Journal of Technology Research (JTR)
Volume 3, Special Issue, (2025), pp 620-626 ISSN 3005-639X
The 7th Conference of Engineering Sciences and Technology
(CEST-2025)

Accurate Segmentation of Imbalanced Medical Images: A Comparative Study

Amaal A, Oshah[1] , Ahmed E, Rgibi[*1] ,Amany Alarbish[1], Hesham H. Amin[2]

[1]Department of Computer Engineering and Information Technology, Sabratha University, Faculty of Engineering, Sabratha, Libya.

[2]Department of Electrical Engineering, Faculty of Engineering, Sohag University, Sohag, Egypt.

*Corresponding author email: hhamin@eng.sohag.edu.eg

Received: 16-10-2025 | Accepted: 02-12-2025 | Available online: 25-12-2025 | DOI:10.26629/jtr.2025.57

ABSTRACT

Medical image segmentation is a fundamental task for accurate automated diagnosis, treatment planning, and clinical decision-making. This study presents a comparative evaluation of the LeViT-UNet model a convolutional encoder decoder network enhanced with transformer blocks on imbalanced computed tomography (CT) datasets. Two loss functions were investigated: the traditional Focal Loss and the composite Focal-Tversky Loss. The model was trained and validated on annotated CT slices exhibiting high class imbalance to assess segmentation accuracy and convergence stability. Experimental results reveal that training with Focal Loss enables faster convergence and achieves higher Dice and Jaccard scores during early epochs by emphasizing challenging samples. In contrast, the Focal-Tversky Loss achieves a better trade-off between sensitivity and specificity, leading to improved stability and generalization across imbalanced data. These findings underscore the importance of selecting task-specific loss functions for medical image segmentation and demonstrate that integrating LeViT-UNet with Focal-Tversky Loss provides a robust and consistent framework suitable for clinical applications demanding high precision.

Keywords: Medical image Segmentation, Focal loss function, Focal-Tversky Loss, LeViT-UNet model.

تحقيق تجزئة دقيقة للصور الطبية غير المتوازنة: دراسة مقارنة

آمال ابوعجيلة أوشاح[1]، أحمد العجيلي الرجيبي[1]، اماني الأربش[1]، هشام حامد أمين[2]

[1]قسم هندسة الحاسوب وتقنية المعلومات، جامعة صبراتة، كلية الهندسة، صبراتة، ليبيا.

[2]قسم الهندسة الكهربائية، كلية الهندسة، جامعة سوهاج، سوهاج، مصر.

ملخـــــــص البحـــــــث.

أن تجزئة الصـــور الطبية مهمة جوهرية في المجال الطبي، إذ تعتمد بشـــكل متزايد على تقنيات التعلم العميق، لكنها تواجه تحديات مستمرة مثل اختلال توازن الفئات، حيث تشـغل الآفات مسـاحة محدودة مقارنة بالخلفية. تعتمد جودة التجزئة بدرجة كبيرة على اختيار دوال الخســارة المناســبة القادرة على معالجة الاختلال بفعالية. تهدف هذه الورقة إلى مقارنة أداء نموذج LeViT–UNet، وهو نموذج هجين يجمع بين الشـبكات التلافيفية وكتل المحولات، باسـتخدام دالتي الخسـارة Focal Loss و Focal–Tversky Loss لمعالجة اختلال توازن الفئات. تم تدريب النموذج وتقييمه على مجموعة بيانات للأشـــعة المقطعية لتجويف البطن المعروفة بعدم توازن فئاتها، وتم قياس دقة التجزئة باســتخدام مؤشــري Dice و Jaccard إلى جانب تحليل اســتقرار التدريب. أظهرت النتائج أن اســتخدام دالة

Focal Loss يســرّع من تقارب النموذج ويحقق أداءً جيدًا في المراحل الأولى من التدريب من خلال التركيز على العينات الصــعبة، بينما توفر دالة Focal–Tversky Loss توازنًا أفضل بين الحساسية والنوعية، مما يعزز استقرار النموذج ويحسّن قدرته على التعميم في البيانات غير المتوازنة. تُبرز النتائج أهمية اختيار دوال خسارة مناسبة لمتطلبات المهمة لتحسين جودة تجزئة الصور الطبية، وتؤكد أن دمج LeViT–UNet مع Focal–Tversky Loss يقدّم إطارًا موثوقًا وفعّالًا لتطبيقات طبية تتطلب دقة عالية.

الكلمات الدالة: تجزئة الصور الطبية، دالة خسارة فوكل، دالة خسارة فوكل-تفِرسكي، نموذج LeViT-UNet.

1. INTRODUCTION

Medical The ability to precisely delineate anatomical structures and pathological regions from medical images is a cornerstone of contemporary clinical practice. This process of medical image segmentation provides the foundational data for accurate diagnosis, tailored treatment planning, and informed clinical decision-making [1]. Despite its importance, achieving robust segmentation remains a formidable challenge, particularly when dealing with imbalanced datasets where clinically critical structures—such as small tumors or fine vessels—occupy only a minute portion of the total image area. This imbalance often causes deep learning models to develop a bias toward the dominant class, ultimately compromising their utility for precise analysis [2]. Addressing this vulnerability has become a central focus in medical AI, driving innovation in both model architectures and training paradigms [3]. In response, the field has witnessed a surge of advanced neural network designs. Among the most promising are hybrid models like LeViT-UNet, which integrate lightweight transformer modules into the classic U-Net architecture [4]. By combining the convolutional neural network's (CNN) proficiency at extracting local features with the transformer's strength in modeling long-range dependencies, these models capture a richer spectrum of contextual information. This synergy has proven especially valuable for complex segmentation tasks, with LeViT-UNet itself demonstrating notable efficacy on imbalanced datasets [3, 4]. However, even the most sophisticated architecture is only as effective as the function used to train it. The choice of loss function is therefore paramount. Traditional loss functions, including those based on the Dice and Jaccard indices, often fall short in imbalanced settings because they treat all pixels equally, allowing the gradient from the vast background to overwhelm the signal from small foreground structures [2]. This limitation has spurred the development of more nuanced alternatives. Loss functions such as Focal Loss and Focal-Tversky Loss are specifically engineered to counteract class imbalance by dynamically adjusting the learning focus. They apply greater weight to hard-to-classify examples and, in the case of Focal-Tversky, can strategically penalize false negatives, thereby enhancing a model's sensitivity and specificity for minority classes [1, 2]. Their potential to significantly improve segmentation performance in challenging contexts is now increasingly recognized. While LeViT-UNet presents a powerful architecture and these advanced loss functions offer compelling theoretical benefits, a clear, empirical evaluation of their combined efficacy is needed. How does the integration of Focal Loss or Focal-Tversky Loss specifically influence the convergence behavior and final segmentation accuracy of LeViT-UNet on imbalanced medical data? This study seeks to answer that question. We present a comprehensive performance evaluation of the LeViT-UNet model applied to imbalanced CT image segmentation, directly comparing the impacts of the Focal and Focal-Tversky loss functions. Through this analysis, we aim to identify optimal training configurations and provide insights that contribute to the development of more reliable and robust automated diagnostic systems.

2. MATERIALS AND METHODS

2.1. Model Selection

This paper employs the LeViT-UNet [5] hybrid model as the foundation due to its capability to combine the traditional UNet architecture with transformer blocks effectively. This integration enables the model to capture global contextual features through transformers, while retaining high-resolution spatial information from UNet. Additionally, the model's skip connections facilitate more precise segmentation by fusing low-level features with the global context, which is particularly advantageous for medical image segmentation tasks. To enhance model performance, the focal loss function is utilized because it counteracts class imbalance by reducing the loss weight assigned to easily classified samples and concentrating more on difficult or misclassified instances. This strategy significantly improves segmentation quality, especially for small or underrepresented anatomical structures in medical images.

2.2. Dataset

Two datasets his study leverages two publicly available medical imaging datasets to evaluate the model's performance under class imbalance. The first is the Synapse multi-organ segmentation dataset [6], which comprises 30 abdominal CT scans. From these scans, a total of 3,779 axial contrast-enhanced clinical CT images were extracted for analysis. The dataset is partitioned into 18 cases for training and 12 for validation. Its annotation covers eight distinct abdominal organs: the aorta, gallbladder, spleen, left kidney, right kidney, liver, pancreas, and stomach, providing a benchmark for multi-class organ segmentation. The second dataset is the Automated Cardiac Diagnostic Challenge (ACDC) [6], collected from cine MRIs of 150 patients. It includes 100 fully annotated volumes, which we divided into 80 training and 20 validation samples, with an additional 50 volumes reserved for evaluation. This dataset presents a distinct and challenging multi-label segmentation scenario. A significant challenge arises from the pronounced anatomical overlap between the stomach, large intestine, and small intestine. This overlap not only creates a natural class imbalance but also demands that the model learn to make fine-grained distinctions between multiple adjacent target regions and the background within the same image area.

2.3 Loss Functions

Loss functions have a critical role in optimizing deep learning models and influencing their convergence during training. They are particularly important in addressing challenges such as class imbalance in training datasets, which forms the core of this study [7]. Among many loss functions designed to tackle these issues, the Focal Loss [8] stands out for its ability to focus training on hard-to-classify samples by down-weighting easy examples. Meanwhile, the hybrid Focal-Tversky Loss [2]. combines the advantages of Focal Loss with the Tversky loss, providing enhanced flexibility to balance false positives and false negatives, which is essential for segmenting imbalanced medical images.

Focal loss: is a form of binary cross-entropy loss that addresses the class imbalance problem with standard cross-entropy loss by reducing the contribution of positive samples [8].

$$(pt) = -\alpha(1 - pt)^{\gamma} \log(pt)$$

where, $\gamma > 0$ and when $\gamma = 1$ Focal Loss works like Cross-Entropy loss function, and α ranges from [0,1] that can be treated as a hyperparameter.

Focal-Tversky loss: this function is inspired by the Focal loss adaptation of the cross-entropy loss, the Focal Tversky loss adapts the Tversky loss by applying a focal parameter.

$$L_{FT} = \sum_{C=0}^{C-1} (L_T^C)^{1/\gamma}$$

where L_T^C represents the Tversky loss of class c. Focal Tversky loss is identical to Tversky loss for $\gamma = 1$.

[1,3] for γ that makes the model focus on misclassified pixels. However, when the training is close to convergence Focal Tversky loss becomes suppressed and prevents the model from pushing for complete convergence.

3. RESULTS AND DISCUSSION

3.1 Implementation Details

The experiments are conducted using Python 3.7.10, PyTorch 1.9.1, and Linux 5.15.154+-x86_64. The optimizer used is Adam, which has a learning rate of 2e-3. All models are trained on a Tesla P100-PCIE GPU with 16GB memory. The input resolution of images is 224x224, with a batch size of 8 for training and 16 for validation. The models used transformer backbones pre-trained on ImageNet-1k, and training is conducted for 30 epochs on the Synapse and ACDC datasets.

3.2 Results

To emphasize the pivotal role of loss functions in enhancing medical image segmentation tasks, two scenarios were assessed: one employing the focal loss function, and the other operating with focal-Tversky loss. The results of this comparison, summarized in Table (1), demonstrate the segmentation accuracy and model robustness, particularly in the presence of imbalanced datasets.

The performance comparison model with different loss functions is illustrated in Fig.1

Table 1. Performance Metrics of LiVeT-UNet model with modified models.

Metries	LiVeT-UNet model	LiVeT-UNet model With Focal loss	LiVeT-UNet model With Focal-Tversky loss
Train Loss	0.0705	0.06794	0.1181
Valid Loss	0.02467	0.14535	0.2372
Valid Dice	0.78214	0.77675	0.7820
Valid Jaccard	0.71328	0.71341	0.7211
Best Dice	07853	**0.79954**	**0.7910**
Best Jaccard	**0.7310**	**0.73653**	0.7310
Best Epoch	29	23	28

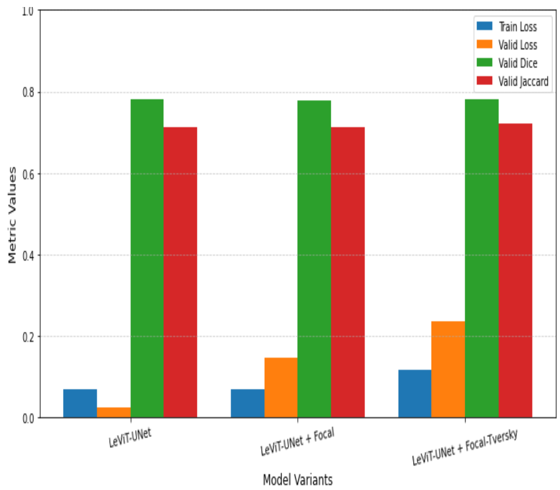

Fig1. Performance comparison for models.

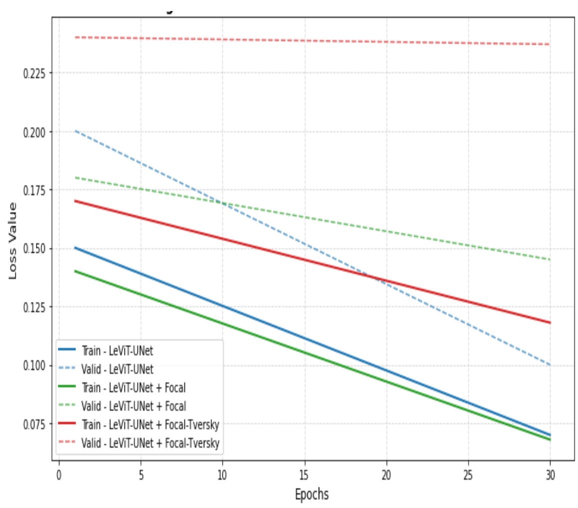

Fig2. Training and validation loss curves for models.

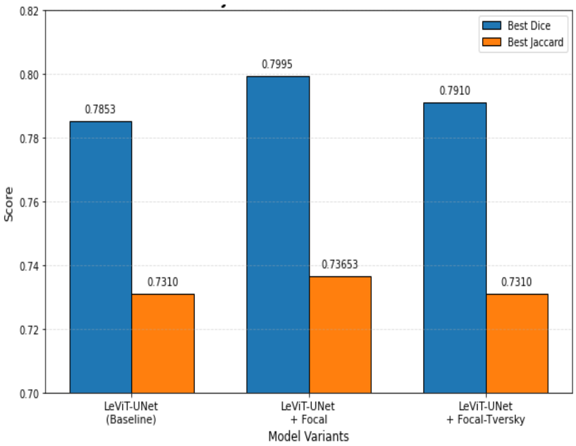

Fig 3. Variant best Dice and Jaccard scores for models.

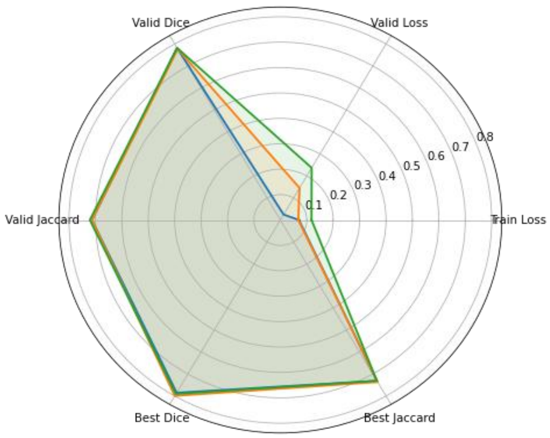

Fig 4. Variant best Dice and Jaccard scores for models.

3.3 DISCUSSION

Table (1) compares the LeViT-UNet model's performance using three different training configurations: the baseline, a model trained with Focal Loss, and one trained with Focal-Tversky Loss. We evaluated each based on their training and validation losses, segmentation accuracy (using Dice and Jaccard indices), and the epoch at which they performed best. A clear pattern emerges from the data: each loss function uniquely shapes how the model converges and generalizes. The baseline model set a training loss benchmark at 0.0705. Switching to Focal Loss yielded a slight but meaningful improvement, driving the training loss down to 0.0679. This lower figure suggests the model converged faster, likely because Focal Loss compels the network to focus on hard-to-classify pixels. In a interesting turn, the Focal-Tversky Loss produced a significantly higher training loss of 0.1181. We can interpret this not as a failure, but as a consequence of its design; by penalizing false negatives more severely, it struggles more during training a trade-off that is particularly valuable for segmenting small structures in imbalanced medical images. When we look at validation loss, the story shifts. The baseline model generalizes most stably, achieving the lowest validation loss of 0.0246. The other two models, trained with Focal Loss (0.1453) and Focal-Tversky Loss (0.2372), show higher validation losses. This is the expected downside of their increased sensitivity to minority classes and harder examples. Crucially, despite this higher loss, both models maintained competitive segmentation accuracy, highlighting a fundamental trade-off between stable generalization and class sensitivity. The accuracy metrics themselves tell a nuanced story. The baseline achieved a Dice score of 0.7821 and a Jaccard of 0.7133. The Focal Loss variant had a slightly lower Dice (0.7768) but a virtually identical Jaccard (0.7134). Most notably, the Focal-Tversky model matched the baseline's Dice (0.7820) while delivering the highest Jaccard index of 0.7211. This points to its superior ability in precisely delineating boundaries and achieving better overlap. Where these models truly diverge is in their convergence timelines. The Focal Loss model peaked early, achieving its best Dice of 0.7995 at epoch 23. The Focal-Tversky model, in contrast, reached a Dice of 0.7910 at a later epoch 28, hinting at a slower but more stable optimization process. The baseline confirmed this trend by converging last, at epoch 29, with a Dice of 0.7853. This sequence confirms that both advanced loss functions enhance performance compared to the base configuration. So, what do we take from this? Focal Loss acts as an accelerator, speeding up convergence and boosting early performance by targeting difficult samples. Focal-Tversky Loss, however, plays a longer game, forging a better balance between sensitivity and specificity that results in more robust segmentation on tricky, imbalanced data. For medical tasks where precision on small structures is paramount, pairing LeViT-UNet with Focal-Tversky Loss appears to be a reliable strategy. Figure Analysis Figure 1 uses a bar chart to visually summarize the core trade-off. It places the Train Loss, Valid Loss, Valid Dice, and Valid Jaccard for each model side-by-side, making the compromise between convergence speed and final accuracy immediately apparent. Figure 2 tracks the loss trends over time. All models decrease their training loss steadily, which is a good sign. However, the baseline and Focal Loss models converge more smoothly and quickly. The Focal-Tversky model's higher validation loss suggests it might be overfitting slightly or simply taking longer to adapt to the class imbalance. In this view, the Focal Loss variant seems to have the most balanced behavior. Figure 3 isolates the peak performance. It clearly shows the Focal Loss model securing the highest Best Dice (0.7995), while the Focal-Tversky model claims the top Jaccard (0.7211). This visually reinforces the conclusion that Focal-Tversky excels at overlap precision. Figure 4 brings it all together on a sample abdominal CT scan. The visual evidence supports the quantitative data: Focal-Tversky loss delivers the most balanced and accurate segmentation. Focal Loss improves on small classes, but Focal-Tversky provides a more comprehensive upgrade. The baseline, while stable, lags in final accuracy.

4. CONCLUSIONS

This evaluation leads us to several key conclusions. The choice of loss function is not a minor detail; it fundamentally directs the LeViT-UNet model's convergence and its ultimate segmentation accuracy on imbalanced medical images. Focal Loss serves as a powerful tool for stabilizing training and speeding up convergence. Focal-Tversky Loss, however, provides a more nuanced control, balancing sensitivity and specificity to enhance the segmentation of small, critical regions. The LeViT-UNet architecture itself proves to be a potent framework, successfully marrying a transformer's attention mechanism with a CNN's efficiency. This makes it well-suited for medical imaging, where both precision and computational practicality are non-negotiable. That said, our results also confirm that the model's performance is still constrained by classic challenges like dataset imbalance and limited diversity, which can hamper generalization. Looking forward, the path seems to point toward hybrid loss functions or adaptive weighting schemes that could offer even greater robustness. Exploring transfer learning and multimodal data fusion would also be logical next steps to improve performance across the wide spectrum of medical imaging modalities. In summary, this work solidifies LeViT-UNet as a highly promising and efficient solution for tackling the complex demands of modern medical image segmentation.

REFERENCES

[1] Xu, G., Wu, X., Zhang, X., & He, X. (2021). LeViT-UNet: Make Faster Encoders with Transformer for Medical Image Segmentation. arXiv preprint arXiv:2107.08623. https://arxiv.org/abs/2107.08623

[2] Abraham, N., & Khan, N. M. (2019, April). 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) (pp. 683-687). IEEE.

[3] Yeung, M., Sala, E., Schönlieb, C.-B., & Rundo, L. (2022). Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalance. Medical Image Analysis, 74, 102217. https://www.sciencedirect.com/science/article/pii/S0895611121001750.

[4] Khan, R. F., & Khan, N. M. (2023). Transformers in medical image segmentation: A narrative review. Journal of Medical Imaging and Health Informatics, 13(6), 1234 1245. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10722011.

[5] G. Xu, X. Zhang, X. He, and X. Wu, LeViT-

UNet: Make Faster Encoders with Transformer for Medical Image Segmentation, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 14432 LNCS, no. December 2023, pp. 42 53, 2024, doi: 10.1007/978-981-99-8543- 2_4.

[6] S. Jadon, A survey of loss functions for semantic segmentation, 2020 IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol. CIBCB 2020, 2020, doi: 10.1109/CIBCB48159.2020.9277638.

[7] Kato, S., & Hotta, K. (2024). Adaptive t-vMF dice loss: An effective expansion of dice loss for medical image segmentation. Computers in Biology and Medicine, 168, 107695.

[8] Azad, R., Heidary, M., Yilmaz, K., Hüttemann, M., Karimijafarbigloo, S., Wu, Y., ... & Merhof, D. (2023). Loss functions in the era of semantic segmentation: A survey and outlook. arXiv preprint arXiv:2312.05391.