# Deep Learning Approach for Three-Dimensional Liver Image Segmentation using 3D UNet and Its Variant with Transformer

**Farhat Gumelar[1], Anindya Apriliyanti Pravitasari[*2], Yusep Suparman[3], Mohammad Hamid Asnawi[4], Chipta Dwi Ramadian[5], Triyani Hendrawati[6], and Gumgum Darmawan[7]**

[1, 2, 3, 4, 5, 6, 7] Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, INDONESIA.
(E-mail: farhat20001@mail.unpad.ac.id, anindya.apriliyanti@unpad.ac.id, yusep.suparman@unpad.ac.id, mohammad19011@mail.unpad.ac.id, chipta20001@mail.unpad.ac.id, triyani.hendrawati@unpad.ac.id, gumgum@unpad.ac.id)

**Keywords:** Liver, Image Segmentation, 3D UNet, 3D UNETR, 3D Swin-UNet, Vision Transformer.

## ABSTRACT

As the largest organ in the abdomen, the liver plays an important role in the body's metabolism. Nonetheless, several kinds of conditions that can be referred to as liver diseases might impair liver function in the body's metabolic processes. In this context, doctors typically do supplementary examination procedures in the form of liver scans using computed tomography (CT) scans to assess a patient's liver condition.

Liver segmentation, which involves identifying the liver as a region of interest (ROI) within the CT scan image, serves as the primary method for evaluating the organ's health [1]. Nonetheless, liver segmentation is a challenging task due to the organ's diverse pathologies, variable shapes, and proximity to other organs [1]. Consequently, a technology that can precisely and effectively segment the liver is needed in order to accelerate the liver segmentation process.

In this study, liver segmentation based on CT scan images was performed using the 3D UNet, a commonly utilized algorithm in the analysis of three-dimensional biomedical imaging data. However, the 3D UNet architecture is inherently limited by its convolutional layers, which have restricted long-term memory capabilities [2]. As a result, this research also explored modified versions of the 3D UNet architecture incorporating Vision Transformer (ViT), specifically 3D UNet Transformer (UNETR) [2] and 3D Swin-UNet Transformer [3].

Upon evaluating these three algorithmic models, their performance was assessed by computing dice scores, which exhibited significant variation. Furthermore, it was concluded that further iterations are needed to achieve optimal model performance. The dice scores for the 3D Swin-UNet Transformer, 3D UNETR, and 3D UNet models were determined to be 95.62%, 76.34%, and 72.88%, respectively. Consequently, it was

ascertained that the 3D Swin-UNet transformer model delivered the best performance in the context of 3D liver image segmentation.

The outcomes of this study offer a comparative analysis of the effectiveness and performance of deep learning algorithms, specifically the 3D UNet and its iterations incorporating vision transformers, in the domain of liver segmentation. This segmentation model may serve as a valuable tool to streamline the liver segmentation process, facilitating its use as a foundation for medical professionals to observe liver conditions.

**Acknowledgment:** The authors are grateful to the Research Center for AI and Big Data Universitas Padjadjaran and the Directorate for Research and Community Service (DRPM) Ministry of Research, Technology, and Higher Education Indonesia which supports this research under Higher Education Research Grant no. 094/E5/PG.02.00.PT/2022.

**References**:

[1] A. Gotra et al., "Liver segmentation: indications, techniques and future directions," Insights Imaging, vol. 8, no. 4, pp. 377–392, Aug. 2017, doi: 10.1007/s13244-017-0558-1.
[2] A. Hatamizadeh et al., "UNETR: Transformers for 3D Medical Image Segmentation," Mar. 2021.
[3] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. Roth, and D. Xu, "Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images," Jan. 2022.