# A quantitative comparison between Focal loss and Binary Cross-Entropy loss in Brain Tumor Auto-segmentation using U-net

1. Introduction

Brain tumors are among the most lethal cancers and cause the death of a significant number of people annually. 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 [1] 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 [1].

Magnetic resonance imaging (MRI) is the primary scan tool for detecting and diagnosing tumors since MRI provides outstanding contrast for soft tissues. In general, the function of MR imaging in the workup of Intra axial tumors can be divided into tumor diagnosis and classification, treatment planning, and post-treatment management [2]. Based on this, MRI images can play an essential role in discovering tumors and saving the life of patients. It is worth mentioning that before the scan, a particular dye called a contrast medium is given to make a more precise image.

Image segmentation is the process of partitioning a digital image into some segments. Image Segmentation is commonly used to find objects and edges in the image [3]. Also, image segmentation is the process of allocating a label to every pixel in an image in which the pixels with the same label share specific visual characteristics [4].

Brain tumor image segmentation is often done by experts to separate different parts of the brain image to identify tumors and differentiate between them and normal parts of the brain. In the medical routine, precise segmentation of brain tumors provides valuable information for treatment planning. Process of tumor localization is very intense work and highly dependent on the doctors’ experience, skills, and their slice-by-slice decisions. Early diagnosis of a cancerous tumor can increase the chance of the patient's recovery after treatment and improve the survival probability of the patient. However, qualified experts need a considerable period of time to segment the brain tumors. In addition, this process is prone to be influenced by individual opinions, and the results may vary between observers.

Many computer-aided diagnosis methods have been developed in recent years to automate the process of segmentation to save the doctors' time and provide trustworthy and accurate results while reducing the exerted efforts of experienced physicians to perform the procedures of diagnosis for every patient. Some of the special segmentation models are:

* Thresholding method: if the value for each pixel exceeds the threshold, it is identified as a tumor. [5].
* Edge-based method: differences in the intensity between edges of pixels are used as the boundaries of the tumors [6].
* Region growing method: first, a pixel input into segmentation and similar pixels will be classified as a tumor. [7].
* Atlas method: An MRI without a tumor will be used to segment MRI with the tumor volume. [8].

Among these methods, the role of Convolutional Neural Networks (CNN) in automated brain tumor segmentation is meaningful and attracted great interest in recent years. CNN was introduced in 1990, inspired by experiments performed by Hubel and Wiesel. One of the first projects to be done with CNN was Yann Lecun's famous handwritten digit identification project. He proposed the LeNet-5, which produced promising results [9]. The central brilliance of the convolution network occurred in 2012 during the ImageNet Large Scale Recognition Challenge. This was an event that caught the attention of the research community. In the 2012 ImageNet Contest, Alex Krizhevsky entered with the AlexNet, which had an excellent performance compared to others in classifying images with 16.4% error.

One of the most helpful CNN architectures is U-Net [10], which introduces skip connections between the layers in the network. The U-net architecture achieves excellent performance on various biomedical segmentation applications. However, Isensee et al. [11], who acquired top performance using a well-trained U-Net, showed that enhancing segmentation performance is not just a matter of adjusting the network architecture. The choice of the loss function, training strategy, and post-processing led had an enormous impact on the segmentation performance.

Semantic segmentation Is a cardinal task in computer vision, and it aims to map each pixel in an image to its associated labels, like a tumor, or normal tissue. Some of the earliest approaches to implementing this method in the medical field suffer from two main problems:

First, the training patches (local regions) are larger than the training samples. As a result, the running time will be significantly prolonged. Second, the accuracy of the segmentation depends on the appropriate size of the patches. Consequently, the U-net architecture has been introduced to conquer these problems [10].

As is evident in Figure (1), U-net includes two main paths: a contracting path (or encoder) and an expansive path (or decoder) [10]. The encoder is a CNN consisting of two consecutive 3×3 convolutional (unpadded) layers, separately followed by a rectified linear unit (ReLU) and a 2×2 max-pooling layer. Conversely, the decoder seeks to up-sample the resulting feature map using deconvolution layers followed by 2×2 up-convolution, a concatenation layer with the corresponding down-sampled layer from the encoder, two 3×3 convolutions, and a ReLU [12].

Finally, the up-sampled features will be conducted to a 1× 1 convolution layer to output the final segmentation map. As a result of this process, the networks can achieve precise segmentation outcomes.

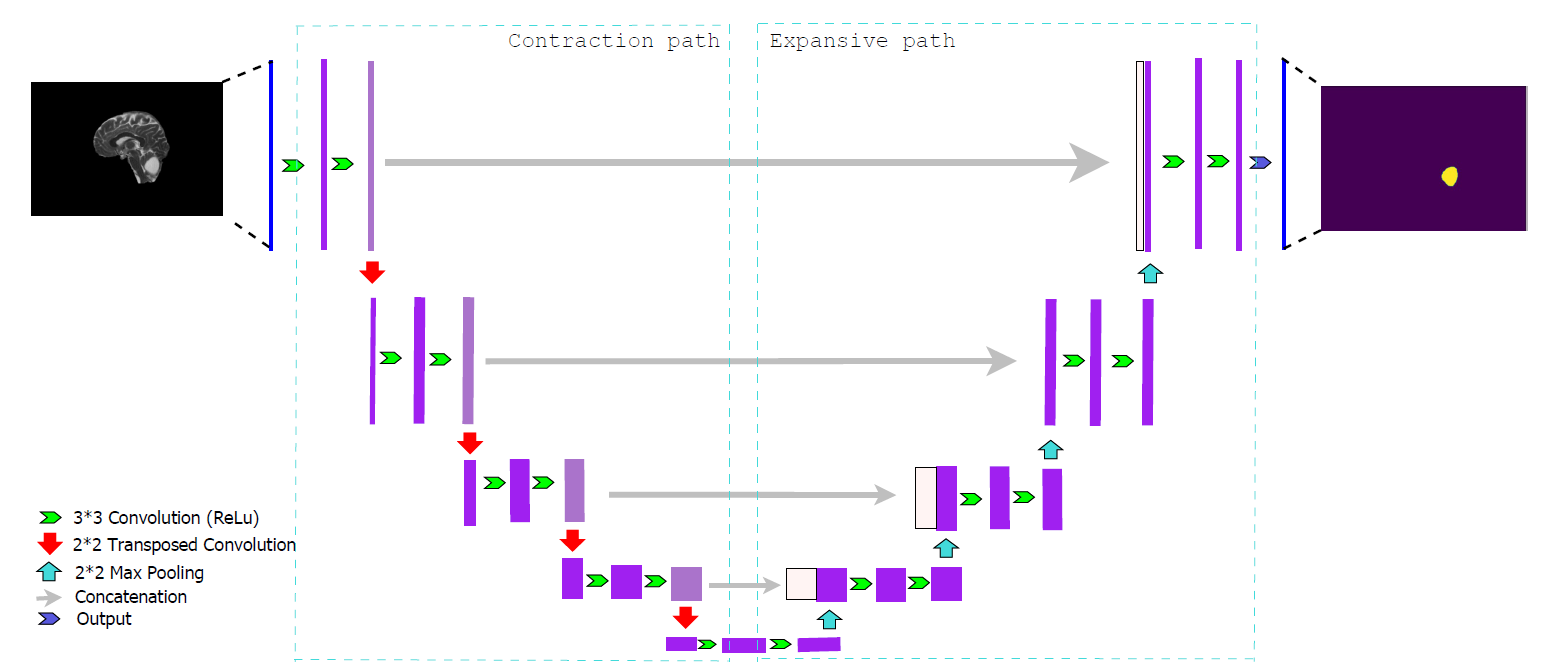


Figure 1. An example U-Net structure with 4 encoding and 4 decoding blocks.

The contributions of this study can be summarized as follows:

1. We have collected an enriched dataset consisting of high-resolution brain MRI images taken after giving a contrast medium to increase the clarity of the images, which can have a massive impact on the discovery procedure. Most of the research on operating automated brain tumor segmentation methods uses the available data on the internet. We gathered the MRI image after long negotiating with radiology experts working with these data in the imaging center. They helped us choose images with more excellent resolution and proper dimensions in all three planes to secure the diversity in the dataset, which plays a significant role in better model training. Moreover, the process of Separating the brain image from other components of the image (Skull stripping) is done image to image instead of implementing an algorithm to the whole dataset, which, overall, brought considerable accuracy.
2. In the study of segmentation tasks, more attention is often placed on enhancing the architecture of the models, while choosing the proper loss function has many benefits in terms of increasing the accuracy of the output annotation mask or discovering the exact location of the region of interest (tumor). In order to achieve this goal, we have compared the most critical evaluation metrics in a segmentation task to find out which loss function (Binary cross-entropy loss or Focal loss) can perform better in this process.
3. Method

In this study, we used two different loss functions in the standard U-net model to determine their result in the semantic segmentation of the brain tumor task.

**2.1 Loss function:**

There is an optimization problem behind the learning of every deep neural network. Generally, a neural network model will train using the stochastic gradient descent optimization method, and weights are updated using the backpropagation of the error algorithm. The gradient descent algorithm aims to change the weights by that the subsequent evaluation reduces the error. In other words, the optimization process is navigating down the gradient of error. In this context, a function used to evaluate a prospective solution is referred to as the objective function. Normally, we aim to minimize this error in deep neural networks. Accordingly, the objective function is often called a cost function or a loss function. In other word, loss function penalizes bad predictions.

Choosing the proper loss function for the semantic segmentation task is highly dependent on the complexity of the task.

**2.2 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 estimated and calculated using the cross-entropy loss. The probability output value of the cross-entropy loss function is between 0 and 1. This value increases when the predicted probability of the pixel in the image belongs to the actual class. This process can be defined as a classification algorithm. The binary cross-entropy is a particular class of cross-entropy, where the target of the prediction is 1 or 0.

The default loss function used for image segmentation is cross-entropy which mathematically is more related to accuracy [13]. In the equation below, we have provided the Log loss formula. The term pi In this formula is the probability of class 1, and (1- pi) 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.



**2.2 Focal loss:**

In medical image segmentation, there is always the possibility of imbalanced data. The imbalance in data happens because the interested region (in the particular tumor) is too smaller than the whole image. To conquer this problem, we can add weight to the cross-entropy loss formula to handle the data imbalance.

It is worth mentioning that these weights will assign to hard-to-classify examples. Hard or Misclassified examples are False negatives, and easy or Correctly Classified examples are the Ture Negatives. In this study, we strive to compare the function of BCE and Focal loss in the brain tumor image segmentation task.

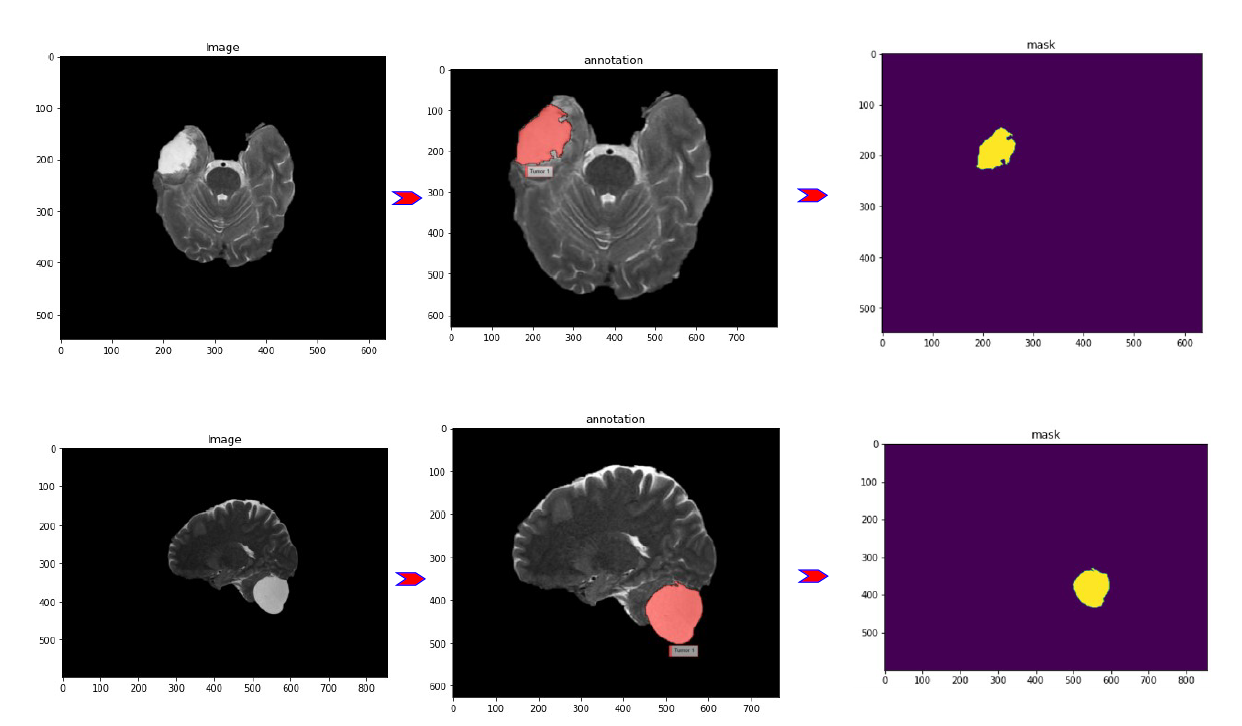


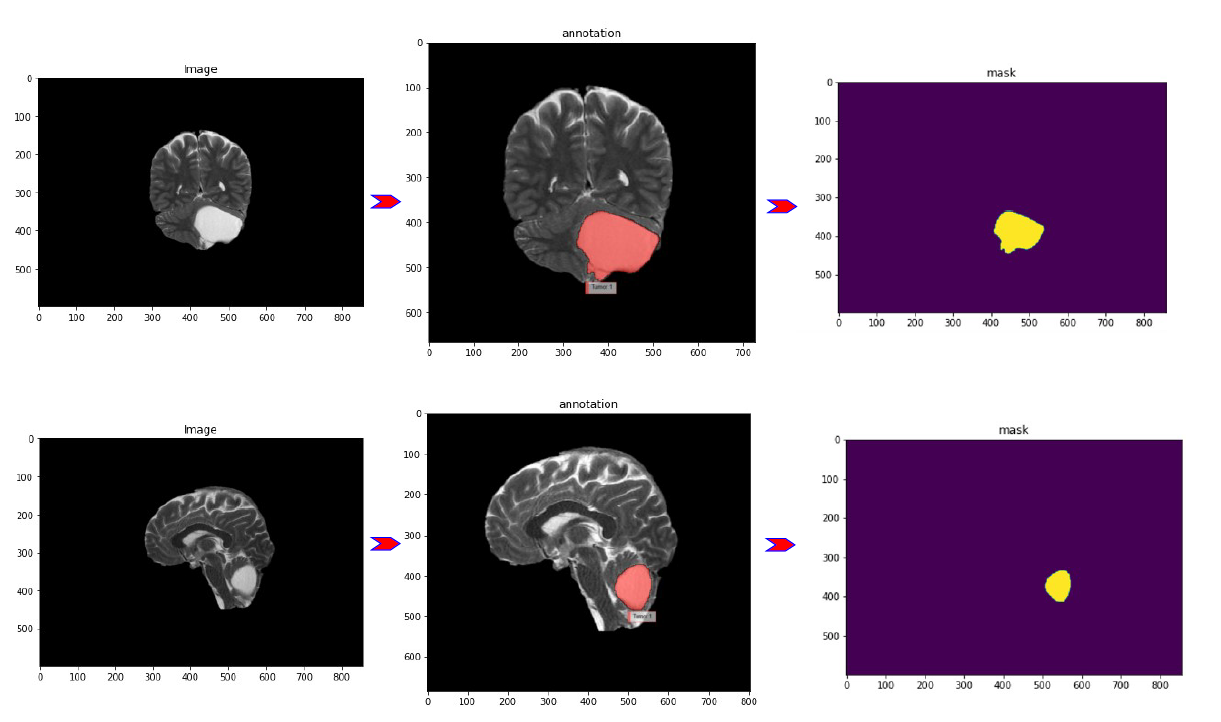
**2.4 Dataset**

The training data is a set of 314 MRI images (800x512 pixels) from the brains of 108 patients diagnosed by a doctor 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. A dedicated study is 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.

Each image comes with a corresponding fully annotated ground truth segmentation map (or mask), determining the tumor's exact location in a separate image with similar dimensions. Associating an annotation mask to each image was one of the essential tasks in this study, which was fully conducted by the diagnosis of a radiologist expert. In order to improve the precision of the segmentation task, images have been entered into some preprocessing stages.

In figure 1, we see several examples of images used in this study, along with their annotation and, finally, the output mask.





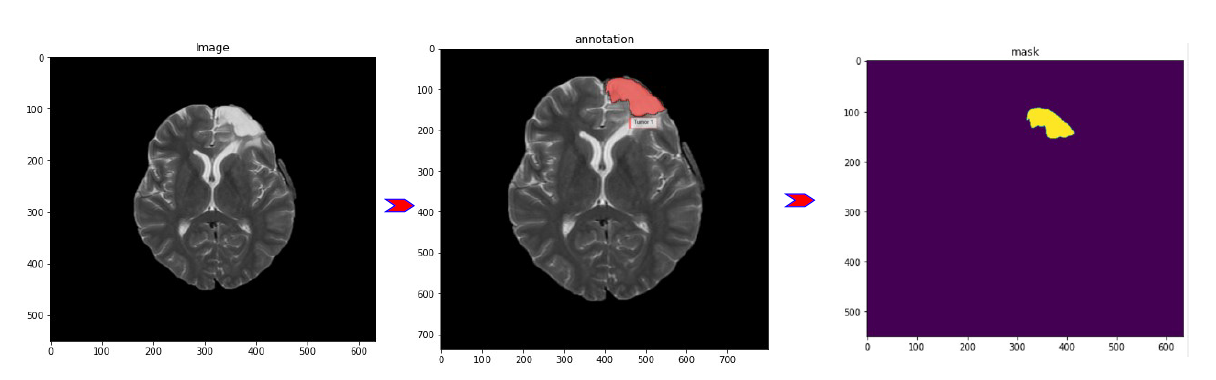


Figure 2 l Five cases from the dataset show the annotating of the tumor area and the final Mask or ground truth. The model will use these masks to detect the tumor region. Radiology experts have done these annotations.

**2.4.1 Skull Stripping**

Skull stripping is one of the initial steps in the path of finding anomalies in the brain. The images taken from an MRI machine include all parts of the head and even the neck of the patient. If these additional parts of MRI images enter the segmentation model, it may give the possibility of misdiagnosis of the model. The process of Skull stripping will isolate the brain tissue from non-brain tissue in an MRI image. Region-based binary mask extraction is used to extract the brain region in an MRI image.

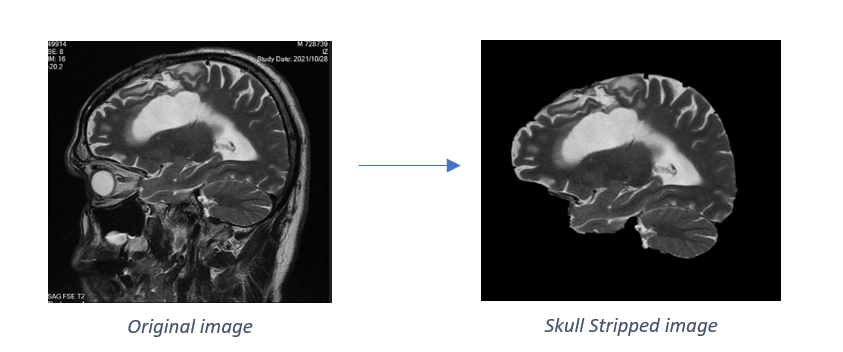
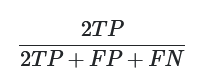


Figure 3 l an example of a raw MRI image transformed into a suitable image for segmentation task through the skull stripping process.

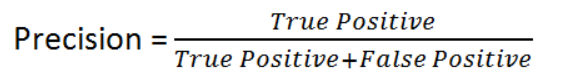
* 1. **Evaluation metrics:**
* Dice Coefficient (F1 Score)

Measures the similarity between predicted pixels and ground truth and ranges from zero to one.



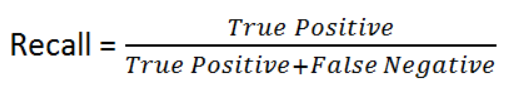
* Precision

Defines as the proportion of pixels in our segmentation model that correspond to pixels in the ground truth.



* Recall (Sensitivity)

Measures the ratio of pixels in the ground truth that were successfully detected by our segmentation model.



3.Results

We trained the U-net model on our high-resolution brain MRI images one time with BCE loss, and next time we trained the model with Focal loss. The model is implemented in python (Spyder 5.0.1) and is trained on the Cloud Tensor Processing Unit (TPU) of google collab environment, which is designed to accelerate machine learning workloads. Lastly, to test the performance of our model, fivefold cross-validation has been implemented, and the result is shown in table 1:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Binary Cross Entropy Loss | *K-folds* | *Accuracy* | *F1 (Dice Coefficient)* | *Precision* | *Sensitivity (Recall)* | *Specificity* |
| Fold-1 | 98.56 % | 50 % | 80.1 % | 71.4 % | 99.8 % |
| Fold-2 | 99.19 % | 76.1 % | 69.7 % | 83.7 % | 99.4 % |
| Fold-3 | 99.01 % | 72 % | 60 % | 90.1 % | 99.1 % |
| Fold-4 | 99.30 % | 80.4 % | 79.3 % | 81.5 % | 99.6 % |
| Fold-5 | 99.32 % | 75.9 % | 72.5 % | 79.7 % | 99.5 % |
| **5-fold Mean** | **99.07 %** | **70.8 %** | **72.32 %** | **81.28 %** | **99.4 %** |
| Focal Loss | Fold-1 | 98.08 % | 82.3 % | 84.1 % | 85.3 % | 99.9 % |
| Fold-2 | 99.16 % | 78.8 % | 86.1 % | 82.5 % | 99.8 % |
| Fold-3 | 99.34 % | 78.1 % | 78.5 % | 87 % | 99.8 % |
| Fold-4 | 99.23 % | 81 % | 88.4 % | 79.5 % | 99.8 % |
| Fold-5 | 99.32 % | 76.1 % | 69.8 % | 83.3 % | 99.4 % |
| **5-fold Mean** | **99.17 %** | **79.26 %** | **81.38 %** | **83.52 %** | **99.7 %** |

*Table 1 l Full model output in terms of all evaluation metrics separately for BCE loss and Focal loss*

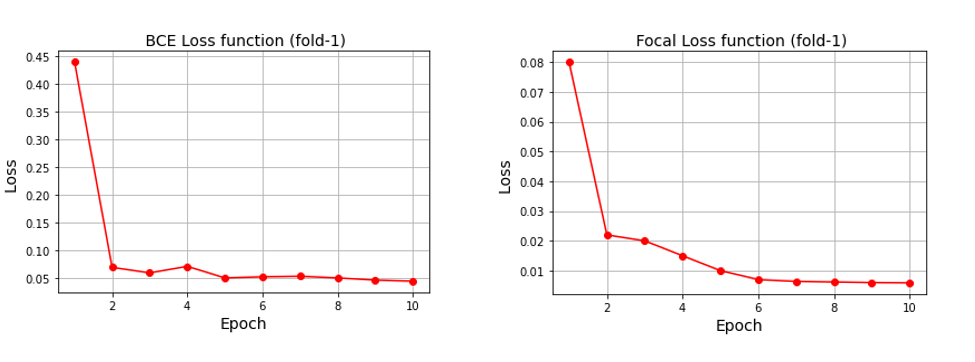


Figure 3 l the amount of loss reduction in each Epoch separately for BCE loss and Focal loss

Figure 3 shows two charts representing the loss reduction after each Epoch in the first fold of BCE loss and Focal loss separately. Obviously, the initial loss in the first Epoch in BCE loss is 0.45, but for Focal loss, it is 0.08. Also, after the tenth Epoch, the amount of loss for Focal loss is under 0.01. meanwhile, this amount for BCE loss is under 0.05.

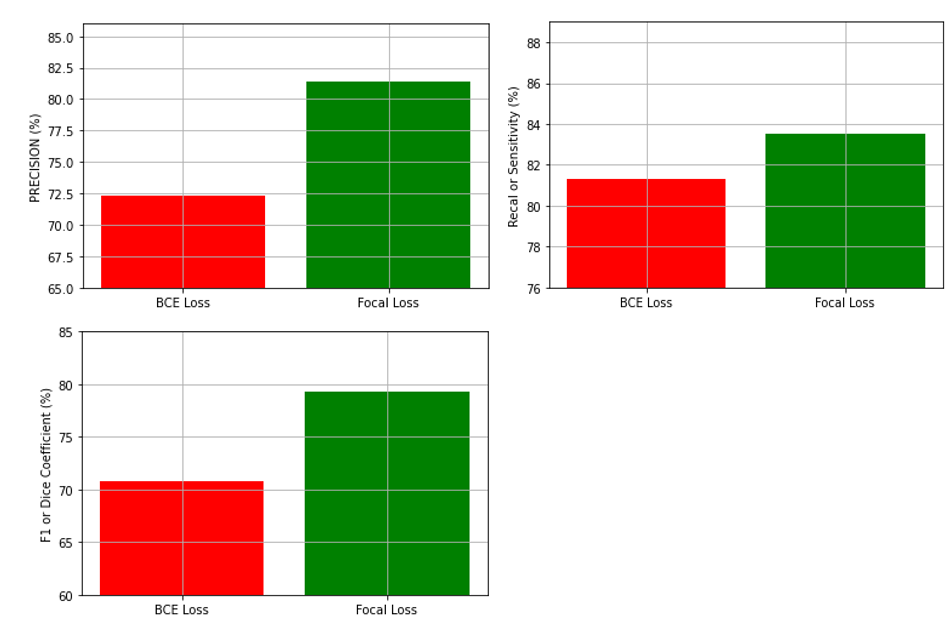
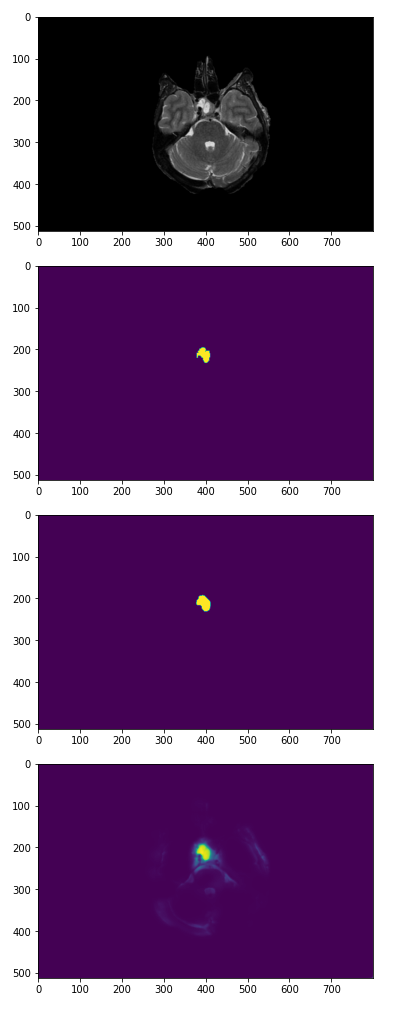
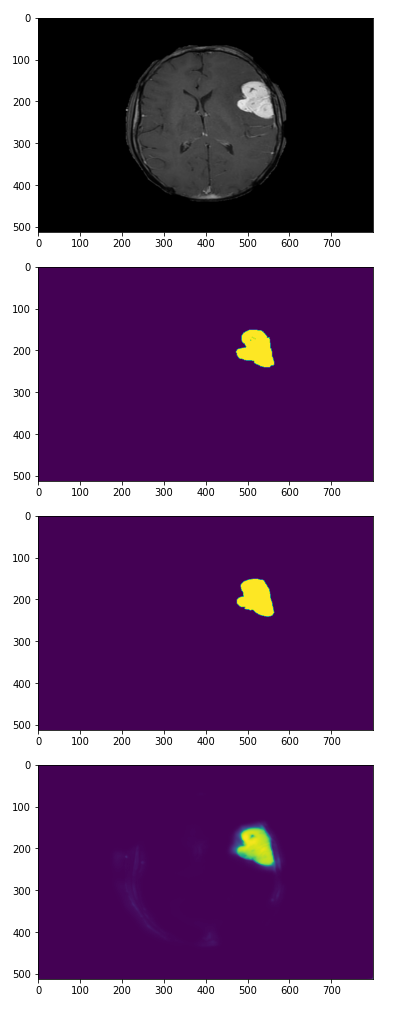


Figure 4 l the performance comparison results of the BCE loss and Focal loss in terms of Precision, F1 score and Recall

In Figure 4, we drew charts to visualize better the performance comparison of the BCE loss and Focal loss in terms of Precision, F1 score, and Recall, which are the essential evaluation metrics in the segmentation task. It can be concluded that Focal loss is outperforming 8.4% in Dice coefficient, 9 % in Precision, and 2.2 % in Recall.

Figure 5 shows the model prediction results for finding tumors on some random MRI images as an example. We can see the raw brain MRI image with the tumor in the first row. In addition, the annotation map or the corresponding mask is displayed in the second row. Our model assumes a probability value for each pixel in terms of the presence or absence of a tumor. This value will be a number between zero and one. The closer the probability is to one, the brighter the point. The images in the last line represent these probability assignments for each pixel in each image drawn by the model. Finally, the ultimate prediction of the model for locating the tumor in an image is shown in the third row. For creating this prediction, the model assumes probabilities of more than 0.5 per pixel equal to 1 and probabilities fewer than 0.5 equal to zero. Therefore, a threshold of 0.5 has been fixed, and then binary values of zero and one are assigned to each pixel. Finally, our final prediction is that pixels with a probability equal to one are tumors and plotted.



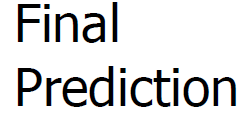


Figure 5 l Six cases of the dataset showing the image, corresponding ground truth, final prediction of tumor done by model, and each pixel's probability of being tumor.

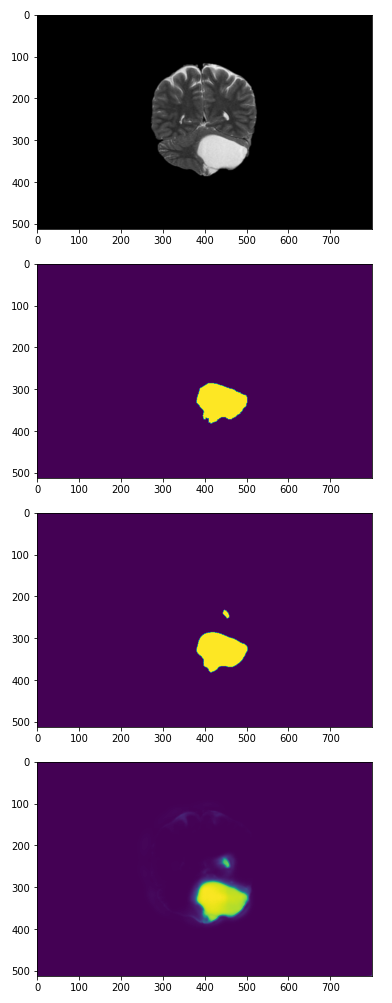
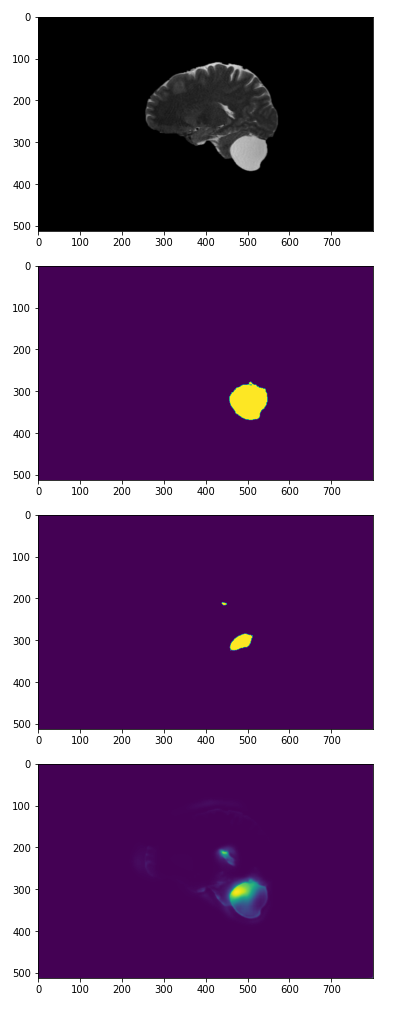
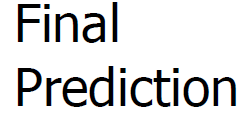
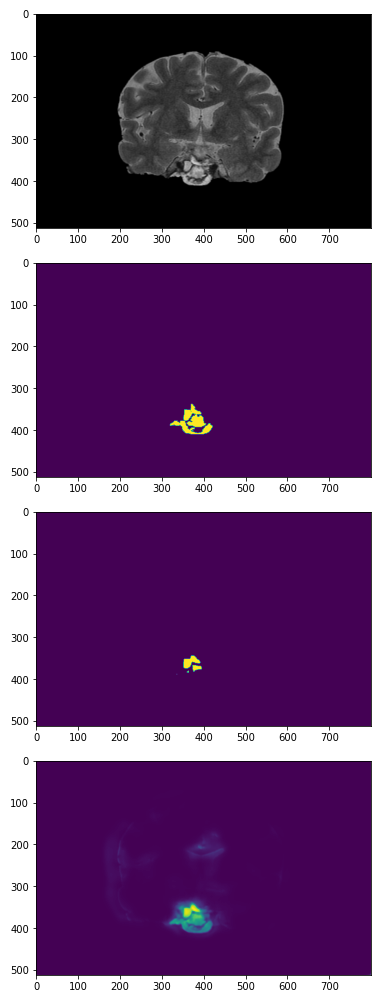
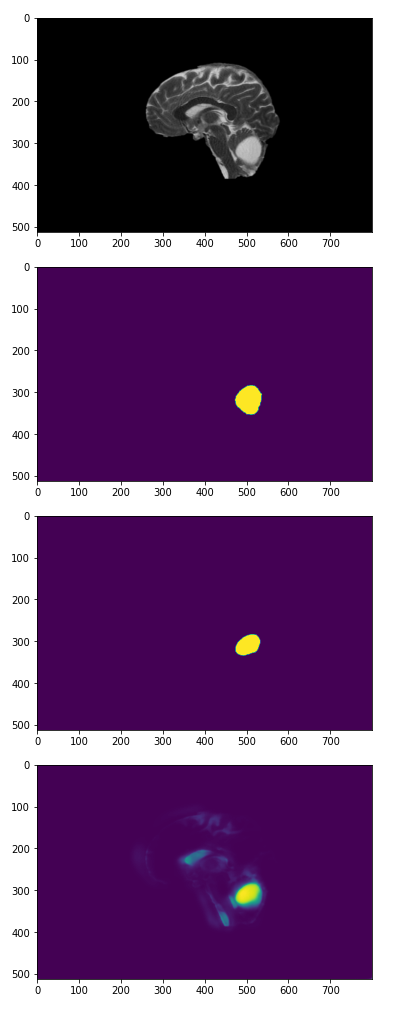


Figure 5 l Six cases of the dataset showing the image, corresponding ground truth, final prediction of tumor done by model, and each pixel's probability of being tumor.







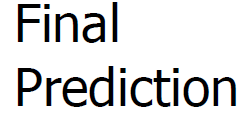




Figure 5 l Six cases of the dataset showing the image, corresponding ground truth, final prediction of tumor done by model, and each pixel's probability of being tumor.

4.Discussion and Conclusion

In this paper, we intended to evaluate an essential factor in the process of automated brain tumor segmentation, which is loss function. We have compared two widely used loss functions in terms of evaluation metrics valuable in a segmentation task, 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 worth mentioning 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 [17]. Moreover, In terms of network structure, we realized 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.

This study shows that Focal loss performs significantly better due to higher precision, F1 score, Recall, and accuracy in brain tumor auto-segmentation tasks using CNN and U-net architecture. Accordingly, we strongly recommend that use of BCE loss just be limited to classification tasks without imbalance in the data. On the other hand, we believe that focal loss is a proper choice for segmentation tasks, especially brain tumor segmentation.

Inspired by the concept of focal loss, Ken C. L. et al. [14] 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.

Daobin Huang et al. [15] 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 %, 67%, and 59% for the three tumor regions. On the other hand, our Dice score is 9% higher than their average Dice.

Mohseni Salehi et al. [16] 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.

5. References

1. 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-iv86. doi: 10.1093/neuonc/noy131. Erratum in: Neuro Oncol. 2018 Nov 17;:null. PMID: 30445539; PMCID: PMC6129949.
2. 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:S173-89. doi: 10.1148/rg.26si065513. PMID: 17050514.
3. Vipin Y. Borole, Sunil S. Nimbhore, Dr. Seema S. Kawthekar , " Image Processing Techniques for Brain Tumor Detection: A Review " , International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 4, Issue 5(2), September - October 2015 , pp. 028-032 , ISSN 2278-6856.
4. Nandita Goyal and Bharti Sharma 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022012011
5. Zaidi H, El Naqa I. PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. Eur J Nucl Med Mol Imaging. 2010 Nov;37(11):2165-87. doi: 10.1007/s00259-010-1423-3. Epub 2010 Mar 25. PMID: 20336455.
6. Aslam, Asra & Khan, Ekram & Beg, M.. (2015). Improved Edge Detection Algorithm for Brain Tumor Segmentation. Procedia Computer Science. 58. 10.1016/j.procs.2015.08.057.
7. Lin GC, Wang WJ, Kang CC, Wang CM. Multispectral MR images segmentation based on fuzzy knowledge and modified seeded region growing. Magn Reson Imaging. 2012 Feb;30(2):230-46. doi: 10.1016/j.mri.2011.09.008. Epub 2011 Nov 30. PMID: 22133286.
8. Bauer S, Seiler C, Bardyn T, Buechler P, Reyes M. Atlas-based segmentation of brain tumor images using a Markov random field-based tumor growth model and non-rigid registration. Annu Int Conf IEEE Eng Med Biol Soc. 2010;2010:4080-3. doi: 10.1109/IEMBS.2010.5627302. PMID: 21096622.
9. Zhang, Quan. "Convolutional neural networks." Proceedings of the 3rd International Conference on Electromechanical Control Technology and Transportation. 2018.
10. Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. <https://doi.org/10.1007/978-3-319-24574-4_28>
11. Isensee, F.; Kickingereder, P.;Wick,W.; Bendszus, M.; Maier-Hein, K.H. No new-net. In International MICCAI Brainlesion Workshop; Springer: Berlin, Germany, 2018; pp. 234–244.
12. Zeineldin RA, Karar ME, Coburger J, Wirtz CR, Burgert O. DeepSeg: deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images. Int J Comput Assist Radiol Surg. 2020 Jun;15(6):909-920. doi: 10.1007/s11548-020-02186-z. Epub 2020 May 5. PMID: 32372386; PMCID: PMC7303084.
13. Beers, Floris van et al. “Deep Neural Networks with Intersection over Union Loss for Binary Image Segmentation.” ICPRAM (2019).
14. Wong, Ken C. L., Mehdi Moradi, Hui Tang and Tanveer F. Syeda-Mahmood. “3D Segmentation with Exponential Logarithmic Loss for Highly Unbalanced Object Sizes.” ArXiv abs/1809.00076 (2018): n. pag.
15. 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. doi: 10.1186/s12911-021-01431-y. PMID: 34330265; PMCID: PMC8323198.
16. Salehi SSM, Erdogmus D, Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: International workshop on machine learning in medical imaging: 2017. Berlin: Springer; 2017. p. 379–87.
17. Kishida, I., Nakayama, H. (2019). Empirical Study of Easy and Hard Examples in CNN Training. In: Gedeon, T., Wong, K., Lee, M. (eds) Neural Information Processing. ICONIP 2019. Communications in Computer and Information Science, vol 1142. Springer, Cham. https://doi.org/10.1007/978-3-030-36808-1\_20