contrast to obtain the highest resolution images and obtained with multiplanar and in different sequences with GE 1.5 tesla machine. These images are in sagittal, axial, and coronal planes and the average age of the patient in the study is 43 years old.

In this study, each image is accompanied by a fully annotated ground truth segmentation map, which precisely identifies the tumor's location in a separate image with similar dimensions. To enhance the accuracy of the segmentation task, the images underwent several preprocessing stages including skull stripping and annotation and labeling. Figure 1 displays a collection of examples used in the study, showcasing the original images, their corresponding annotations, and the resulting output masks.

We analyzed brain tumor images to quantify the tumor-to-brain area ratio and classify tumors as either High-Grade Glioma (HGG) or Low-Grade Glioma (LGG). To enhance image quality, we applied Gaussian blurring for noise reduction, followed by segmentation of brain and tumor regions using Otsu's thresholding and binary thresholding, respectively (Figure 1). The mean ratio of tumor regions to non-tumor areas in the images is 0.22, with a median of 0.19, indicating a class imbalance between the two regions.

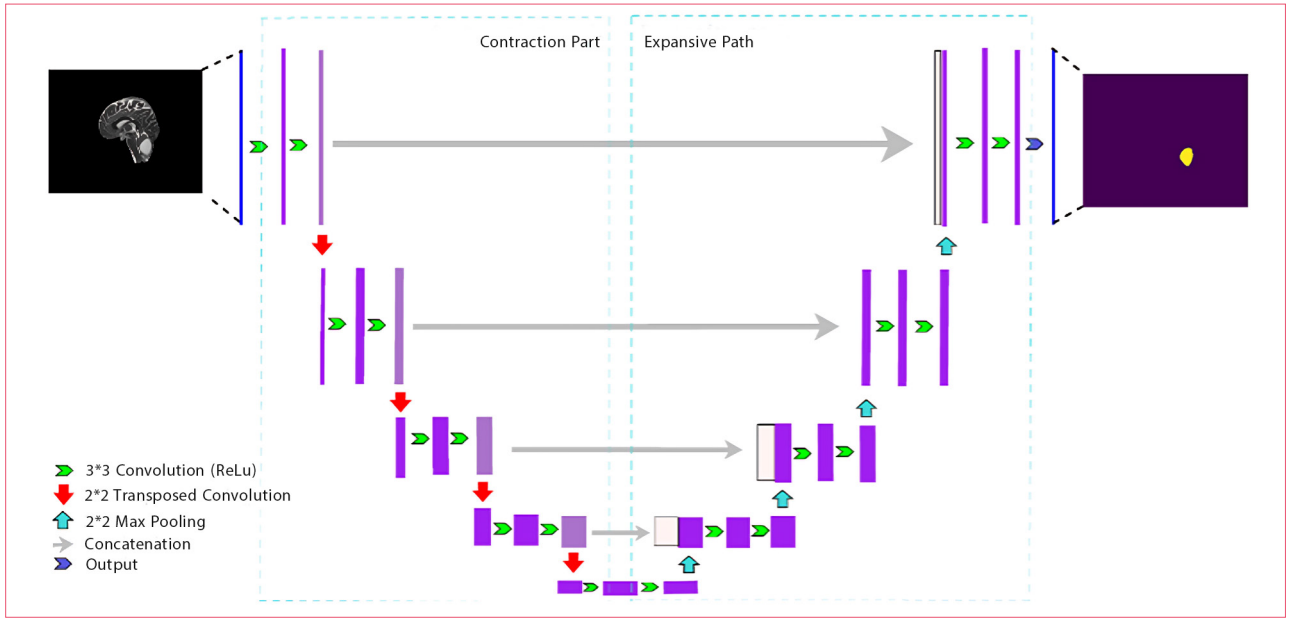

Figure 1. U-Net structure with 4 encoding and 4 decoding blocks

Skull Stripping

Skull stripping, also known as brain extraction or brain masking, is a critical preprocessing step in the analysis of MRI data. The purpose of skull stripping is to remove non-brain tissues from the MRI

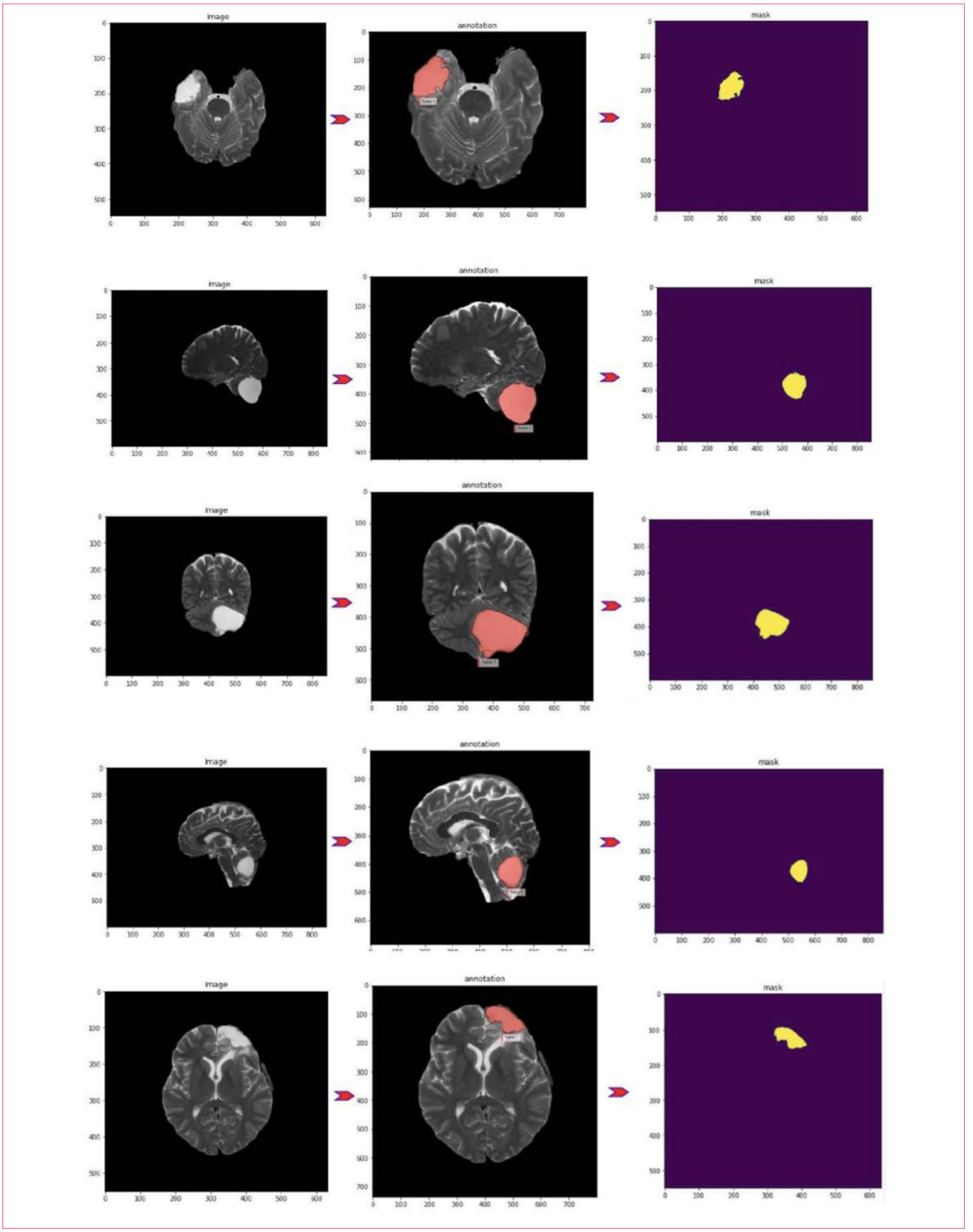

Figure 2. Five cases from the dataset show the annotating of the tumor area and the final Mask or 21 ground truth. The model will use these masks to detect the tumor region. Radiology experts have done these annotations.

images, such as the skull, scalp, and other surrounding tissues, leaving only the brain region for further analysis. This step is crucial for various neuroimaging applications, including brain segmentation, volumetric analysis, cortical thickness measurement, and functional connectivity studies, among others. Simply, the process of skull stripping will isolate the brain tissue from non-brain tissue in an MRI image.[23] We used region-based binary mask extraction for this task. Figure 3 demonstrates the transformation of a raw MRI image into a suitable format for the segmentation task, achieved through the skull stripping process.

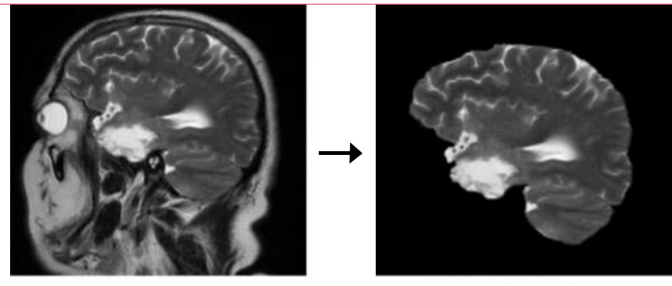

Figure 3. An example of a raw MRI image transformed into a suitable image for segmentation task through the skull stripping process.

Evaluation metrics

To comprehensively assess the performance of our semantic segmentation task, it is crucial to define the following metrics: True Negatives (TN), True Positives (TP), False Negatives (FN), and False Positives (FP). These definitions lay the foundation for the evaluation metrics that we will subsequently introduce.

- True Positive (TP): The model correctly identifies the tumoral pixel.
- False Positive (FP): The model incorrectly identifies a pixel as tumoral.
- False Negative (FN): The model incorrectly identified the pixel as non-tumoral.
- True Negative (TN): The model correctly identified a normal pixel as non-tumoral.

Dice Coefficient (F1 Score): The Dice coefficient, also known as the Sørensen–Dice coefficient, is a similarity metric used to compare the similarity or overlap between two sets. In the context of image segmentation and medical image analysis, it is often used to evaluate the performance of image segmentation algorithms by measuring the agreement between the segmented image and the ground truth (manually annotated) image.[24] Basically, we can also say Dice coefficient measures the similarity between predicted pixels and ground truth and ranges from zero to one. Let TP, FP, and FN be variables representing true positives, false positives, and false negatives, respectively.

$$\frac{2.TP}{2 \cdot TP + FP + FN}$$

The Dice coefficient is defined as follows: $\frac{2 \cdot |A \cap B|}{|A| + |B|}$

where:

- A is the first set (e.g., the pixels of the segmented image).
- B is the second set (e.g., the pixels of the ground truth image).
- |A ∩ B| is the cardinality of the intersection of A and B (the number of pixels that are correctly classified).
- |A| is the cardinality of set A (the total number of pixels in the segmented image).
- |B| is the cardinality of set B (the total number of pixels in the ground truth image).
- The Dice coefficient ranges from 0 to 1, with 0 indicating no overlap between the sets (completely dissimilar) and 1 indicating perfect overlap (complete similarity).

Precision

Precision is a performance metric used to evaluate the accuracy of a segmentation algorithm. Precision measures the proportion of the correctly classified pixels for a particular class among all the pixels that the algorithm has classified as belonging to that class. We can define it as the proportion of pixels in our segmentation model that correspond to pixels in the ground truth.[25]

$$\frac{TP}{TP + FP}$$

Recall (Sensitivity)

Is a performance metric used to evaluate the completeness of a segmentation algorithm. Recall measures the proportion of correctly classified pixels for a particular class among all the pixels that belong to that class in the ground truth (manually annotated) image. To be more accurate: Recall measures the ratio of pixels in the ground truth that were successfully detected by our segmentation model[26]

$$\frac{TP}{TP + FN}$$

Accuracy

The easiest way to understand the percent of pixels in the image that are segmented correctly is accuracy. Because most of the pixels in an MRI image are background and the region of interest or tumor includes a small area of the image, accuracy is often high in brain tumor segmentation tasks and is not an interested metric The best accuracy is 1.0, whereas the worst is 0.0.

$$\frac{TP + TN}{TP + TN + FN + FP}$$

RESULTS

We conducted two separate training sessions for our U-net model using high-resolution brain MRI images. In the first session, we utilized the BCE loss, while in the second session, we employed the

Focal loss. The implementation of the model was carried out in Python using Spyder 5.0.1. To leverage powerful machine learning capabilities, we trained the model on the Cloud Tensor Processing Unit (TPU) available in the Google Colab environment, specifically designed for accelerating machine learning tasks. Throughout the training process, we relied on Keras and TensorFlow versions 2.8.0. Lastly, to assess the performance of our model, we implemented a fivefold cross-validation technique and the obtained results are presented in Table 2.

Table 2. Full model output in terms of all evaluation metrics separately for BCE loss and Focal Loss

Loss K-folds	Accuracy	F1(Dice Coefficient)	Precision	Sensitivity (Recall)	Specificity
Focal Loss					
Fold-1	98.98 %	78 %	88.5 %	62.3 %	99 %
Fold-2	99.87 %	81.1 %	86.1 %	79.7 %	99.1 %
Fold-3	99.21 %	79.8 %	77.6 %	89.7 %	99.7 %
Fold-4	99.84 %	79.4 %	81.3 %	75.6 %	99.2 %
Fold-5	99.33 %	86.7 %	81.1 %	89.3 %	99.4 %
5-Fold Mean	99.44 %	81 %	82.92 %	79.32 %	99.2 %
Blnary Cross Entropy Loss					
Fold-1	99.32 %	72.3 %	74.2 %	68.3 %	98.3 %
Fold-2	98.6 %	77.8 %	76.1 %	78.5 %	98.8 %
Fold-3	98.63 %	63.5 %	78.5 %	74.3 %	99.2 %
Fold-4	99.15 %	76 %	87.4 %	63.9 %	99.6 %
Fold-5	99.45 %	83 %	64.6 %	74.5 %	98.7 %
5-Fold Mean	99.03 %	74.52 %	76.16 %	71.9 %	98.9 %

Figure 4 presents performance comparison charts for the BCE loss and Focal loss using essential evaluation metrics, Precision, F1 score, and Recall, in the segmentation task. It is evident that the Focal loss outperforms the BCE loss, showing improvements of 6.4% in Dice coefficient, 6.76% in Precision, and 7.42% in Recall.

In Figure 5, we demonstrate model prediction results for tumor detection on random MRI images. In the first row, the raw brain MRI image with the tumor is displayed, while the second row shows the corresponding annotation map or mask. The model assigns a probability value to each pixel, indicating the likelihood of a tumor's presence or absence. These probabilities range from zero to one, with higher values representing brighter points in the images displayed in the last row. The final prediction of the model for tumor localization is depicted in the third row.

To create this prediction, the model applies a threshold of 0.5 to the probabilities. Pixels with probabilities greater than 0.5 are set to one, representing tumors, while pixels with probabilities lower than 0.5 are set to zero, indicating the absence of tumors. The pixels with a probability equal to one are identified as tumors and are plotted accordingly as our final prediction.

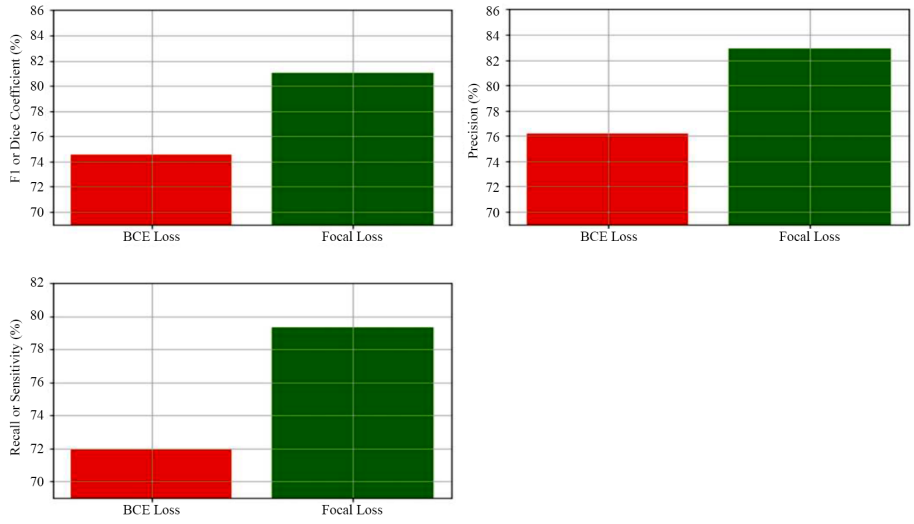

Figure 4. The performance comparison results of the BCE loss and Focal loss in terms of Precision, F1 score and Recall.

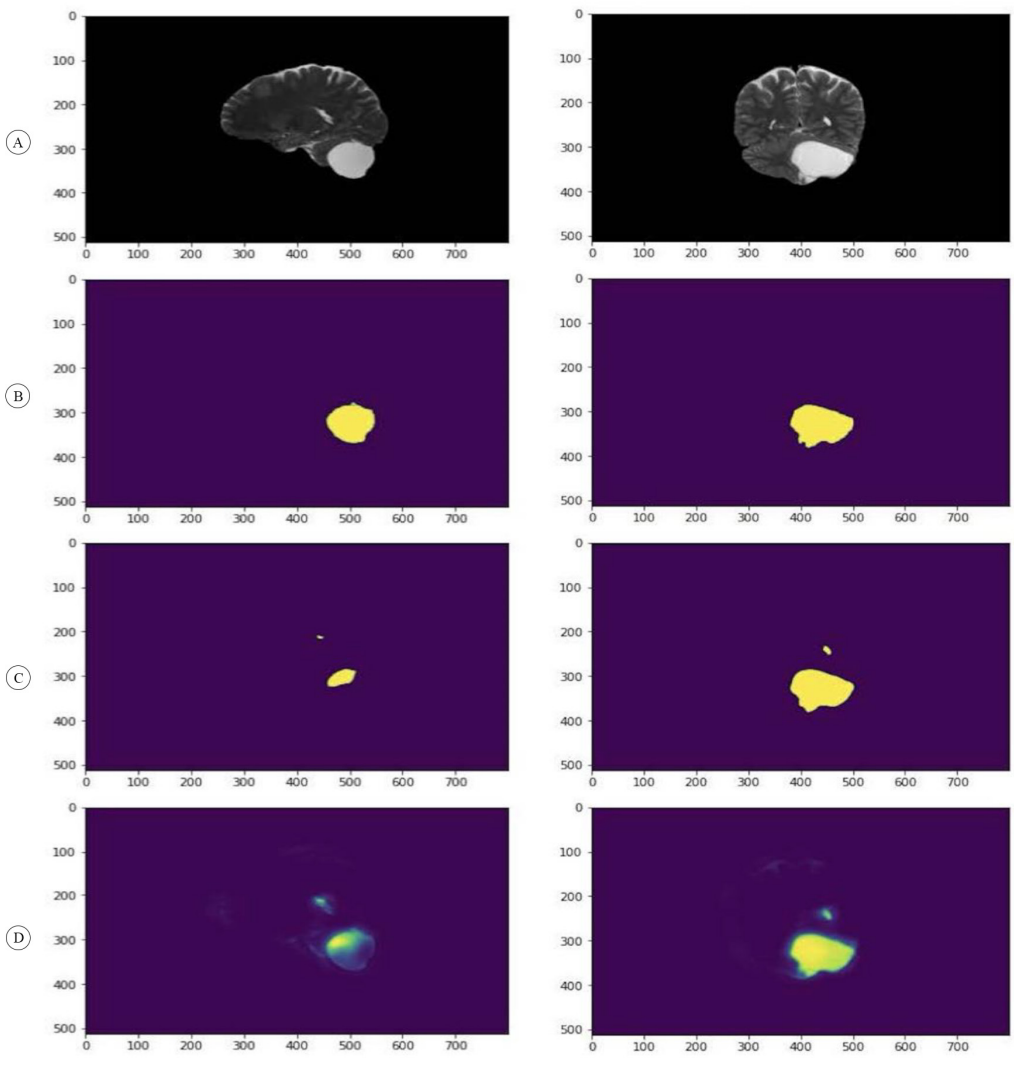

Fig 5

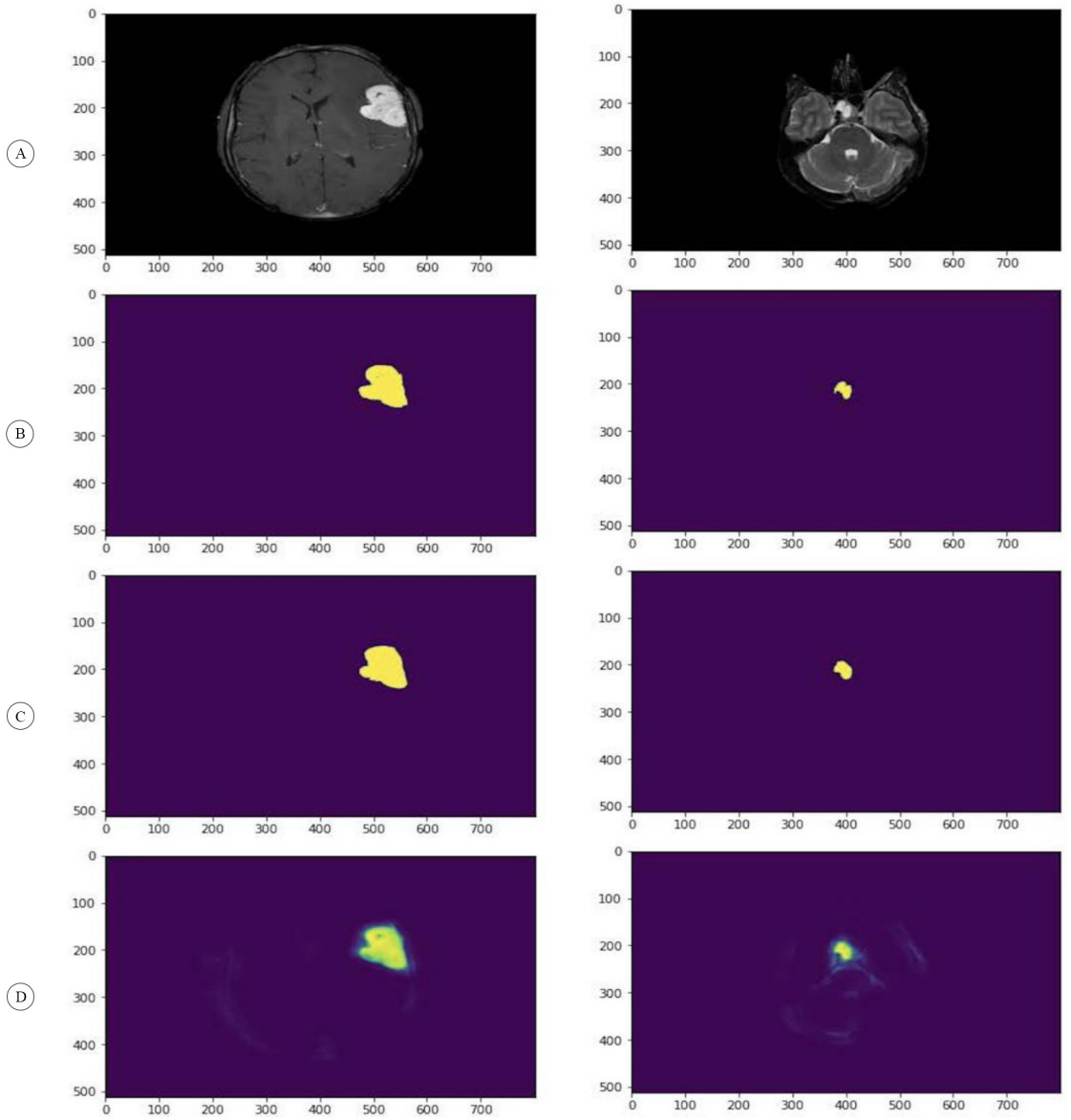

Fig 5

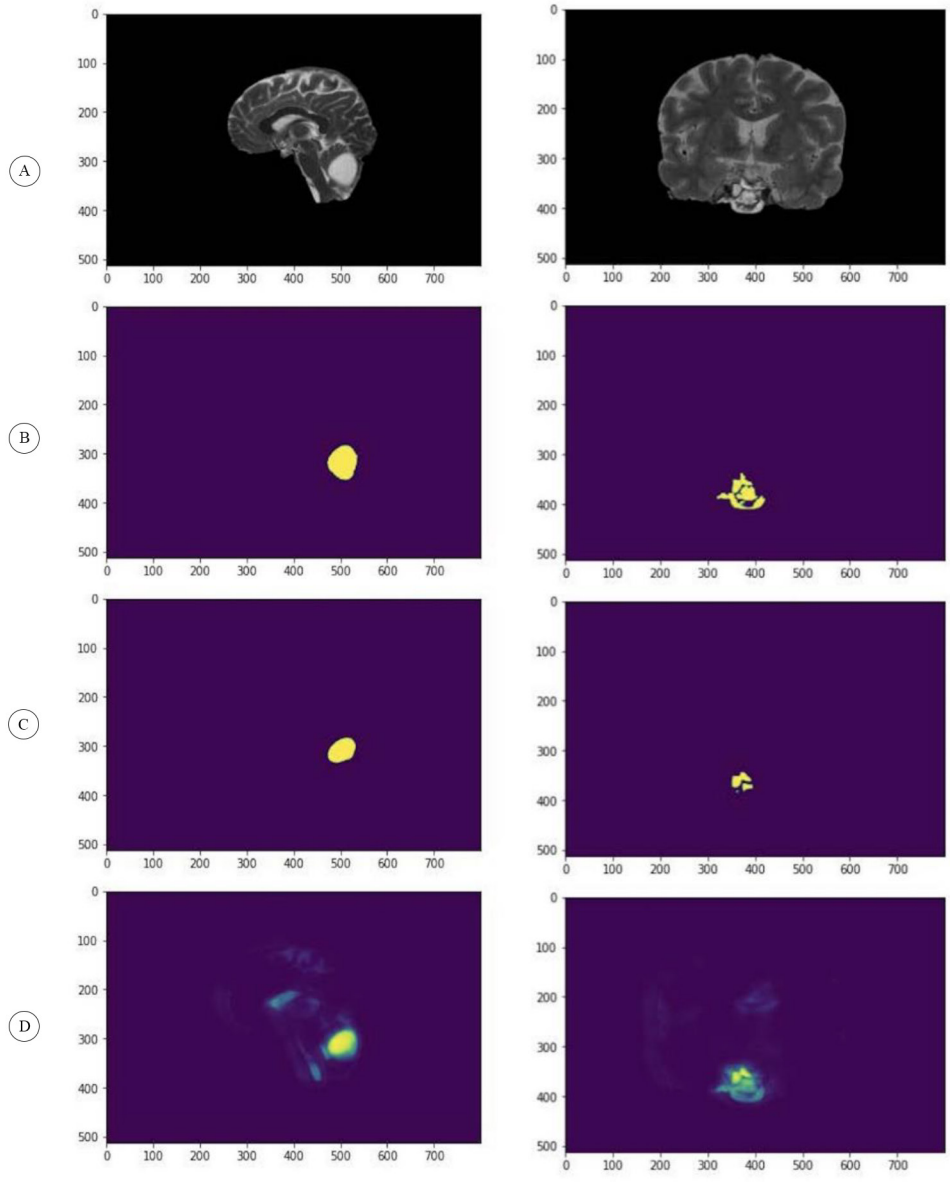

Fig 5

Figure 5. Six cases of the dataset showing the A) Image, B) Corresponding ground truth, C)Final prediction of tumor done by model, and D) Each pixel's probability of being tumor.

Hausdorff distance

As the ultimate evaluation criterion, we computed the Hausdorff distance for the image segmentation task, specifically focusing on the performance of the focal loss, which we selected as our best-performing loss function. Hausdorff distance plays a crucial role in image segmentation by providing a quantitative measure of the similarity between the segmented regions produced by an algorithm and the ground truth or reference boundaries. It is an important evaluation metric that helps assess the accuracy and quality of the segmentation results. The best range for the Hausdorff distance depends on various factors, such as the resolution and scale of the images, the complexity of the segmented regions, and the specific application of the segmentation. Generally, a lower Hausdorff distance indicates better accuracy and similarity between the segmented boundaries and the ground truth boundaries. However, the acceptable range can vary depending on the specific task and the domain[27,28] Table 3 displays the results of the Hausdorff distance values obtained for each fold using the Focal loss function.

Table 3. The Hausdorff distance values obtained for each fold using the Focal loss function.

K-folds in Focal Loss	Hausdorff Distance (mm)
Fold 1	46.42811
Fold 2	42.54174
Fold 3	52.68178
Fold 4	50.63595
Fold 5	48.08201

%95 CI = (43.22mm,52.92mm)

DISCUSSION

In this study, we intended to evaluate an essential factor in the process of automated brain tumor segmentation, which is loss function. We compared two widely used loss functions using evaluation metrics relevant to segmentation tasks, and we realized that Focal loss is performing significantly better. The acceptable performance of focal loss compared to BCE loss maybe because of its better encounter with imbalanced data. The loss function is a dynamically moderated version of cross entropy loss, where the scaling factor tends to zero as confidence in the correct class increases. Intuitively, this scaling factor can automatically down-weight the contribution of easy examples during training and rapidly focus the model on hard examples. It is important to note that examples (pixels) that are consistently well-classified at the early training stage are easy examples and examples (pixels) that are misclassified are hard examples.[29] Moreover, in terms of network structure, we found that U-net performs best among others in segmentation tasks, and changing any hyper-parameter or model would decrease the precision and recall. Hence, we used standard U-net architecture with four encoding and four decoding paths and different loss functions.

Inspired by the concept of focal loss, Ken C. L. et al.[30] have presented an exponential, logarithmic loss that balances the labels by their relative sizes and their segmentation difficulties. They have improved the dice coefficient due to using logarithmic Dice loss. However, the resolution of the images they

used was less than ours. Also, instead of data augmentation, we used a more diverse dataset to train the model, which can directly affect the generalization of the model.

Training a U-net on MRI images is typically not a very computationally intensive task, yet it often yields highly promising results. On the other hand, Ullah et al.[31] introduced a patch-based convolutional neural network (PBCNN) approach for accurate brain tumor segmentation from MRI scans. The authors utilize the BraTS datasets from 2012 to 2018, preprocessing the data with techniques like Gaussian filtering and intensity normalization. The images are divided into smaller patches to improve computational efficiency and allow the model to focus on local features. Their proposed PBCNN architecture outperforms existing state-of-the-art methods, achieving a Dice coefficient of 0.91 and an accuracy of 0.96. The study emphasizes the importance of big data analysis in medical imaging and highlights the potential of PBCNN models for improving brain tumor diagnosis and treatment. One weakness of the paper is the computational cost associated with training the proposed PBCNN model. This suggests that the method may not be feasible for researchers or clinicians without access to significant computing resources.

One limitation of this study is its reliance on single-center data, which could introduce potential demographic biases (e.g., age, ethnicity) and scanner-specific artifacts in longitudinal studies due to the use of uniform imaging protocols and equipment. Since the dataset is sourced exclusively from one imaging center, it may not fully capture patient populations diversity or genetic diversity that typically seen in real-world clinical settings. However, for the segmentation task in this study, this limitation is less concerning as we are primarily focused on computational accuracy. Longitudinal studies, however, may not be fully applicable to this dataset.

Additionally, there is room for exploring variations of the standard U-Net architecture, such as U-Net++[32] or Attention U-Net[33], which incorporate features like nested skip connections or attention gates to improve feature extraction and boundary precision.

Daobin Huang et al.[34] in 2021 have used a hybrid loss for brain tumor segmentation to optimize their proposed network for the class imbalance problem. They offered two hybrid loss functions comprising contributions from different losses, including recall loss, combined Dice loss, and cross-entropy loss, which has improved their DSC value. Accordingly, their model achieved Dice scores of 85.

Mohseni Salehi et al.[35] propose a generalized loss function to deal with unbalanced data. Their loss function is based on the Tversky index to handle the issue of data imbalance and achieve a much better trade-off between precision and recall. Their results in multiple sclerosis lesion segmentation on magnetic resonance images show improved F2 score and Dice coefficient.

The automated and semi-automated methods that have progressed through recent years can be classified into two groups: Generative and discriminative. Generative methods model the joint probability distribution of the image and its corresponding segmentation labels. In other words, they try to learn the statistical relationship between the input image and the output segmentation labels.[36,37]

On the other hand, Discriminative methods, focus on modeling the conditional probability distribution of the segmentation labels given the input image. Instead of trying to model the entire joint distribution, they directly learn the decision boundary between different segments based on the input features. Popular discriminative models for image segmentation include Support Vector Machines (SVMs), Random Forests, and Deep Learning-based models like Convolutional Neural Networks (CNNs). Moreover, CNNs for brain tumor segmentation can hardly be grouped into single-label prediction and dense prediction architectures. Single label prediction architectures take preprocessed patches as input and predict the label of the central pixel in the patch.[38] To bring an example, Pereira et al.[39] proposed a 2D convolutional neural network that creates a single-label prediction and fulfills the best performance in the BRATS 2015 challenge. In addition, a two-phase training scheme was proposed to deal with the class imbalance problem. Since only the label of the central voxel is predicted each time, the single-label prediction networks are very slow during the inference phase.[11]

CONCLUSION

In conclusion, the findings of this study demonstrate the superiority of Focal loss over BCE loss in the context of brain tumor auto-segmentation tasks utilizing CNN and U-net architecture. The remarkable improvements observed in precision, F1 score, Recall, and accuracy clearly indicate that Focal loss is an optimal choice for addressing imbalanced data in segmentation tasks, particularly in the challenging domain of brain tumor segmentation. Therefore, we strongly advocate for the cautious use of BCE loss solely in classification tasks without data imbalance, while urging the adoption of Focal loss as the preferred approach for achieving more accurate and reliable results in brain tumor segmentation endeavors. By embracing Focal loss, researchers and practitioners can unlock the full potential of deep learning techniques, leading to more effective medical image analysis and ultimately contributing to enhanced patient care and outcomes.

Disclosures

Conflicts of interest

The authors of this paper declare that there is no conflict of interest associated with the research, data analysis, and findings presented herein. All aspects of this study were conducted with utmost integrity and adherence to scientific standards.

The authors have no financial or personal relationships that could have influenced the research process, data interpretation, or the content of this paper. Additionally, they have not received any financial support or sponsorship from organizations or individuals that could be perceived as having a vested interest in the outcome of this study.

Code, Data, and Materials Availability

The dataset used in this study will be made available to the journal upon the request. Additionally,

to promote reproducibility and transparency, the complete codebase employed for data analysis and results generation will be accessible on GitHub: https://github.com/mahdishafiei/BrainTomur-Semantic-segmentatio

Ethical consideration

This research conducted entirely observational and did not involve any interventions or manipulations on human subjects or animals. The study aimed to analyze existing data sets cross sectionally, and at no point were living beings, either human or animal, subjected to any form of experimentation or interaction. I would like to clarify that the data utilized in the study did not involve any personal information and, therefore, did not breach any confidentiality or privacy issues. The research primarily focused on using data collected from the image center, which was already anonymized and de-identified before it was made available for public use. As such, no individuals were directly involved, and the research was conducted in full compliance with ethical guidelines and regulations. It is worth mentioning that the research described in the manuscript was reviewed and approved by the Institutional Review Board and ethics Committee at "Tehran University of Medical science", which is accredited to ensure the protection and welfare of human data images. The approval was obtained before the commencement of any data analysis or processing.

ACKNOWLEDGMENT

I extend my heartfelt appreciation to the team at" Bahar Imaging Center" for their invaluable assistance and support in gathering the necessary data for this research.

Their cooperation and dedication were instrumental to the success of this study. I am also deeply thankful to my esteemed professors at Tehran University of Medical Sciences for their guidance, mentorship, and encouragement throughout this academic endeavor and research. Their expertise and insightful feedback significantly enriched the quality of this research, and I am truly thankful for their contributions to my academic growth. It is important to note that this research received no external funding, and I have no financial interest in this manuscript.

REFERENCES

1. Khazaei Z, Goodarzi E, Borhaninejad V, Iranmanesh F, Mirshekarpour H, Mirzaei B, et al. The association between incidence and mortality of brain cancer and human development index (HDI): an ecological study. BMC Public Health. 2020 Nov 12;20(1):1696.

2. Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2011-2015. Neuro Oncol. 2018 Oct 1;20(suppl_4):iv1–86.

3. Al-Okaili RN, Krejza J, Wang S, Woo JH, Melhem ER. Advanced MR imaging techniques in the diagnosis of intraaxial brain tumors in adults. Radiographics. 2006 Oct;26 Suppl

1(suppl_1):S173-89.

4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data. 2017 Sep 5;4(1):170117.

5. Borole VY, Nimbhore S, Kawthekar SS. Image Processing Techniques for Brain Tumor Detection: A Review. 2015 [cited 2024 Dec 18]; Available from: https://www.semanticscholar.org/paper/Image-Processing-Techniques-for-Brain-Tumor-A-Borole-Nimbhore/3975ab9ca18c1ef9d38c2a0b39efc703e6ebcd67.

6. Thaha MM, Kumar KPM, Murugan BS, Dhanasekeran S, Vijayakarthick P, Selvi AS. Brain tumor segmentation using convolutional neural networks in MRI images. J Med Syst. 2019 Jul 24;43(9):294.

7. Liu Z, Tong L, Chen L, Jiang Z, Zhou F, Zhang Q, et al. Deep learning based brain tumor segmentation: a survey. Complex Intell Syst. 2023 Feb;9(1):1001–26.

8. Corso JJ, Sharon E, Dube S, El-Saden S, Sinha U, Yuille A. Efficient multilevel brain tumor segmentation with integrated bayesian model classification. IEEE Trans Med Imaging. 2008 May;27(5):629–40.

9. Clark MC, Hall LO, Goldgof DB, Velthuizen R, Murtagh FR, Silbiger MS. Automatic tumor segmentation using knowledge-based techniques. IEEE Trans Med Imaging. 1998 Apr;17(2):187–201.

10. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017 May 24;60(6):84–90.

11. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain tumor segmentation with Deep Neural Networks. Med Image Anal. 2017 Jan;35:18–31.

12. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2015. p. 3431–40.

13. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Lecture Notes in Computer Science. Cham: Springer International Publishing; 2015. p. 234–41. (Lecture notes in computer science).

14. Milletari F, Navab N, Ahmadi S-A. V-net: Fully convolutional Neural Networks for volumetric medical image segmentation [Internet]. arXiv [cs.CV]. 2016. Available from: https://www.semanticscholar.org/paper/V-Net%3A-Fully-Convolutional-Neural-Networks-for-

Milletar%C3%AC-Navab/50004c086ffd6a201a4b782281aaa930fbfe6ecf.

15. Ciresan D, Giusti A, Gambardella L, Schmidhuber J. Deep neural networks segment neuronal membranes in electron microscopy images. Pereira F, Burges CJ, Bottou L, Weinberger KQ, editors. Neural Inf Process Syst. 2012 Dec 3;25:2852–60.

16. Li G, Zheng Y, Cui J, Gai W, Qi M. DIM-UNet: Boosting medical image segmentation via diffusion models and information bottleneck theory mixed with MLP. Biomed Signal Process Control. 2024 May;91(106026):106026.

17. Liu Y, Wang H, Chen Z, Huangliang K, Zhang H. TransUNet+: Redesigning the skip connection to enhance features in medical image segmentation. Knowl Based Syst. 2022 Nov;256(109859):109859.

18. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH. Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Cham: Springer International Publishing; 2018. p. 287–97. (Lecture notes in computer science).

19. Liaw R, Liang E, Nishihara R, Moritz P, Gonzalez JE, Stoica I. Tune: A research platform for distributed model selection and training [Internet]. arXiv [cs.LG]. 2018. Available from: http://arxiv.org/abs/1807.05118.

20. van Beers F, Lindström A, Okafor E, Wiering M. Deep neural networks with intersection over union loss for binary image segmentation. In: Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods [Internet]. SCITEPRESS - Science and Technology Publications; 2019. Available from: https://research.rug.nl/files/87088047/ICPRAM_2019_35.pdf.

21. Kingma DP, Ba J. Adam: A method for stochastic optimization [Internet]. arXiv [cs.LG]. 2014. Available from: http://arxiv.org/abs/1412.6980.

22. Lin T-Y, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE; 2017. p. 2999–3007.

23. Smith SM. Fast robust automated brain extraction. Hum Brain Mapp. 2002 Nov;17(3):143–55.

24. Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index. Acad Radiol. 2004 Feb;11(2):178–89.

25. Smeulders AWM, Chu DM, Cucchiara R, Calderara S, Dehghan A, Shah M. Visual tracking: An experimental survey. IEEE Trans Pattern Anal Mach Intell. 2014 Jul;36(7):1442–68.

26. Tan M, Wu F, Kong D, Mao X. Automatic liver segmentation using 3D convolutional neural networks with a hybrid loss function. Med Phys. 2021 Apr;48(4):1707–19.

27. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. IEEE Trans Pattern Anal Mach Intell. 1993;15(9):850–63.

28. Pike VW. Considerations in the development of reversibly binding PET radioligands for brain imaging. Curr Med Chem. 2016;23(18):1818–69.

29. Kishida I, Nakayama H. Empirical study of easy and hard examples in CNN training [Internet]. arXiv [cs.CV]. 2019. p. 179–88. Available from: https://www.researchgate.net/publication/337808063_Empirical_Study_of_Easy_and_Hard_Examples_in_CNN_Training.

30. Wong KCL, Moradi M, Tang H, Syeda-Mahmood T. 3D segmentation with exponential logarithmic loss for highly unbalanced object sizes [Internet]. arXiv [cs.CV]. 2018. Available from: https://www.semanticscholar.org/paper/3D-Segmentation-with-Exponential-Logarithmic-Loss-Wong-Moradi/88b3a6c3900b189f02cb8a3bfae95214f90ca585.

31. Ullah F, Nadeem M, Abrar M, Al-Razgan M, Alfakih T, Amin F, et al. Brain Tumor Segmentation from MRI images using handcrafted convolutional neural network. Diagnostics (Basel). 2023 Aug 11;13(16):2650.

32. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation [Internet]. arXiv [cs.CV]. 2018. Available from: http://arxiv.org/abs/1807.10165.

33. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: Learning where to look for the pancreas [Internet]. arXiv [cs.CV]. 2018. Available from: http://arxiv.org/abs/1804.03999.

34. Huang D, Wang M, Zhang L, Li H, Ye M, Li A. Learning rich features with hybrid loss for brain tumor segmentation. BMC Med Inform Decis Mak. 2021 Jul 30;21(Suppl 2):63.

35. Salehi SSM, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: Machine Learning in Medical Imaging. Cham: Springer International Publishing; 2017. p. 379–87. (Lecture notes in computer science).

36. Bauer S, Nolte L-P, Reyes M. Segmentation of brain tumor images based on atlas-registration combined with a Markov-Random-Field lesion growth model. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE; 2011. p. 2018–21.

37. Menze BH, Van Leemput K, Lashkari D, Riklin-Raviv T, Geremia E, Alberts E, et al. A generative

probabilistic model and discriminative extensions for brain lesion segmentation--with application to tumor and stroke. IEEE Trans Med Imaging. 2016 Apr;35(4):933–46.

38. Zhang W, Li R, Deng H, Wang L, Lin W, Ji S, et al. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. Neuroimage. 2015 Mar;108:214–24.

39. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using Convolutional Neural Networks in MRI images. IEEE Trans Med Imaging. 2016 May;35(5):1240–51.

Supplement

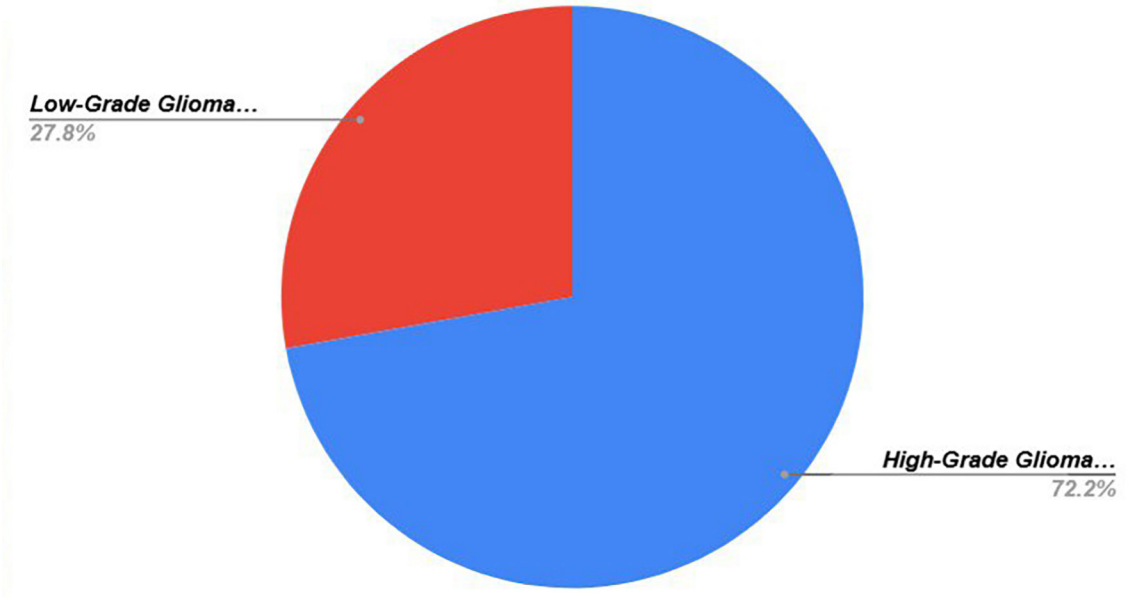

Figure S1. Classification tumors as either High-Grade Glioma (HGG) or Low-Grade Glioma (LGG)