Optimising Semantic Segmentation of Tumor Core Region in Multimodal Brain MRI: A Comparative Analysis of Loss Functions

Ceena Mathews*
Prajyoti Niketan College, Kerala, India .

Abstract

Complete removal of tumor core tissues is paramount to prevent the recurrence of brain tumors. Effective automated brain tumor segmentation is challenging due to the heterogeneous nature of gliomas and the class imbalance problem that is common in brain MR images. Class imbalance in the segmentation of brain tumor subregions occurs when the tumor subregion classes have a smaller volume than the background classes representing healthy brain tissues in brain MR images. From the literature, it is evident that deep learning models are extensively used to effectively segment brain tumors. A crucial component of any deep learning model is the loss function, which optimizes the model's parameters during training. Recent studies show that region-based and compound loss functions help achieve better optimisation in dealing with the class imbalance in medical images. In this work, we explored the performance of a brain tumor segmentation framework using nested 2D U-Net model optimised with region-based and compound loss functions on the BraTS 2019 dataset. The model is evaluated using metrics such as dice score and Hausdorff distance.

Key Words: Brain Tumor; Class Imbalance; Region-based Loss; Compound Loss; Nested U-Net; BraTS 2019.

1 Introduction

Necrotic core (NCR), enhancing tumor (ET), and non-enhancing tumor (NET) regions form the tumor core of glioma which is removed surgically and treated with radiation and chemotherapy. During surgical procedures, the complete resection of tumor core tissues (ET, NET, NCR) is paramount to prevent tumor recurrence.

Automatic segmentation of tumors from brain MRI can help reduce subjective errors caused due to the manual segmentation. However, due to the heterogeneous nature of gliomas and the class imbalance problem that is common in brain MR images, effective automated brain tumor segmentation is challenging. Consequently, the prediction accuracy for tumor core subregions may decrease, potentially misleading physicians in removing tumor core tissues and raising the risk of tumor recurrence.

Class imbalance in brain MR images arises when the tumor subregion classes have a smaller volume than the background classes that represent healthy brain tissues. This can lead to potential problems when training a machine learning model. When a dataset has a smaller representation of one class, the model may not have sufficient exposure to that class, resulting in an underperforming model that cannot achieve high accuracy of prediction for the underrepresented class.

Recently, many research works are being carried out in brain tumor segmentation using deep learning [6, 7, 9, 12, 22, 24]. A crucial component of any deep learning model is the loss function, which is used to optimise the model's parameters during training. In the case of class imbalance, a standard loss function such as cross-entropy may not perform well because it is biased towards the majority class [18]. This can lead to a model that incorrectly labels the minority class as the majority class.

To overcome this issue, several loss functions have been proposed specifically for imbalanced datasets, such as region-based loss functions eg. focal loss and the Dice coefficient loss. These loss functions give more weight to the minority class during training, encouraging the model to focus more on correctly classifying the minority class. Recent studies show that region-based and compound loss functions help achieve better optimisation in dealing with class imbalance in medical images.

In this study, to enhance the segmentation accuracy of brain tumor subregions, we explored the performance of brain tumor segmentation model using nested U-Net model optimised with different region-based loss and compound loss functions. The model's performance was evaluated on the BraTS 2019 dataset using dice score and Hausdorff distance metrics.

2 Related Studies

In the literature, various studies have been conducted to compare optimisation achieved by different loss functions in dealing with class imbalance. [17] compared seven loss functions on the CVC-EndoScenestill dataset and observed that the region-based losses give better performance than the cross-entropy loss. They have used U-Net and LinkNet with VGG-16 and Densnet121 as backbones.

[10] compared fifteen losses using NBFS Skull-stripped dataset and found that the region-based losses such as focal Tversky loss and Tversky loss outperformed the other loss functions. Simple 2D U-Net model architecture for segmentation has been used for comparison. The author also observed that the binary cross-entropy loss function works well

*Prajyoti Niketan College, Kerala, India. Email: ceenamathews@prajyotiniketan.edu.in.

with balanced datasets. Additionally, the author also discovered that the Dice loss or generalised Dice loss works well with low skewed dataset (where there are more samples of the background class than the foreground class) because it helps the model to focus on the minority class, which is the object of interest.

[14] compared twenty loss functions using four datasets for liver, liver tumor, pancreas and multi-organ segmentation and concluded that the compound loss functions perform better than the region-based and distribution-based loss functions. 3D U-Net with nnU-Net V1 as backbone has been used to study the impact of the loss functions. [23] recommended the compound loss functions such as TopK loss, focal loss and focal Dice loss, which force the network training on hard samples. They also suggest that compound loss functions are the great choices to deal with complicated situations. [26] proposes a new loss function combining Dice loss and focal loss to facilitate the training of the neural model which segments organs-at-risk from head and neck CT images.

[2] uses a three-layer deep U-Net based encoder-decoder architecture for semantic segmentation. In three layer deep U-Net architecture, each layer of the encoding side includes dense modules and the decoding side uses convolutional modules. The network was trained on the BraTS 2019 dataset using soft Dice loss and focal loss function. [5] uses MCA-ResUNet for the segmentation of brain tumor MR images. The network was trained on the BraTS 2019 dataset using Dice loss.

[19] proposes the Generalised Dice (GD) overlap as a loss function for highly unbalanced segmentations. It discusses the challenges posed by class imbalance in medical image segmentation and evaluates the GD overlap against other commonly used loss functions, such as Dice loss, sensitivity-specificity loss, and weighted cross-entropy loss.

[20] proposes a combo loss which is a weighted sum of Dice loss and modified cross entropy loss. It was evaluated on PET multi-organ, ultrasound echocardiography and prostate MRI dataset. When the combo loss was included, 3D U-Net and 3D SegNet performed better.

[21] proposed the unified focal loss that generalises Dice and cross entropy-based losses for handling class imbalance. The proposed loss function was evaluated on five publicly available, class imbalanced medical imaging datasets such as CVC-ClinicDB, Digital Retinal Images for Vessel Extraction (DRIVE), Breast Ultrasound 2017 (BUS2017) etc. The literature shows that compound or region-based loss functions have consistent performance than distribution-based loss functions.

3 Materials and Methods

3.1 BraTS Dataset

The BraTS 2019 training dataset includes pre-operative multimodal MRI scans of 335 patients, of which 259 are HGG and 76 are LGG cases. This work uses 3D MR images of 150 HGG patients from the BraTS 2019 training dataset. Each patient case has four MRI sequences such as T1-weighted (T1), T1-weighted with gadolinium contrast (T1Gd), T2-weighted (T2), fluid-attenuated inversion recovery (FLAIR), and ground truth. The ground truths in these datasets are manually segmented and annotated by the experts as background (label 0), NET (NCR/NET) (label 1), ED (label 2), and ET (label 4). Label 3 is not used by the experts.

The dimension of each MRI is (240 x 240 x 155) where 240 x 240 indicates the height and width of a slice and 155 specifies the number of slices. These MRI scans were acquired with different clinical protocols and various scanners from multiple (n=19) institutions. Since the MR images are acquired using different scanners, they are of different resolution. The images are co-registered, skull-stripped, and re-sampled to $1mm^3$. The 3D images are in NIfTI format with the '.nii.gz' extension [3, 4, 15].

Each MRI sequence is significant in identifying different tumor subregions. In T1Gd, ET appears brighter whereas the subregions NET and NCR appear darker. In FLAIR images, ET, NET, and edema appear brighter.

Additionally, for evaluating the performance of the segmentation of brain tumor, three subregions are suggested by the dataset providers:

1) tumor core (TC), which includes NCR, NET and ET;

2) ET area

3) whole tumor (WT), where WT comprises of TC and edema.

3.2 Loss Function

Loss functions used in deep learning segmentation frameworks may be classified into distribution-based loss, region-based loss, and compound loss functions.

3.2.1 Distribution Loss

Distribution-based loss function is used to reduce the disproportion between two distributions. The most fundamental function in this category is cross-entropy. Binary cross-entropy, weighted cross-entropy, balanced cross-entropy, focal loss, and distance map derived loss penalty terms are some of the other distribution-based loss functions. The cross-entropy is a measure of the difference between two probability distributions. Binary cross-entropy used for binary classification is defined as in Eq.1.

Weighted binary cross-entropy (WBCE) defined in Eq. 2 is a variant of binary cross-entropy. It assigns different weights to different classes, enabling to distinguish regions of different classes and learn significant patterns in the image.

$$Loss_{bce} = \frac{-(T * log(P) + (1-T) * log(1-P))}{N} \quad (1)$$

$$Loss_{wbce} = \frac{-(T * log(P) * w + (1-T) * log(1-P))}{N} \quad (2)$$

where T indicates ground truth values, P indicates predicted values, N indicates the number of samples and w is a hyperparameter which enables a tradeoff between false positives and false negatives. In order to reduce the number of false negatives, set $w > 1$, similarly to decrease the number of false positives, set $w < 1$.

3.2.2 Region-based Loss

Region-based loss functions aim to minimise the mismatch or maximise the overlap regions between ground truth and predicted segmentation. Dice loss, Tversky loss, Focal Tversky loss are popular region-based loss functions used in this study.

Dice Loss

Dice loss is calculated using DSC (Eq. 4), the most commonly used metric for evaluating brain tumor segmentation accuracy. It is calculated as in Eq. 3 since it assigns equal weights to each class, it is only partly suitable for dealing with class imbalance.

$$Loss_{Dice} = 1 - DSC \tag{3}$$

where

$$DSC = \frac{2 * |T \cap P|}{|T| + |P|} \tag{4}$$

where T represents ground truth values, P represents predicted values, and DSC is the Dice similarity coefficient.

Tversky Loss

Tversky loss is a variant of the commonly used loss function in machine learning known as cross-entropy loss. It is used in applications where the data is imbalanced and skewed towards one class. Due to Dice loss's equal weighting of false positives and false negatives, the segmentation outcomes of a highly unbalanced class dataset frequently show high precision but a low true positive rate. The Tversky loss in Eq. 5, helps achieve a more balanced trade-off between precision and recall. It is based on the Tversky Index given in Eq. 6 where the labels are weighted with α and β parameters.

$$Loss_{tversky} = 1 - TI \tag{5}$$

where $Loss_{tversky}$ indicates tversky loss and

$$TI = \frac{\sum_{i=1}^{N} p_{ic} * g_{ic} + \varepsilon}{\sum_{i=1}^{N} p_{ic} * g_{ic} + \alpha * \sum_{i=1}^{N} p_{i\bar{c}} * g_{ic} + \beta * \sum_{i=1}^{N} p_{ic} * g_{i\bar{c}} + \varepsilon} \tag{6}$$

where p_{ic} is predicted value of the pixel i of the tumor class c and $p_{i\bar{c}}$ is the predicted value of pixel i of the non-lesion class $\bar{c}$; g_{ic} and $g_{i\bar{c}}$ represent the ground truth value of the pixel i of the tumor class c and non-tumor class $\bar{c}$, respectively. ε is a small constant to avoid division by zero error [21].

The hyperparameters α and β enables a trade-off between false positives and false negatives to achieve better recall in the case of large class imbalance. The values of α and β are chosen from the range 0 and 1, such that α+ β should be equal to 1. In this loss, in order to reduce false negatives, greater weight is assigned to α potentially impacting the true positive rate indirectly [11, 16]. When α and β are set to 0.5, the resulting loss function is equivalent to the Dice loss function.

Focal Tversky Loss

Focal Tversky loss proposed by [1] is an extension of Tversky loss with the hyperparameter γ. It is beneficial to concentrate and accurately predict the challenging classes within the region of interest. The value of the hyperparameter is chosen in such a way that there is a balance between precision and recall. The focus on hard-to-classify classes can be increased with $\gamma \leq 1$. Focal Tversky loss is defined in Eq. 7.

$$Loss_{ft} = (Loss_{tversky})^{\gamma} \tag{7}$$

where $Loss_{ft}$ specifies focal Tversky loss and γ indicates the hyperparameter whose value may range between 0 and 1.

3.2.3 Compound Loss

For better optimisation of the network model, a wider range of attributes of different loss functions can be combined. Compound losses combine multiple, independent loss functions [21]. Normally, it is formed by combining a region-based loss function and a distribution-based loss function. Combo loss [20] and Dice focal loss [26] are examples of compound loss functions.

3.3 Nested U-Net

A nested U-Net typically consists of multiple U-Net modules stacked together, forming a hierarchical structure. It tends to have a larger number of trainable parameters compared to a traditional U-Net architecture. In general, a model with more trainable parameters has a greater capacity to learn complex patterns and may be able to achieve higher accuracy on the training dataset. This is because the model has more degrees of freedom to fit the training data, and can represent more complex relationships between the input features and output targets.

The deep learning architecture used in this work to investigate the performance of the region-based loss functions is nested U-Net model since it outperforms U-Net in biomedical image segmentation [25]. However, due to the computational limitations in training the original model with the BraTS MRI dataset, we have used a lesser number of filters which thereby decreases the number of trainable parameters. However, the nested structure of the U-Net used in our study has a larger number of total trainable parameters compared to a traditional U-Net.

The nested U-Net pyramid used in the work has 5 levels. The number of convolutional blocks is dependent on the pyramid level. The top level (fifth) of the pyramid has five convolutional blocks. Each convolutional block consists of two 2D convolutional layers followed by ReLU activation and batch normalisation. L2 regulariser is applied and a dropout of 0.2 is included after two convolutional blocks to avoid overfitting. Since He normal initialisation works better with ReLU activation layers, we have used it as the kernel initialiser. We have used 32, 64, 128, 256, and 512 filters with kernel size as 3 x 3. Maxpooling with stride 2 is applied to the output of the convolutional block. Each convolutional block is preceded by a concatenation layer which concatenates the output from the previous convolutional block at the same level with the corresponding upsampled output of the lower dense block. In the output layer, softmax activation is applied. The model has 9.17 million trainable parameters.

3.4 Evaluation Metrics

We tested the performance of the model using evaluation metrics such as DSC indicated in Eq. 4 and Hausdorff distance (HD). DSC values range from 0 to 1 indicating the degree of overlap between the segmented mask and the ground truth. A DSC score of 1 indicates perfect segmentation.

3.4.1 Hausdorff Distance

HD computes the distance between the set of non-zero pixels of two images according to Eq. 8. It determines the degree of similarity between two images superimposed on one another by measuring the distance of the point of A that is farthest from any point of B and vice versa [8]. HD is one of the most informative and useful criteria because it is an indicator of the largest segmentation error [13]. The HD is measured in millimeters if the brain MRI is represented in 2D space whereas HD is measured in cubic millimeters if the brain images are represented in 3D space.

$$HD(T,P) = max(h(T,P),h(P,T)) \tag{8}$$

where $h(T,P) = \max_{t\in T} \min_{p\in P} \|t-p\|$ and $\|t-p\|$ is the euclidean distance on the points t and p of T and P respectively. T represents the ground truth pixels and P represents the predicted pixels.

4 Results and Discussion

4.1 Performance Comparison of Region-based Loss Functions

From the literature, it is apparent that region-based and compound loss functions perform better than distribution based functions. Therefore, we investigated the performance of some region-based loss functions used in the segmentation of brain tumor. The nested U-Net architecture is used for the experiment. 3D MR images from 150 HGG patients in the benchmark BraTS 2019 challenge dataset are used. The dimension of each MRI is (240 x 240 x 155) where 240 x 240 indicates the height and width of a slice and 155 specifies the number of slices.

However, due to computational and memory limitations, we extracted only the middle 90 slices from each 3D MRI sequence, yielding 2D images. We considered the middle slices from each modality since other slices may not provide much information about tumor. Additionally, each of this image is cropped to a size of 192x192 due to various factors such as computational and memory constraints. Also such a size is chosen considering the max pooling and upsampling process.

Each MRI sequence is proficient in identifying different tumor regions, for instance, the area with ET appears brighter and TC appears darker in T1Gd, whereas WT appears brighter in FLAIR images. Therefore for the effective segmentation of all tumor subregions, we combined all the four sequences of MRI. Thus, the dataset for the study contains 13500 slices each of dimension 192 x192 x 4. 60% (8100 samples) of the dataset are used for training and 20% (2700 samples) each for validation and testing. The data is then preprocessed using a simple normalisation technique to scale the pixel values in the range of 0 and 1.

With a learning rate of $1e^{-2}$, we initially trained the model using the Adam optimiser. However, the loss converged too quickly as a result of overfitting, and the predicted image was blank. It does not even look like that the model is training. As a result, we decreased the learning rate and finally chose a $3e^{-5}$ learning rate for improved training.

The batch size is a crucial hyperparameter in deep learning models. It determines the number of samples that are processed together in a single forward and backward pass during training. However, the batch size directly impacts the memory requirements of the model. Larger batch sizes require more memory to store the activations and gradients during training. If the batch size is too large, it may exceed the available GPU memory, leading to out-of-memory errors. Thus, the batch size needs to be chosen carefully based on the available resources. Due to computational and memory limitations, our network model was trained and tested with a batch size of 16.

Experiment using Dice Loss

Table 1 shows the mean DSC score obtained for the different tumor subregions using the Dice loss based nested U-Net model. It has been observed that the model when optimised with Dice loss produced a mean Dice score for TC and ET. The segmentation results show that 88% of ET and 91.19% of TC are accurately segmented. However, the ED segment has an accuracy of only 85%. As a result, the overall accuracy of the predicted WT, which comprises TC and ED, is 88.26%.

Experiment using Tversky loss

For the experiment, the values of α are selected with a difference of 0.1. The difference in the values of α as small

Table 1: DSC score for nested U-Net model using Dice loss function

Tumor subregions	Mean DSC
WT	0.8826
ET	0.8876
TC	**0.9119**

as 0.05 is not used since the result might not be substantial, ie., it may not lead to significant changes. It is possible that the change in α may have a minimal effect, especially if the dataset has a balanced distribution of false positives and false negatives.

All value combinations for the hyperparameters α and β were tried in the experiment and the DSC score achieved for the tumor subregions WT, ET and TC are depicted in Table 2. From the table, it is observed that the Dice score for all values of α and β gives comparable results for both ET and TC. However, for some values of α the model gives low Dice score for WT since the similarity score for ED is low. It is apparent from the experiment that the model may have learned to compensate for the weight given to false positives and false negatives by adjusting the weights of the convolutional layers and the connections between them. Therefore the model produces accurate segmentations without a significant difference in the values of the similarity score. Comparable results of the model for small and large values of α may be due to the combination of the complexity of the dataset and the learned patterns in the data.

Table 2: Tversky Loss : Mean DSC score for different values of α and β. WT - whole tumor, ET - enhancing tumor, TC - tumor core

α	β	Mean DSC		
		WT	ET	TC
0.1	0.9	0.8182	0.8600	0.8808
0.2	0.8	0.8705	0.9032	0.9178
0.3	0.7	0.7118	0.9006	0.9172
0.4	0.6	0.8724	0.9068	**0.9256**
0.5	0.5	**0.8812**	0.9033	0.9199
0.6	0.4	0.8808	**0.9077**	0.9223
0.7	0.3	0.8740	0.8897	0.9207
0.8	0.2	0.7219	0.8910	0.9111
0.9	0.1	0.6498	0.9005	0.9085

Experiment using Focal Tversky loss

All possible combinations of α, β, and γ are not tested due to computational constraints. For evaluating the performance of the focal Tversky loss, the values of α and β that produced better results for Tversky loss are used. Table 3 shows the experimental results for the chosen values of the hyperparameters $\alpha, \beta,$ and γ. It is apparent from the experiment that the focal Tversky loss function based model predicts ET and TC pretty well for all the experimented values of α, β and γ. It is observed that better prediction for all tumor subregions, was obtained for the values $\alpha = 0.7$, $\beta = 0.3$, $\gamma = 0.75$.

Table 3: Focal Tversky Loss : Mean DSC score for different values of α, β and γ.

α	β	γ	Mean DSC		
			WT	ET	TC
0.6	0.4	0.75	0.8912	0.9119	0.9261
0.6	0.4	0.9	0.8920	0.9130	0.9227
0.7	0.3	0.75	**0.8941**	**0.9147**	**0.9270**
0.7	0.3	0.9	0.8869	0.9121	0.9267

As shown in Table 4, among the region-based loss functions, the network model optimised using focal Tversky loss function produces better mean DSC score for small tumor structures ET, NCR and NET.

Table 4: Performance comparison of different region-based loss functions. WT - whole tumor, ET - enhancing tumor, TC - tumor tore

Loss Function	**Mean DSC**		
	WT	ET	TC
Dice	0.8826	0.8876	0.9119
Tversky (α= 0.6 and β= 0.4)	0.8808	0.9077	0.9223
Focal Tversky (α= 0.7,β= 0.3,γ= 0.75)	**0.8941**	**0.9147**	**0.9270**

Experiment using Compound Loss Function

The settings of the hyperparameters that gave better outcomes for all tumor regions were taken into consideration when

comparing the experimented loss functions. Although there are alternative α, and β values for the best WT, ET, and TC Dice scores in Tversky loss, the α, and β values chosen are 0.6 and 0.4, respectively, since they yield good results for all tumor regions than the other options. It is evident from Table 4, the Dice score obtained for TC using focal Tversky loss function based brain tumor segmentation model outperforms other region-based loss function. To avoid tumor recurrence, it is essential to resect all tumor core structures as much as possible. To achieve this objective, it is necessary to segment all tumor subregions from MRI with greater accuracy.

From the literature [14, 20, 21, 23, 26] it is observed that compound loss function based models gain better segmentation results. In a compound loss function, a distribution based loss function and a region-based loss functions are combined to optimise the parameters of a model in order to achieve better performance. The advantage of using multiple loss functions is that it can lead to better generalisation and more robustness as the model is trained on multiple criteria at the same time.

Therefore for better optimisation of the segmentation framework, we use a compound loss function that is formed by adding focal Tversky loss and WBCE loss (Eq. 2) functions as shown in Eq. 9.

$$Loss = Loss_{ft} + Loss_{wbce} \tag{9}$$

WBCE loss assigns more importance to specific classes by applying weights to them and is hence very useful to handle imbalanced datasets. Focal Tversky loss focuses on false positives and false negatives that are more difficult to predict. It assigns more weight to difficult instances and reduces the impact of easy instances. Thus by combining these two loss functions, the resulting loss function will assign more weight to difficult instances through the focal loss component and assign more importance to specific classes through the WBCE loss. This can lead to a more accurate model that can handle imbalanced classes and difficult predictions.

We investigated the performance of the nested 2D U-Net model optimised using the compound loss function on the same preprocessed dataset. The model was trained and tested with a learning rate of $3e^{-5}$ using Adam as optimiser. We experimented with the same hyperparameter values (α= 0.7, γ= 0.75) of the focal Tversky loss function which derived optimal results. The value of w used for experiment is two. The segmentation results of the experiment are evaluated using DSC and HD metrics.

Table 5 shows the performance comparison of nested U-Net model using the compound loss against WBCE and region-based loss functions. The table shows that the mean DSC for TC which comprises the tumor core tissues, outperforms other loss functions. Fig. 1 shows the segmentation results in 5 patients achieved using nested U-Net model optimised with novel compound loss function.

Fig. 2 and the Table 5 shows that the mean Dice score produced for the WT using the compound loss is slightly better than that produced by the weighted binary-cross entropy loss function whereas the mean Dice score obtained for TC which comprises the tumor core tissues outperforms the other loss functions.

However, the compound loss-based model achieved comparable DSC score for ET compared to that obtained using focal Tversky loss function. It is clear from the illustrated chart

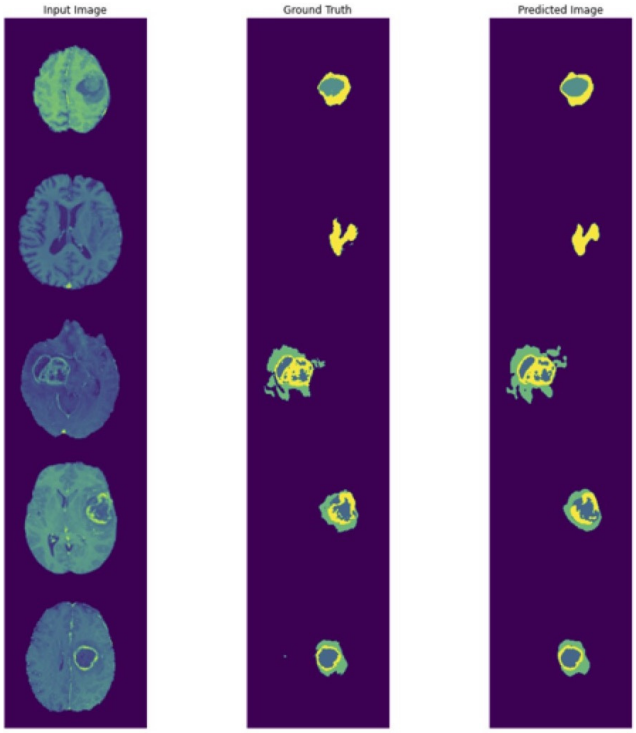

Figure 1: Segmentation results of nested U-Net trained using Compound loss

Table 5: Performance comparison of compound loss against other loss functions using nested 2D U-Net model. WT - whole tumor, ET - enhancing tumor, TC - tumor core

Loss Function	**Mean DSC**		
	WT	ET	TC
Dice	0.8826	0.8876	0.9119
Tversky (α=0.6 and β=0.4)	0.8808	0.9077	0.9223
Focal Tversky (α=0.7,β=0.3,γ=0.75)	0.8941	**0.9147**	0.9270
Weighted BCE	0.8989	0.8750	0.9182
Compound	**0.8994**	0.9140	**0.9311**

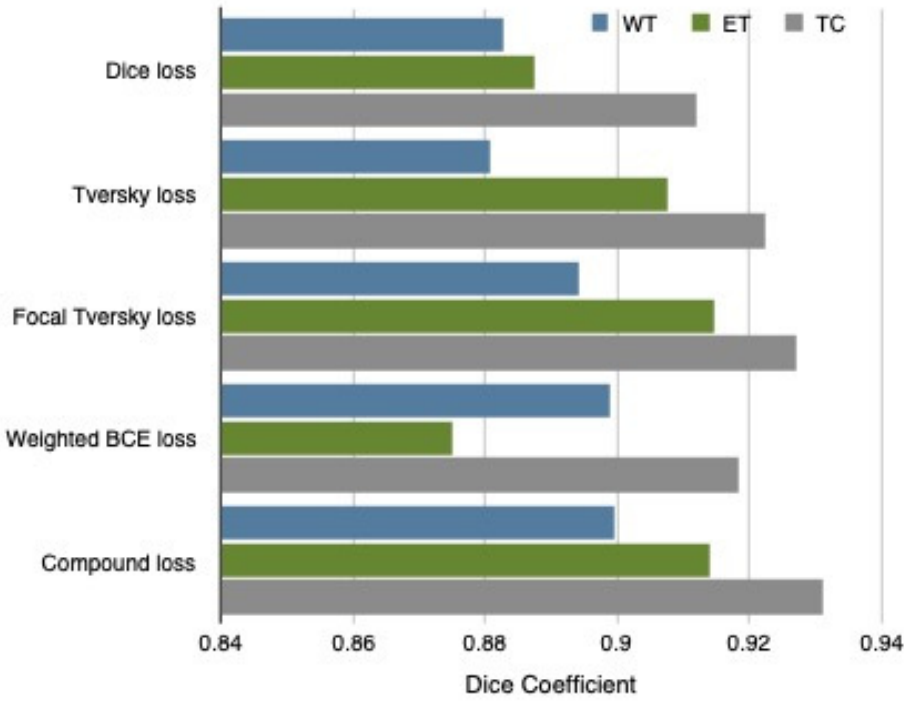

Figure 2: Graphical illustration of the prediction analysis of nested U-Net trained using compound loss and other loss functions for the tumor subregions WT, ET and TC using the metric DSC. WT - whole tumor, ET -enhancing tumor, TC - tumor core

that the nested U-Net model optimises well with the compound loss function and outperforms other loss functions in terms of TC and WT. This is due to the inclusion of the hyperparameter w because it gives weightage to the minority class (tumor tissues). The values of w and α are chosen to increase TPR, and $\gamma < 1$ is chosen such that it catalyses the model to concentrate on foreground pixels representing tumor in MRI.

We have also tested the performance of the compound loss function and other loss functions using the 2D U-Net model and the results are shown in Table 6. We have used the same hyperparameters which showed better performance with the nested U-Net model, for investigating the performance of the compound loss function with the 2D U-Net model. From the Tables 5 and 6, it is evident that the compound loss function gives better prediction for tumor core tissues (ET and TC) than the other loss functions. Table 7 shows the performance of the compound loss function against the state-of-the-art methods using other loss functions to deal with the class imbalance problem.

However, selecting optimal values for parameters such as w, α, β, and γ is indeed a great challenge due to

1) determining the correct balance between false positives and false negatives is often challenging, as it may require trade-offs based on subjective considerations.

2) the need to perform multiple experiments, varying the values of w, α, β, and γ, and evaluate the performance using different metrics. This process is time-consuming and computationally expensive.

5 Conclusions

Accurate diagnosis of the tumor core structures such as ET, NET, and NCR is critical for the maximum removal of those structures through surgery. From the literature, it is apparent that the segmentation of glioma subregions derives less segmentation accuracy for these core tissues, primarily because of the infiltrating nature of glioma and the class imbalance problem in brain MR images. Loss functions play a vital role in optimising the performance of a deep learning model and also in dealing with the class imbalance problem seen dominantly in medical images. The literature indicates that compound or region-based loss functions generally perform more consistently than distribution-based loss functions. Therefore, we investigated and compared the performance of the nested U-Net model using popular region-based loss functions and a compound loss function in the prediction of different tumor subregions. DSC score and HD metrics were used to evaluate the segmentation model's performance. Models trained with the compound loss function outperform those using region-based loss functions like Dice loss, Tversky loss, and focal Tversky loss in predicting all tumor core subregions, including enhancing tumor (ET), tumor core (TC), and whole tumor (WT). The nested U-Net model optimised with the compound loss function surpasses state-of-the-art methods that use other loss functions to address the class imbalance problem.

The enhancing tumor regions may have more heterogeneous

Table 6: Performance comparison of 2D U-Net model trained using compound loss and other loss functions using metrics DSC and HD

Loss	Mean DSC			Mean HD		
	WT	ET	TC	WT	ET	TC
Dice Loss	0.8915	0.8852	0.9165	8.1	5.9	4.6
Tversky Loss (α=0.6)	0.8783	0.8885	0.9127	9.2	5.5	4.4
Focal Tversky Loss (α=0.7 and γ=0.75)	0.8757	0.8874	0.9144	8.05	6.0	4.7
WBCE (w=2)	**0.9129**	0.8963	0.9275	**6.03**	5.3	4.2
Compound loss	0.9041	**0.9127**	**0.9299**	6.9	**4.8**	**4.09**

Table 7: Performance comparison of the compound loss function with state-of-the-art 2D segmentation frameworks using other loss functions on BraTS 2019 dataset

Method	Loss Function	DSC		
		WT	ET	TC
[2]	Dice Loss	0.89	0.74	0.85
[2]	Focal Loss	**0.92**	0.79	0.90
[5]	Dice Loss	0.849	0.784	0.865
Nested 2D U-Net	Compound Loss	0.899	**0.914**	**0.931**

appearances, and the boundaries between the tumor and the surrounding tissue can be less clear, making it more challenging to accurately segment enhancing tumor areas. By integrating attention gates with encoder-decoder architecture, the segmentation model can better emphasise important regions of the image while reducing the influence of irrelevant regions.

References

[1] Nabila Abraham and Naimul Mefraz Khan. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 683–687. IEEE, 2019.

[2] Rupal R Agravat and Mehul S Raval. Brain tumor segmentation and survival prediction. In *International MICCAI Brainlesion Workshop*, pages 338–348. Springer, 2019.

[3] S Bakas et al. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. nature sci. data 4, 170117 (2017), 2017.

[4] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The cancer imaging archive*, 286, 2017.

[5] Tianyi Cao, Guanglei Wang, Lili Ren, Yan Li, and Hongrui Wang. Brain tumor magnetic resonance image segmentation by a multiscale contextual attention module combined with a deep residual unet (mca-resunet). *Physics in Medicine & Biology*, 67(9):095007, 2022.

[6] Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In *annual conference on medical image understanding and analysis*, pages 506–517. Springer, 2017.

[7] Farnaz Hoseini, Asadollah Shahbahrami, and Peyman Bayat. An efficient implementation of deep convolutional neural networks for mri segmentation. *Journal of digital imaging*, 31:738–747, 2018.

[8] Daniel P Huttenlocher, Gregory A. Klanderman, and William J Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, 15(9):850–863, 1993.

[9] Fabian Isensee, Paul F Jäger, Peter M Full, Philipp Vollmuth, and Klaus H Maier-Hein. nnu-net for brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6*, pages 118–132. Springer, 2021.

[10] Shruti Jadon. A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pages 1–7. IEEE, 2020.

[11] Uday Kamal, Thamidul Islam Tonmoy, Sowmitra Das, and Md Kamrul Hasan. Automatic traffic sign detection and recognition using segu-net and a modified tversky loss function with l1-constraint. *IEEE Transactions on Intelligent Transportation Systems*, 21(4):1467–1479, 2019.

[12] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.

[13] Davood Karimi and Septimiu E Salcudean. Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Transactions on medical imaging*, 39(2):499–513, 2019.

[14] Jun Ma, Jianan Chen, Matthew Ng, Rui Huang, Yu Li, Chen Li, Xiaoping Yang, and Anne L Martel. Loss odyssey in medical image segmentation. *Medical Image Analysis*, 71:102035, 2021.

[15] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

[16] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *International workshop on machine learning in medical imaging*, pages 379–387. Springer, 2017.

[17] Luisa F Sánchez-Peralta, Artzai Picón, Juan Antonio Antequera-Barroso, Juan Francisco Ortega-Morán, Francisco M Sánchez-Margallo, and J Blas Pagador. Eigenloss: combined pca-based loss function for polyp segmentation. *Mathematics*, 8(8):1316, 2020.

[18] Boris Shirokikh, Alexey Shevtsov, Anvar Kurmukov, Alexandra Dalechina, Egor Krivov, Valery Kostjuchenko, Andrey Golanov, and Mikhail Belyaev. Universal loss reweighting to balance lesion size inequality in 3d medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23*, pages 523–532. Springer, 2020.

[19] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 240–248. Springer, 2017.

[20] Saeid Asgari Taghanaki, Yefeng Zheng, S Kevin Zhou, Bogdan Georgescu, Puneet Sharma, Daguang Xu, Dorin Comaniciu, and Ghassan Hamarneh. Combo loss: Handling input and output imbalance in multi-organ segmentation. *Computerized Medical Imaging and Graphics*, 75:24–33, 2019.

[21] Michael Yeung, Evis Sala, Carola-Bibiane Schönlieb, and Leonardo Rundo. Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, 95:102026, 2022.

[22] Dingwen Zhang, Guohai Huang, Qiang Zhang, Jungong Han, Junwei Han, and Yizhou Yu. Cross-modality deep feature learning for brain tumor segmentation. *Pattern Recognition*, 110:107562, 2021.

[23] Yue Zhang, Shijie Liu, Chunlai Li, and Jianyu Wang. Rethinking the dice loss for deep learning lesion segmentation in medical images. *Journal of Shanghai Jiaotong University (Science)*, 26:93–102, 2021.

[24] Xiaomei Zhao, Yihong Wu, Guidong Song, Zhenye Li, Yazhuo Zhang, and Yong Fan. 3d brain tumor segmentation through integrating multiple 2d fcnns. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3*, pages 191–203. Springer, 2018.

[25] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018.

[26] Wentao Zhu, Yufang Huang, Liang Zeng, Xuming Chen, Yong Liu, Zhen Qian, Nan Du, Wei Fan, and Xiaohui Xie. Anatomynet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical physics*, 46(2):576–589, 2019.

Authors

Ceena Mathews is an Associate Professor in the Department of Computer Science, Prajyoti Niketan College, University of Calicut, Kerala. She obtained her Ph.D. in Computer Science from Mahatma Gandhi University, Kottayam, Kerala in 2023. Her research interests include machine learning, deep learning, medical image analysis.