UNIVERSITAS
GADJAH MADA
3D Lung Cancer Segmentation: A Study of CNN and Vision Transformer Architectures
Sholy Putri Fauzya, Prof. Ir. Hanung Adi Nugroho, S.T., M.Eng., Ph.D., IPM., SMIEEE. ; Dr.Eng. Ir. Igi Ardiyanto, S.T
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRAK

Kanker paru-paru merupakan salah satu penyebab utama kematian akibat kanker di seluruh dunia. Segmentasi kanker paru pada citra CT scan 3D berperan penting dalam diagnosis dan penanganan dini, namun tetap menjadi tantangan besar dalam analisis citra medis. Meskipun berbagai metode berbasis *Convolutional Neural Network* (CNN) telah banyak diterapkan, pendekatan ini masih menghadapi keterbatasan, seperti rendahnya kemampuan generalisasi, interpretabilitas yang terbatas, dan ketahanan terhadap variasi data yang kurang memadai. Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi alternatif berbasis *Vision Transformer* (ViT) yang lebih baru dan membandingkannya dengan metode CNN dalam segmentasi kanker paru. Penelitian ini menggunakan data dari $Task06\_Lung$.

Model yang dievaluasi meliputi CNN, UNet, ViT, UNetr, dan Swin-UNetr. Semua model dilatih dari awal tanpa menggunakan bobot *pre-trained*, dan evaluasi dilakukan menggunakan *7-fold cross-validation* untuk meningkatkan reliabilitas hasil. Metrik utama yang digunakan adalah *Dice Score*. Hasil penelitian menunjukkan bahwa Swin-UNetr berhasil mencapai kinerja terbaik dengan rata-rata *Dice Score* sebesar 0.73 ± 0.08, mengungguli model lainnya. UNetr mengikuti dengan *Dice Score* sebesar 0.69 ± 0.08, sedangkan UNet mencapai 0.66 ± 0.03. Model ViT memperoleh performa lebih rendah sebesar 0.54 ± 0.03, dan model CNN mendapatkan skor terendah yaitu $0.50 \pm (0.01 \times 10^{-2})$.

Hasil ini diperkuat oleh analisis hasil prediksi segmentasi, dimana Swin-UNetr menghasilkan segmentasi yang paling akurat dan menyerupai *ground-truth*, dengan batas tumor yang presisi. Sebaliknya, model CNN dan ViT menunjukkan segmentasi yang kurang akurat. Pengujian hipotesis yang dilakukan menunjukkan bahwa seluruh hipotesis penelitian diterima pada tingkat signifikansi 5%. Hal ini membuktikan bahwa model berbasis ViT secara statistik menghasilkan *Dice Score* yang lebih tinggi dibandingkan CNN, bahwa model *hybrid* seperti UNetr dan Swin-UNetr secara signifikan mengungguli arsitektur berbasis CNN, dan bahwa Swin-UNetr menunjukkan performa terbaik diantara semua model yang dievaluasi.

Ukuran dataset yang relatif kecil serta pelatihan tanpa bobot *pre-trained* menyebabkan keterbatasan dalam kemampuan generalisasi model dan rentan terhadap *overfitting*. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan model *pre-trained*, memperbesar jumlah dan variasi data, menerapkan teknik augmentasi data yang lebih luas, serta mengeksplorasi arsitektur *hybrid* yang lebih efisien. Secara keseluruhan, penelitian ini menyimpulkan bahwa pendekatan berbasis ViT, khususnya Swin-UNetr, memiliki potensi besar dalam meningkatkan akurasi dan efektivitas segmentasi kanker paru. Hasil ini diharapkan dapat menjadi dasar untuk mengembangkan sistem analisis kanker paru yang lebih baik dan aplikatif.

Kata kunci—Kanker Paru, Segmentasi, Uji Komparasi, CNN, ViT

UNIVERSITAS
GADJAH MADA
3D Lung Cancer Segmentation: A Study of CNN and Vision Transformer Architectures
Sholy Putri Fauzya, Prof. Ir. Hanung Adi Nugroho, S.T., M.Eng., Ph.D., IPM., SMIEEE. ; Dr.Eng. Ir. Igi Ardiyanto, S.T
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Lung cancer is one of the leading causes of cancer-related mortality worldwide. The segmentation of lung cancer in 3D CT scan images plays a crucial role in early diagnosis and treatment planning; however, it remains a major challenge in medical image analysis. Although various methods based on Convolutional Neural Networks (CNN) have been widely applied, these approaches still face limitations, such as low generalization ability, limited interpretability, and insufficient robustness to data variations. Therefore, this study explores a newer alternative based on the Vision Transformer (ViT) architecture and compares it with CNN methods for lung cancer segmentation. This study utilizes data from the $Task06_Lung$ dataset.

The models evaluated include CNN, UNet, ViT, UNetr, and Swin-UNetr. All models were trained from scratch without pre-trained weights, and evaluation was performed using 7-fold cross-validation to improve result reliability. The primary metric used was the Dice Score. The results showed that Swin-UNetr achieved the best performance with an average Dice Score of 0.73 ± 0.08, outperforming the other models. UNetr followed with a Dice Score of 0.69 ± 0.08, while UNet achieved 0.66 ± 0.03. The ViT model obtained a lower performance of 0.54 ± 0.03, and the CNN model recorded the lowest score of $0.50 \pm (0.01 \times 10^{-2})$.

These findings were supported by a qualitative analysis, where Swin-UNetr produced segmentation results most closely resembling the ground truth, with precise tumor boundary delineation. In contrast, CNN and ViT models produced less accurate segmentations. Hypothesis testing confirmed that all research hypotheses were accepted at the 5% significance. This demonstrates that ViT-based models statistically achieved higher Dice Scores than CNN models, that hybrid models such as UNETR and Swin-UNetr significantly outperformed CNN-based architectures, and that Swin-UNetr exhibited the best performance among all evaluated models.

However, the relatively small dataset size and the training of models from scratch without pre-trained weights led to limitations in generalization ability and increased susceptibility to overfitting. Therefore, future studies should utilize pre-trained models, expand the dataset size and diversity, apply more extensive data augmentation techniques, and explore more efficient hybrid architectures. Overall, this study concludes that ViT-based approaches, particularly Swin-UNetr, have great potential to improve the accuracy and effectiveness of lung cancer segmentation. These findings will be a foundation for developing more advanced and practical lung cancer analysis systems.

Keywords—Lung Cancer, Segmentation, Comparative Study, CNN, Vision Transformer