Proceedings of CONF-SEML 2025 Symposium: Machine Learning Theory and Applications
DOI: 10.54254/2755-2721/2025.TJ23590

# *Key Technologies Analysis of Deep Learning-based Medical Image Segmentation*

**Ruochen Feng**

*Faculty of Science and Technology, Beijing Normal-Hong Kong Baptist University, Beijing, China*
*s230026036@mail.uic.edu.cn*

***Abstract:*** This study explores Deep Learning (DL) methods in medical image segmentation. By precisely delineating anatomical structures and focal areas, it enhances the precision of diagnostic and treatment strategies. With an emphasis on the architecture design, performance advantages, and clinical application scenarios of Convolutional Neural Network (CNN), Fully Convolutional Network (FCN), and U-shaped Convolutional Network (U-Net) models, the research work examined the fundamentals of DL-based medical image segmentation technology. The working mechanism of these models, their adaptability to multimodal images, and their practical value in clinical scenarios. DL models perform better in precision as well as effectiveness when compared to traditional methods. Specifically, the hierarchical feature extraction capability of CNN shows robustness in feature extraction for organ segmentation, the pixel-level prediction mechanism of FCN is outstanding in fine-grained segmentation tasks, and the encoder-decoder structure of U-Net, combined with skip connections, improves segmentation accuracy. Compared to traditional methods, DL models demonstrate superior performance in terms of accuracy and efficiency. Specifically, CNN's hierarchical feature extraction capability exhibits robustness in feature extraction for organ segmentation. The pixel-level prediction mechanism of FCN stands out in fine-grained segmentation tasks. Additionally, the encoder-decoder architecture of U-Net, combined with skip connections, improves segmentation accuracy, although this requires significant computational resources.

***Keywords:*** Deep Learning (DL), Convolutional Neural Network (CNN), Fully Convolutional Network (FCN), U-shaped Convolutional Network (U-Net).

## 1. Introduction

Medical image segmentation, which involves separating organ or lesion pixels from background images, is crucial for assessing anatomical or pathological structures, providing key information on volumes and shapes [1]. This process is vital in computer-aided diagnosis and smart medicine by enhancing diagnostic accuracy and efficiency [2]. Artificial Intelligence (AI), which mimics human intelligence for problem-solving and learning, has revolutionized image segmentation [3]. AI is broadly categorized into Machine Learning (ML) and Deep Learning (DL). ML techniques have been essential in medical image processing, including segmentation, interpretation, and image-guided therapy [4]. DL, in particular, offers improved segmentation performance, helping in feature selection, extraction, and generation [4]. By reducing the burden on healthcare professionals, DL-

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.TJ23590

based segmentation aids in tumor size verification and post-therapy assessment, contributing to more efficient medical decision-making [5].

Recent advancements in medical image segmentation have been dominated by ML and DL techniques, which overcome the limitations of traditional model-driven approaches. While early methods relied on manual feature engineering (e.g., edge detection, template matching, and hybrid SVM-level set models), their dependency on expert knowledge and sensitivity to noise or low contrast hindered scalability [4,6]. The rise of DL has revolutionized the field, with Convolutional Neural Networks (CNNs) like U-Net leading the charge. CNNs automate hierarchical feature extraction directly from raw data, achieving robustness against artifacts such as blur and noise [7, 8]. Modern frameworks extend beyond basic architectures: Generative Adversarial Networks (GANs) enhance segmentation through synthetic data generation, while transformer-based models improve long-range dependency modeling [8, 9]. Taghanaki et al. systematized DL solutions into six categories, including loss-optimized and weakly supervised methods, addressing challenges like scarce annotations. For instance, weakly supervised techniques leverage incomplete or inexact labels (e.g., Zhou et al.'s work on imperfect supervision), reducing reliance on fully annotated datasets [10].

Current research prioritizes multi-modality fusion and real-time clinical integration. Akkus et al. [9] demonstrated DL's efficacy in brain Magnetic Resonance Imaging (MRI) segmentation, achieving radiologist-level precision. Similarly, transformer architectures are being adapted for 3D medical volumes, enhancing contextual awareness [11]. Industry-academia collaborations further accelerate clinical adoption, with AI tools now assisting in tumor quantification, therapy assessment, and surgical planning [12]. Despite progress, challenges persist in interpretability, cross-domain generalization, and computational efficiency, driving innovations like Neural Architecture Search (NAS) and lightweight models [13]. The main objective of this research is to analyze the key DL technologies in medical image segmentation, concentrating on the structures, characteristics, and clinical applicability of models. This study provides a fundamental reference for future research and clinical implementation by comparing their distinct features, application scopes, and respective strengths and weaknesses in medical image segmentation.

## 2. Methodology

### 2.1. Dataset description

Medical image analysis relies on diverse datasets tailored to specific anatomical structures and clinical tasks. Key datasets are shown in Table 1. These datasets originate from academic challenges (e.g., brain tumor segmentation (BraTS)), hospital collaborations (e.g., Lung Nodule Analysis (LUNA)16), and public repositories such as The Cancer Imaging Archive (TCIA) [14]. They typically include raw medical images, expert annotations (segmentation masks, labels), and relevant clinical metadata. For instance, the BraTS dataset provides multi-modal MRI scans with tumor segmentation labels for brain tumor research, while LUNA16 focuses on lung nodule detection using Computed Tomography (CT) scans [14]. Open Artwork System Interchange Standard (OASIS) offers longitudinal MRI data for brain structure segmentation, aiding studies on neurodegenerative diseases [15]. STARE supports retinal vessel segmentation in ophthalmology, and Camelyon enhances breast cancer diagnostics with whole-slide histopathology images [15]. Emerging datasets like CryoNuSeg cater to specialized tasks such as nuclear instance segmentation in histopathology [14, 15]. The type of datasets is shown in Table 1.

DOI: 10.54254/2755-2721/2025.TJ23590

Table 1: The introduction of datasets

| Dataset Name | Segmentation Task | Data Scale | Content Description |
|---|---|---|---|
| BraTS | Brain tumor segmentation | ~300-400 cases | Multi-modal MRI scans (T1, T1Gd, T2, FLAIR) with tumor segmentation labels |
| LUNA16 | Lung nodule segmentation | 888 CT scans | CT scans with lung nodule segmentation annotations |
| OASIS | Brain structure segmentation | Varies (e.g., OASIS-3: 1098 subjects) | Longitudinal MRI scans with segmentation labels for brain structures |
| STARE | Retinal vessel segmentation | 20 images | Color fundus photographs with vessel segmentation labels |
| Camelyon | Metastasis region segmentation | ~400 whole-slide images | Histopathology whole-slide images with metastasis region annotations |

## 2.2. Methods

### 2.2.1. Introduction of Machine Learning (ML) and Deep Learning (DL)

ML is a branch of AI that allows systems to automatically learn and predict independently from data patterns, without explicit programming. By automatically identifying and sketching organs or diseased areas in medical image segmentation, ML improves the speed and precision of diagnosis and therapy planning. Its primary attributes are data-driven, highly adaptive, and automatically extracting features. Its main techniques include k-Nearest Neighbors (k-NN), which classifies based on adjacent instances; Random Forests, which combine decision trees for robust predictions; and Support Vector Machines (SVM), which use a hyperplane to distinguish classes. The significance of machine learning is seen in its ability to lessen radiologists' workload, reduce errors, and improve consistency. Though it is less effective than DL for large and complicated images, it performs well with smaller datasets.

As a subset of ML, DL models intricate patterns in data using multi-layered neural networks; the main technical feature of DL is its ability to extract features automatically. It directly extracts features from raw inputs, which is well-suited for tasks involving high-dimensional data like medical images. Compared with traditional ML, the representation of deep features is more flexible and highly adaptable. Models can be adapted to various segmentation tasks and medical image types, including ultrasound, CT, MRI, and X-ray images, by modifying their architecture or using transfer learning. Hence, the essence of deep learning is featuring representation learning. Deep neural networks are essentially a series of nested nonlinear transformations. DL provides previously unheard-of accuracy and efficiency in medical imaging segmentation by automating identifying features like tumors, organs, or lesions.

### 2.2.2. CNN for medical image segmentation

CNN is a type of DL architecture specifically designed for processing grid-like data. The structure of CNN is shown in Figure 1. Start with the input layer, which is linked to the input image, where the number of neurons equals the number of pixels in the input image. The next set of layers is convolutional layers; they extract spatial features through filters, also called kernels. The size of filters is arbitrary, depending on the kernel's size. Each neuron responds only to the receptive field of the previous layer. The activation layer maps the feature maps to the nonlinear space via a nonlinear

DOI: 10.54254/2755-2721/2025.TJ23590

function (Rectified Linear Unit (ReLU)), so every convolutional layer's output is treated as activation maps, improving the model's expressiveness and enabling the network to learn complex relationships.

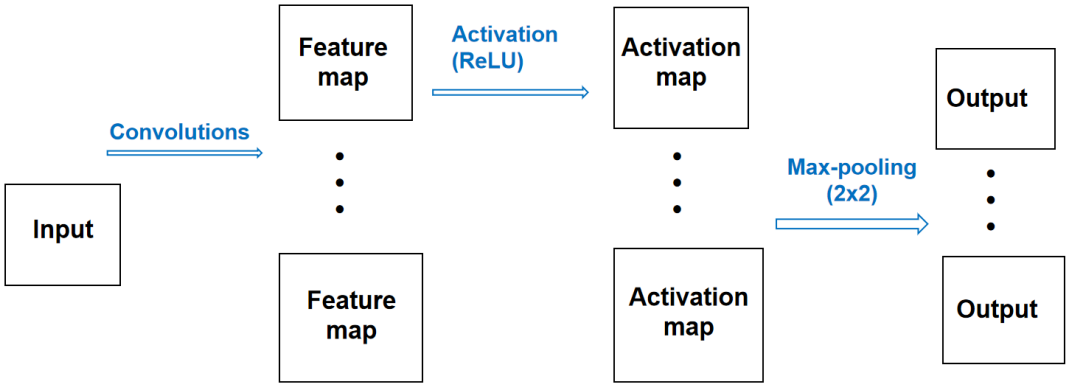


Figure 1: The CCN architecture (photo credit: original)

To lower the feature map's dimension, a subsampling layer is added in between the convolutional layers. This is typically achieved through implementing a maximum pooling operation that maintains the maximum pixel values constant for every non-overlapping 2×2 blocks within the feature map. The final fully connected layers translate the extracted features into the final classification output, where all neurons are connected to each other. CNN gradually transitions from shallow edge texture extraction to deep complex semantic features (e.g., organ morphology) through stacking convolution and pooling operations. In the backpropagation process during training, the weights of neural connections and kernels are iteratively refined to enhance performance.

General CNN is a 2D CNN used to process two-dimensional images (e.g., X-rays), and the convolution kernel slides in the plane, which is suitable for slice-level analysis or tasks with small interlayer dependence. 3D CNN uses 3D filters to analyze stereoscopic data, and the convolution kernel slides in spatial dimensions to capture stereoscopic context in detail, making it particularly crucial for 3D scanning and segmentation like CT or MRI. CNN is widely used in medical imaging for heart and liver segmentation and pneumonia classification. It excels in automatic high-precision feature extraction, which can raise the precision and effectiveness of diagnostic segmentation.

### 2.2.3. FCN for medical image segmentation

FCN is an extension of CNN. The FCN architecture is shown in Figure 2. It uses a fully convolutional layer to replace the last fully connected layer, making it run a dense pixel-wise prediction. To achieve image segmentation. The first block of the FCN structure is the encoder, which is similar to a general CNN, including an input layer, convolutional layers, an activation layer, and a pooling layer. However, convolutional layers compose the first five layers of the typical CNN structure, whereas fully connected layers comprise the sixth through eighth layers [5]. FCN turns the three layers from layers 5 to 7 into convolutional layers [5]. The SoftMax layer generates classification information for each pixel, effectively resolving the partition problem. An encoder is used to extract hierarchical features of images and reduce the spatial dimensions of feature maps. Next is the decoder. In FCN, the deconvolution layer serves to upsample the feature map produced by the last convolution layer, restoring it to the original image size. Hence, pixel-level prediction is realized. Then came skip connections. FCN introduced skip connections, which directly connected the corresponding layers of

DOI: 10.54254/2755-2721/2025.TJ23590

the encoder and decoder. This helps to preserve spatial details and improve segmentation accuracy. Lastly, the upsampled feature map is subjected to pixel-by-pixel classification to finish the image segmentation.

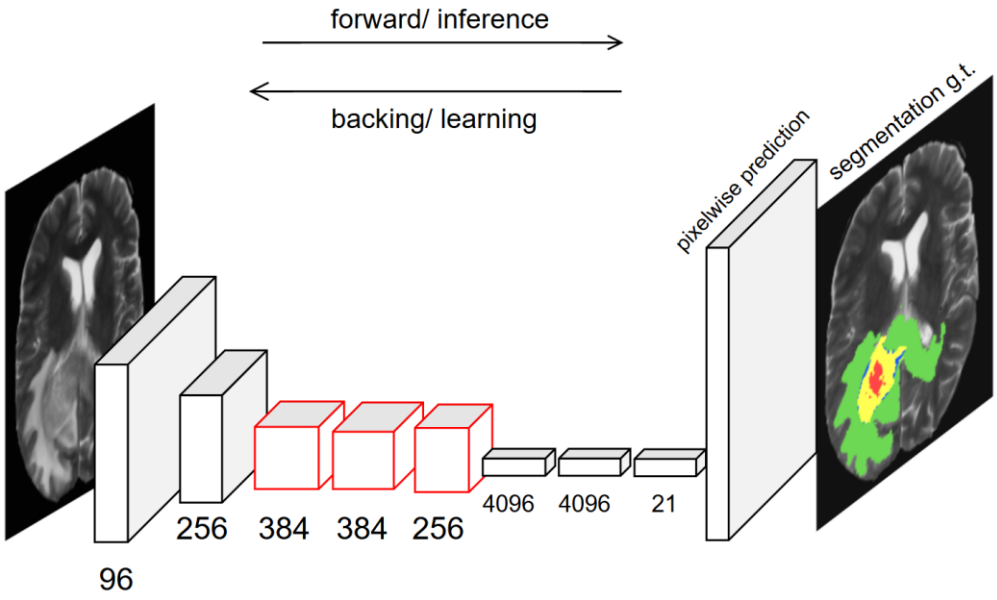


Figure 2: The FCN architecture (photo credit: original)

FCN can learn the spatial structure of different tissues and distinguish them from the surrounding background, which is suitable for organ and tumor segmentation, especially liver segmentation. Due to its ability to recognize delicate structures, FCN is widely used in retinal blood vessel segmentation (e.g., OCT and fundus images) and cell segmentation in pathological sections. FCN can capture fine features, so it has obvious advantages in the accurate location of lesions and regional examination.

### 2.2.4. U-Net for medical image segmentation

U-Net is a CNN architecture derived from FCN, designed to enhance medical image segmentation through its U-shaped structure. The architecture of U-Net (Figure 3) consists of two main components: an encoder (contracting path) and a decoder (expansive path), linked via skip connections, making it well-suited for small sample medical image segmentation. Like traditional CNNs, the encoder comprises four sub-modules, each featuring two convolutional layers followed by a max pooling layer for downsampling. This structure progressively reduces the spatial dimensions of the feature maps while simultaneously increasing the number of feature channels, thereby enabling the capture of high-level contextual information. The decoder mirrors the encoder, with four sub-modules that use upsampling (e.g., deconvolution) to restore spatial resolution. After each upsampling, two convolutional layers optimize feature maps, reduce channel numbers, and restore the size to match the input image, producing the final segmentation output. Skip Connections Connect layers of the exact resolution as the encoder and decoder directly. In the encoder, high-resolution low-level features are combined with the decoder's high-level semantic features through splicing. This way, fine-grained spatial information is preserved and segmentation accuracy is improved. The output layer employs a 1×1 convolution followed by a SoftMax activation function, assigning each pixel to a distinct category to produce the final segmentation map.

DOI: 10.54254/2755-2721/2025.TJ23590

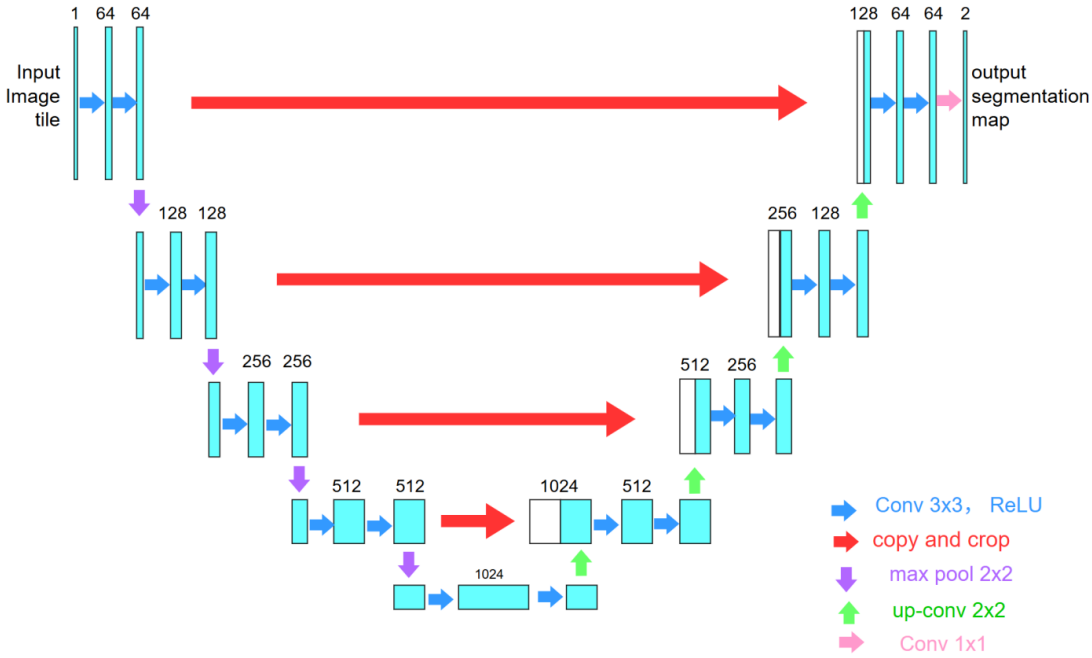


Figure 3: The U-Net architecture (photo credit: original)

U-Net solves the challenges of noise and blurred boundaries in the segmentation of medical images. U-Net has excellent performance in organ segmentation in CT, MRI, and X-ray modes and is often used to segment medical images of lungs, liver, kidney, brain, and other organs, and is suitable for detecting and segmenting small structures (e.g., lesions, tumors). Moreover, U-Net's adaptability to various image sizes has led to the development of subsequent models such as 3D U-Net and V-Net. 3D U-Net extends the U-Net architecture to three dimensions, using 3D convolution and pooling layers. It can directly process three dimensions of volume data, capture spatial relationships, and is suitable for 3D segmentation. V-Net adds residual connections at each stage to improve training efficiency and uses the Dice loss function, which is ideal for dealing with class imbalance problems.

## 3. Discussion

### 3.1. Advantages and disadvantages analysis

#### 3.1.1. CNN

It extracts hierarchical features through convolution and pooling layers, suitable for medical image segmentation tasks such as the heart or liver. However, the fully connected layer limits the pixel-level prediction and reduces precision, especially when the deep spatial information is lost; it is challenging to deal with small structures.

#### 3.1.2. FCN

Enabling dense pixel-level prediction. Skip connections and upsampling preserve spatial detail, making it suitable for organ and tumor segmentation, such as liver or retinal blood vessels. However, FCN is produced with lower accuracy when dealing with smaller organs such as the pancreas. The problem with FCN is that the receive size is fixed, so if the object size changes, FCN will have difficulty detecting all images. Due to inadequate upsampled resolution recovery, FCN may lose subtle features in complex images.

DOI: 10.54254/2755-2721/2025.TJ23590

#### 3.1.3. U-Net

Based on FCN, U-Net uses a symmetric encoder-decoder structure and skip connections, blending low-level and high-level features. Its design excels at preserving spatial detail, making it ideal for segmentation of small and medium-sized structures in CT, MRI, or X-ray images. However, U-Net has a high demand for computing resources, and regularization is required on small data sets to prevent overfitting.

### 3.2. Future direction

In the future, a hybrid model integrating the attention mechanism or Transformer with the above models can improve the segmentation accuracy of noisy or low-contrast images. Additionally, it is possible to extend 3D U-Net or V-Net to optimize real-time 3D segmentation of volume data for clinical CT or MRI. Furthermore, for transfer learning, pre-training and fine-tuning on large data sets could ease the challenge of limited labeled medical data. FCN and U-Net are used for liver, lung, and kidney segmentation to assist in surgical planning. CNN and U-Net excel in identifying tumors or lesions to aid in early diagnosis. Finally, FCN and U-Net segmented retinal vessels to support the diagnosis of diabetic retinopathy. There are some bottlenecks, but the thesis can find solutions. For example, medical data sets are small. It can be mitigated through data enhancement, synthetic data, and semi-supervised learning for class imbalance, where background pixels dominate segmentation tasks. Dice loss or focus loss can reduce this challenge.

## 4. Conclusion

This study details fundamental DL techniques applied to medical image segmentation, focusing on their fundamental structures, characteristics, and applications in medical image segmentation. These are essential for improving diagnostic accuracy and operational efficiency in healthcare. Three core architectures, CNN, FCN, and U-Net, are proposed for analysis. The hierarchical feature extraction capability of CNN is evaluated, and the pixel-level segmentation capability of FCN is enhanced by upsampling and skip connection. U-Net has attracted attention for its encoder-decoder framework's excellent performance in segmenting small anatomical structures. A synthesis of prior studies reveals that U-Net is generally superior to CNN and FCN in segmentation accuracy, especially in small structure segmentation, due to its capability to preserve spatial details. However, U-Net requires substantial computational resources. CNN performs well in feature extraction but lacks precision at the pixel level. At the same time, FCN excels in organ-level pixel prediction but presents challenges when dealing with objects of different sizes. In the future, A will be considered as the research objective for the next stage. The research will focus on hybrid models incorporating attention mechanisms or Transformers to enhance segmentation in complex situations. At the same time, there are plans to extend these architectures to 3D frameworks such as 3D U-Net for real-time clinical integration and to utilize transfer learning to address limited labeled medical images.

## References

[1] *Hesamian, M.H., Jia, W., He, X. et al. (2019). Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. J Digit Imaging 32, 582–596.*

[2] *Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., & Nandi, A. K. (2022). Medical image segmentation using deep learning: a survey. IET Image Processing, 16(5), 1243-1267.*

[3] *Barragán-Montero, Umair, J., Gilmer, V., Dan, N.Y. et al. (2021). Artificial intelligence and machine learning for medical imaging: A technology review, Physica Medica, 83, 242-256.*

[4] *Razzak, M.I., Naz, S., Zaib, A. (2018). Deep Learning for Medical Image Processing: Overview, Challenges and the Future. In: Dey, N., Ashour, A., Borra, S. (eds) Classification in BioApps. Lecture Notes in Computational Vision and Biomechanics, 26.*

DOI: 10.54254/2755-2721/2025.TJ23590

[5] *Liu, X., Song, L., Liu, S., Zhang, Y. (2021). A Review of Deep-Learning-Based Medical Image Segmentation Methods. Sustainability, 13, 1224.*

[6] *Masood, S., Sharif, M., Masood, A., Yasmin, M., Raza, M. (2015). A survey on medical image segmentation. Curr. Med. Imaging Rev. 11(1), 3–14.*

[7] *aye, M. M. (2023). Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions. Computation, 11(3), 52.*

[8] *Taghanaki, S.A., Abhishek, K., Cohen, J.P., Cohen.Adad, J., Hamarneh, G. (2021). Deep semantic segmentation of natural and medical images: a review. Artif. Intell. Rev. 54, 137–178.*

[9] *Panayides, A. S. et al. (2020). AI in Medical Imaging Informatics: Current Challenges and Future Directions. in IEEE Journal of Biomedical and Health Informatics, 24(7), 1837-1857.*

[10] *Wang, Y., Albrecht, C. M., Braham, N. A. A., Mou, L., and Zhu, X. X. (2022). Self-Supervised Learning in Remote Sensing: A review. in IEEE Geoscience and Remote Sensing Magazine, 10(4), 213-247.*

[11] *Hesamian, M.H., Jia, W., He, X., Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. J. Digit Imaging 32(4), 582–596.*

[12] *Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X. (2020). Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. Med. Image Anal. 63, 101693.*

[13] *Eelbode, T., Bertels, J., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R., et al. (2020). Optimization for medical image segmentation: theory and practice when evaluating with dice score or jaccard index. IEEE Trans. Med. Imaging 39(11), 3679–3690.*

[14] *Zhou, S. K., Greenspan, H., Davatzikos, C., et al. (2021). A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises. Proceedings of the IEEE, 109(5), 820-838.*

[15] *Esteva, A., Chou, K., Yeung, S., et al. (2021). Deep learning-enabled medical computer vision. NPJ Digital Medicine, 4(1), 1-9.*