U-Net in Medical Imaging: A Practical Pathway for AI Integration in Healthcare

Martin Kryl[a], Pavel Košan, Petr Včelák[b] and Jana Klečková[c]
Department of Computer Science and Engineering, University of West Bohemia, Univerzitni 8, Plzen, Czech Republic
{kryl, vcelak, kleckova}@kiv.zcu.cz

Keywords: Medical Imaging, Deep Learning, U-Net, Clinical AI, Image Segmentation, Healthcare Technology.

Abstract: As AI transforms medical imaging, this paper positions U-Net as a practical and enduring choice for segmentation tasks in constrained clinical environments. Despite rapid advancements in architectures like transformers and hybrid models, U-Net remains highly relevant due to its simplicity, efficiency, and interpretability, particularly in settings with limited computational resources and data availability. By exploring modifications such as residual connections and the Tversky loss function, we argue that incremental refinements to U-Net can bridge the gap between current clinical needs and the potential of more advanced AI tools. This paper advocates for a balanced approach, combining accessible enhancements with hybrid strategies, such as radiologist-informed labeling and advanced preprocessing, to ensure immediate impact while building a foundation for future innovation. U-Net's adaptability positions it as both a cornerstone of today's AI integration in healthcare and a stepping stone toward adopting next-generation models.

1 INTRODUCTION

In recent years, deep learning has revolutionized medical imaging, offering advanced tools that enhance diagnostic support and improve the accuracy of medical data analysis. The healthcare sector, which generates vast volumes of data through modalities like CT, MRI, and X-ray, presents an ideal opportunity for AI applications to improve diagnostic efficiency and reliability. However, practical integration within clinical environments remains challenging, often requiring a balance between advanced model capabilities and healthcare's data and infrastructure limitations (Ronneberger et al., 2015).

Among the widely adopted models in medical imaging, U-Net has become foundational for image segmentation, especially due to its efficient architecture and success even with limited data. Initially designed for biomedical tasks, U-Net has been adapted to various medical imaging applications, consistently demonstrating reliable segmentation results (Azad et al., 2024). Despite its strengths, the rapid development of alternative architectures—such as transformers, GANs, and hybrid models—has raised questions about U-Net's continued relevance. Nonetheless, U-Net and its variants are still favored in settings constrained by limited data, computational power, and interpretability needs, making it a practical choice in many clinical contexts (Ronneberger et al., 2015; Azad et al., 2024) .

This paper assesses a modified U-Net model tailored for brain CT scan segmentation, focusing on its efficacy and clinical viability. This model utilizes the Tversky loss function to address class imbalance (Salehi et al., 2017). Achieved results are indicative of general effectiveness but with limitations in boundary precision . These findings suggest that U-Net's architecture, even with its limitations, offers a balanced and pragmatic approach for clinical use. This is particularly relevant as healthcare facilities continue to face significant barriers in adopting more complex architectures, underlining the ongoing relevance of U-Net in real-world medical imaging.

As the field evolves, architectures like Vision Transformers and advanced CNNs hold promise for greater accuracy and flexibility (Shamshad et al., 2023). However, their requirements for extensive computational resources and large datasets may hinder clinical feasibility (Shamshad et al., 2023). Consequently, this paper advocates for continuous refinement of U-Net-based models, emphasizing an approach that prioritizes clinical accessibility. By en-

[a] https://orcid.org/0000-0001-8077-7298
[b] https://orcid.org/0000-0003-4415-790X
[c] https://orcid.org/0000-0003-2050-6925

Kryl, M., Košan, P., Včelák, P. and Klečková, J.
U-Net in Medical Imaging: A Practical Pathway for AI Integration in Healthcare.
DOI: 10.5220/0013314600003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, pages 828-833
ISBN: 978-989-758-731-3; ISSN: 2184-4305

hancing U-Net's robustness and adaptability, healthcare providers can leverage AI advancements within current infrastructural constraints while building a foundation for future, more sophisticated integrations. By progressively enhancing foundational models, healthcare systems can lay the groundwork for incorporating more complex models, facilitating AI-driven improvements in medical imaging.

2 BACKGROUND AND RELEVANCE

The evolution of deep learning has significantly shaped medical imaging, enabling precise analysis and insights through models trained on extensive datasets. Early convolutional neural networks (CNNs), such as U-Net, were specifically designed to address the complexities of biomedical image segmentation. U-Net's encoder-decoder structure, complemented by skip connections, allows the model to capture both high-level features and fine-grained details, making it highly effective for various medical segmentation tasks. (Ronneberger et al., 2015)

While deep learning continues to progress, and newer models are emerging with potential improvements in accuracy and generalization, the accessibility of these models remains limited. Transformers, for instance, introduce self-attention mechanisms that enable the model to dynamically assess the importance of different image regions. Generative Adversarial Networks (GANs) offer potential for generating high-fidelity images, useful for data augmentation and enhancement. Hybrid models that combine CNNs with transformer-based layers have also been explored to leverage the strengths of both architectures. (Pu et al., 2024)

However, these advanced models often require significant memory and computational resources, demanding high-performance hardware that may be unavailable in many clinical settings. Additionally, their reliance on large, diverse datasets poses a challenge in medical imaging, where data access is often constrained due to privacy considerations and limited variability in available datasets. This makes complex architectures less feasible in many clinical settings, where interpretability and accountability are also critical for diagnostic decision-making. (Ronneberger et al., 2015)

Despite these recent advances, U-Net and its derivatives continue to hold relevance, particularly in constrained environments. U-Net's simplicity makes it feasible to implement on accessible hardware, yet it still produces reliable segmentation results. By focusing on incremental improvements, such as the Tversky loss function or selective attention mechanisms, healthcare providers can leverage U-Net's capabilities as a bridge toward integrating more advanced architectures over time. (Ronneberger et al., 2015)

This paper advocates for a balanced approach that prioritizes the refinement and application of U-Net-based models in real-world clinical contexts. By focusing on incremental improvements to the U-Net architecture, such as enhanced loss functions and increased robustness, healthcare providers can leverage deep learning's benefits within existing infrastructural limits, paving the way for gradual adoption of cutting-edge models as technology and data accessibility improve.

3 METHODOLOGY AND MODEL ARCHITECTURE

To address the specific needs of brain CT scan segmentation in a clinical setting, we utilized a modified U-Net model designed to handle challenges related to class imbalance and constrained data resolution. U-Net's encoder-decoder structure, with skip connections that preserve spatial information across layers, provides a strong foundation for medical image segmentation tasks where capturing both detailed and high-level features is critical. This architectural choice is particularly advantageous in settings with limited computational resources and data, making it an accessible yet effective option for clinical applications.

3.1 Modified U-Net Architecture

The U-Net model was adapted in several ways to improve its performance on the task of brain CT segmentation. One modification was the inclusion of residual connections (He et al., 2016) within the encoder and decoder blocks. These residual connections allow the network to add activations from earlier layers directly to the outputs of deeper layers, enhancing the model's ability to retain and propagate contextual information. By summing the activations, this modification allows the U-Net model to capture both local (edges and small structures) and global features (overall context of the brain scan) more effectively, making it suitable for identifying subtle structures in medical images, such as lesions or infarctions.

Another modification was the use of the Tversky loss function instead of the standard cross-entropy loss. The Tversky loss addresses class imbalance by allowing for fine-tuning of false positives and false negatives, which is especially beneficial in medical segmentation tasks where certain regions may be less prominent. By adjusting this balance, the model becomes more effective in capturing smaller regions that may otherwise be overlooked in traditional loss function setups (Sudre et al., 2017; Abraham and Khan, 2019).

In addition to architectural modifications, data augmentation techniques inspired by (Shorten and Khoshgoftaar, 2019; Nemoto et al., 2021) were applied to improve model robustness given the limited dataset size. This preprocessing approach aligns with the model's goal of achieving high segmentation accuracy without requiring an extensive dataset, which is often impractical in clinical environments due to data access, privacy concerns or the amount of effort needed to prepare large quality datasets for model learning.

3.2 Rationale for U-Net Selection

The decision to utilize a modified U-Net over more recent architectures, such as Vision Transformers or hybrid models, was driven by several practical considerations. Unlike more complex models, U-Net's architecture is relatively lightweight and can be deployed on standard hardware configurations commonly available in healthcare facilities. This simplicity, combined with U-Net's demonstrated effectiveness in segmentation tasks, offers a feasible approach to introducing AI-driven diagnostics in clinical settings without extensive infrastructure upgrades.

Furthermore, U-Net's interpretability provides an additional advantage over newer architectures. In a clinical setting, where transparency is crucial, U-Net's straightforward encoder-decoder structure allows for greater model interpretability, making it easier for clinicians to understand and trust the segmentation outputs. Given that interpretability and accountability are critical for clinical adoption as noted in (Siddique et al., 2021), U-Net's design strikes a balance between accuracy and comprehensibility that newer, more complex architectures may not offer as readily.

3.3 Dataset Preparation and Training

The model was trained on a curated local dataset of 50 brain CT image series. Each series included core and penumbra segmentation masks derived from automated and semi-automated techniques. Specifically, penumbra regions were automatically labeled using a custom script that analyzed cerebral blood flow (CBF) and cerebral blood volume (CBV) maps, leveraging standard clinical thresholds to identify ischemic but salvageable tissue. These initial masks were subsequently reviewed and validated by an experienced radiologist to ensure that the segmentations aligned with clinical expectations. Series with disputed or ambiguous regions were excluded from the dataset, ensuring high-quality annotations.

Images in each series were downsampled to a resolution of 256x256 pixels to optimize processing efficiency while retaining the essential features for segmentation. This resolution was selected to align with realistic data limitations in clinical environments, where high-resolution images may not always be feasible to handle due to storage and processing constraints. Furthermore, the chosen resolution and model design ensure that segmentation tasks can be performed swiftly, an important consideration in clinical workflows where timely results are crucial.

Data preprocessing included standard normalization to ensure consistent intensity ranges across images, improving model stability during training. Given the relatively small size of the dataset, augmentation techniques including random rotations, image translations and offsets, horizontal and vertical flips, and brightness adjustments were applied to enhance model robustness in accordance to findings in (Shorten and Khoshgoftaar, 2019; Siddique et al., 2021). Augmentations were only applied to slices containing regions of interest to maximize the relevance of the augmented data while avoiding unnecessary transformations of non-informative slices. These augmentations expanded the effective size of the dataset and mitigated the risk of overfitting.

The model was trained over 35 epochs, with the stopping point determined empirically based on the progression of the loss function on the validation set. This early stopping criterion was chosen to prevent overfitting while ensuring adequate convergence.

4 RESULTS AND PERFORMANCE EVALUATION

The modified U-Net model's performance was evaluated using key metrics standard in medical image segmentation: the Dice coefficient and the Tversky coefficient. These metrics assess the overlap accuracy between the predicted segmentation and the reference labels, providing insights into general segmentation accuracy and the model's handling of class imbalances.

4.1 Performance Metrics and Outcomes

On the validation set, the model achieved a Dice coefficient of 0.61 and a Tversky coefficient of 0.67. The Dice coefficient reflects the model's overall performance in matching target segmentation regions, while the Tversky coefficient indicates its effectiveness in managing class imbalances. The application of the Tversky loss function during training enhanced the model's sensitivity to less prominent regions in the CT images, such as smaller lesions, which might otherwise have been underrepresented.

These metrics suggest that the model can approximate the general area of the segmentation target accurately, though challenges remain in achieving precise boundary alignment. While these metrics may appear modest, they reflect the inherent difficulty of the task: segmenting small, subtle structures like ischemic penumbra regions from noisy CT images. These challenges are further compounded by the constrained dataset size (50 series) and the necessity of downsampling images to 256x256 resolution for practical deployment. This trade-off aligns with known limitations of U-Net architectures in high-precision medical applications, where complex structures may require more refined model adjustments or larger, higher-resolution datasets.

Additionally, manual annotations of medical images often exhibit variability between radiologists, with inter-annotator Dice scores sometimes falling within similar ranges in comparable tasks. This model's performance aligns with the general accuracy achieved by a domain expert annotating a new dataset, though precise quantitative comparisons were unavailable. Nonetheless, the results indicate that the modified U-Net, even with its relatively simple structure, can deliver meaningful outcomes in scenarios with limited data and computational resources.

4.2 Comparative Analysis with Local Delineation Tool

To evaluate the modified U-Net model's effectiveness, a comparative analysis was performed using segmentation outputs from a locally developed tool designed for infarct core delineation in brain CT imaging. The tool is successor to the Delineator published in (Maule et al., 2013). Although the tool has not become widely adopted outside its original setting, it provides a baseline segmentation that facilitates data labeling. This allowed us to generate more labeled training data than would have been feasible with manual radiologist labeling alone.

The use of the tool enabled approximate delineation of the infarct core regions in the dataset, providing a reference standard against which the U-Net model could be evaluated. However, it's important to acknowledge that both the software tool and manual radiologist annotations may contain inaccuracies. Without a rigid, multi-annotator labeling process, objectively determining the segmentation accuracy of either approach remains challenging.

Despite these limitations, the U-Net model's outputs closely aligned with the broad areas identified by the tool, capturing the primary regions of interest with reasonable accuracy. This consistency suggests that the U-Net model, even with its simpler architecture, is suitable for approximate segmentation in cases where precise boundary conformity may be secondary to general region identification. In scenarios where exact segmentation is not strictly required, U-Net offers a viable alternative to more complex software solutions, especially in settings where computational and resource constraints are significant considerations.

4.3 Interpretation of Results and Trade-Offs

The performance of the modified U-Net model reflects both the strengths and trade-offs of using a U-Net-based approach in medical imaging. The model succeeded in identifying the regions of interest broadly, providing a valuable tool for clinicians seeking approximate segmentation. However, its limitations in fine-grained boundary alignment indicate that, while U-Net can approximate the segmentation task, it may not be able to fully replace specialized software without further enhancements.

These results highlight a pragmatic pathway for using U-Net in real-world clinical settings: the model can offer reliable, interpretable segmentation without extensive infrastructure requirements, but as noted in (Isensee et al., 2021), additional adjustments or hybrid approaches may be necessary for applications requiring high precision. In constrained environments, where data access, computational power, and interpretability are significant considerations, this U-Net-based model demonstrates that effective segmentation is achievable with thoughtful modifications, even as the field of medical imaging continues to evolve.

5 POSITION STATEMENT

As AI integrates more deeply into healthcare, U-Net's practical architecture and strong performance make it highly suited for clinical applications, especially under the infrastructural limitations that constrain many medical settings. While novel architectures—such as transformers and hybrid networks—offer higher precision and richer context through attention mechanisms, they demand substantial computational resources and interpretability solutions, challenging immediate adoption (Henry et al., 2022). This work does not argue that U-Net supersedes more advanced architectures; rather, it highlights U-Net's enduring relevance and adaptability in settings where simplicity, efficiency, and interpretability are critical.

This paper advocates for incrementally enhancing U-Net-based models, emphasizing simplicity, clinical interpretability, and targeted modifications like the Tversky loss function to manage class imbalances. By refining U-Net within these limits, healthcare providers can implement AI-based segmentation today without the extensive resources newer models often require.

Moreover, as a foundational model, U-Net can be further developed to incorporate aspects of advanced architectures such as attention mechanism in (Pu et al., 2024). This hybridization pathway enables healthcare facilities to integrate transformer-based attention or other advanced techniques selectively, building a bridge to complex, data-intensive models while meeting current clinical needs. This progressive approach enables real-world impact today while paving the way for advanced model integration as data, computational resources, and clinical AI familiarity expand.

6 FUTURE WORK

We acknowledge the potential for further strengthening our findings through additional experiments and comparisons. An ablation study examining the contributions of the proposed modifications, such as residual connections and the Tversky loss function, could provide deeper insights into their individual and combined impacts on segmentation performance. Similarly, an empirical comparison between U-Net and modern architectures, such as Vision Transformers or hybrid models, would help to better demonstrate the trade-offs between computational efficiency, data requirements, and segmentation accuracy. Finally, detailed experiments incorporating other state-of-the-art approaches, particularly those leveraging hybrid strategies or advanced preprocessing techniques, could contextualize U-Net's performance within a broader framework. These lead into several specific research directions highlighted below.

6.1 Hybrid Labeling with Radiologist Input

Combining radiologist oversight with automated segmentation creates a hybrid labeling approach that enhances data quality and enables valuable model refinements. Allowing experts to adjust AI outputs during the labeling process improves segmentation reliability and provides feedback loops that contribute to ongoing model improvements. This collaborative strategy leverages the strengths of both human expertise and machine efficiency, which could lead to more accurate and clinically relevant outcomes.

6.2 Advanced Preprocessing with Complex Models

Leveraging sophisticated architectures, such as transformers, for preprocessing can enrich datasets for simpler models like U-Net. This tiered approach provides high-quality features that simpler models can efficiently utilize, allowing computationally feasible models to benefit from the strengths of cutting-edge feature extraction. By integrating advanced preprocessing techniques, the performance of established models can be significantly enhanced without necessitating substantial computational resources.

6.3 Navigating Diagnostic Uncertainty

Acknowledging the absence of absolute truth in medical imaging, future work should address inherent inaccuracies in both human and software assessments. Developing confidence metrics and quality-assurance feedback loops, particularly with radiologist input, can enhance reliability, helping to mitigate biases across human and AI judgments. Implementing these measures ensures that diagnostic processes are transparent and that uncertainties are systematically managed, leading to more trustworthy clinical decisions.

6.4 Integration of Hybrid Models and Long-Term Deployment Studies

Research should explore the gradual integration of hybrid U-Net models into clinical workflows, incorporating components from advanced architectures like transformers without overwhelming clinical resources. By deploying these hybrid models in clinical settings for extended studies, researchers can address practical deployment challenges, contributing insights that prepare healthcare facilities for eventual transitions to fully advanced models. This phased approach allows for the assessment of real-world performance and the identification of necessary adjustments, facilitating a smoother adoption of AI technologies in healthcare.

7 CONCLUSION

This paper has emphasized the enduring relevance and adaptability of U-Net-based architectures in medical imaging, highlighting their effectiveness and practicality in clinical environments often constrained by limited resources and data. U-Net's simplicity, interpretability, and robustness make it particularly well-suited to meet healthcare's immediate needs, offering reliable segmentation with manageable computational demands. By integrating targeted enhancements, U-Net-based models serve as a bridge between traditional diagnostic tools and the transformative potential of deep learning.

Incremental enhancements, such as attention mechanisms and refined loss functions, allow U-Net to improve without requiring significant infrastructure upgrades. These modifications provide a practical pathway for increasing segmentation accuracy while preparing for the eventual integration of more advanced architectures.

Additionally, recognizing the inherent lack of an objective truth in medical imaging, this paper advocates for hybrid approaches that incorporate radiologist feedback and advanced preprocessing methods to enhance data quality and model accuracy. These pragmatic strategies facilitate AI adoption in clinical workflows while supporting the development of robust, quality-assurance frameworks to reduce biases in both AI outputs and clinician interpretations.

Ultimately, this paper supports a balanced, progressive approach to AI integration in healthcare. U-Net serves as a practical bridge between traditional tools and next-generation AI, enabling real-world impact today while laying the groundwork for sophisticated, data-intensive models in the future.

REFERENCES

Abraham, N. and Khan, N. M. (2019). 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.

Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J. P., Adeli, E., and Merhof, D. (2024). Medical image segmentation review: The success of u-net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Henry, E. U., Emebob, O., and Omonhinmin, C. A. (2022). Vision transformers in medical imaging: A review. *arXiv preprint arXiv:2211.10043*.

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H. (2021). nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211.

Maule, P., Klečková, J., Rohan, V., and Tupỳ, R. (2013). Automated infarction core delineation using cerebral and perfused blood volume maps. *International journal of computer assisted radiology and surgery*, 8:787–797.

Nemoto, T., Futakami, N., Kunieda, E., Yagi, M., Takeda, A., Akiba, T., Mutu, E., and Shigematsu, N. (2021). Effects of sample size and data augmentation on u-net-based automatic segmentation of various organs. *Radiological Physics and Technology*, 14:318–327.

Pu, Q., Xi, Z., Yin, S., Zhao, Z., and Zhao, L. (2024). Advantages of transformer and its application for medical image segmentation: a survey. *BioMedical Engineering OnLine*, 23(1):14.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer.

Salehi, S. S. M., Erdogmus, D., and Gholipour, A. (2017). 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.

Shamshad, F., Khan, S., Zamir, S. W., Khan, M. H., Hayat, M., Khan, F. S., and Fu, H. (2023). Transformers in medical imaging: A survey. *Medical Image Analysis*, 88:102802.

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48.

Siddique, N., Paheding, S., Elkin, C. P., and Devabhaktuni, V. (2021). U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE access*, 9:82031–82057.

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Jorge Cardoso, M. (2017). 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: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 240–248. Springer.