

Recent advances in deep learning for lymphoma segmentation: Clinical applications and challenges

Wanru Liang¹ , Feiyang Yang² , Peihong Teng³, Tianyang Zhang⁴ 
and Weizhang Shen¹ 

Abstract

Lymphoma is a prevalent malignant tumor within the hematological system, posing significant challenges to clinical practice due to its diverse subtypes, intricate radiological and metabolic manifestations. Lymphoma segmentation studies based on positron emission tomography/computed tomography (PET/CT), CT, and magnetic resonance imaging represent key strategies for addressing these challenges. This article reviews the advancements in lymphoma segmentation research utilizing deep learning methods, offering a comparative analysis with traditional approaches, and conducting an in-depth examination and summary of aspects such as dataset characteristics, backbone networks of models, adjustments to network structures based on research objectives, and model performance. The article also explores the potential and challenges of translating deep learning-based lymphoma segmentation research into clinical scenarios, with a focus on practical clinical applications. The future research priorities in lymphoma segmentation are identified as enhancing the models' clinical generalizability, integrating into clinical workflows, reducing computational demands, and expanding high-quality datasets. These efforts aim to facilitate the broad application of deep learning in the diagnosis and treatment monitoring of lymphoma.

Keywords

Lymphoma, medical image segmentation, clinical application, deep learning, artificial intelligence

Received: 25 March 2025; accepted: 7 July 2025

Introduction

Lymphoma is a malignant tumor that occurs in the lymphatic system and is one of the most common malignancies of the hematological system globally.¹ Lymphoma exhibits complex classification (Figure 1), with each subtype demonstrating distinct biological behaviors and clinical manifestations. Due to the widespread distribution of the lymphatic system throughout the body, lymphomas can occur in various locations, including lymph nodes, spleen, liver, bone marrow, and other organs.² The heterogeneity of their radiological presentations further complicates matters; even within the same lymphoma type, differences in imaging characteristics such as size, shape, density, or signal intensity may exist. Imaging examinations serve as the primary noninvasive method for lymphoma assessment. However, the variability in radiological features poses significant challenges to the manual evaluation and interpretation of lymphoma images.³ Furthermore, the

lack of objective evaluation criteria results in assessment outcomes being influenced by subjective factors of the evaluating physicians. Therefore, research into computer-assisted lymphoma segmentation methods has emerged to address these issues.

¹Department of Hematology and Oncology, The Second Hospital of Jilin University, Changchun, China

²College of Computer Science and Technology, Jilin University, Changchun, China

³Department of Radiology, China-Japan Union Hospital of Jilin University, Changchun, China

⁴Department of General Surgery Center Hepatobiliary and Pancreatic Surgery, The First Hospital of Jilin University, Changchun, China

Corresponding author:

Weizhang Shen, Department of Hematology and Oncology, The Second Hospital of Jilin University, Yatai Street, Changchun, 130012, Jilin Province, China.

Email: shenweizhang2015@126.com



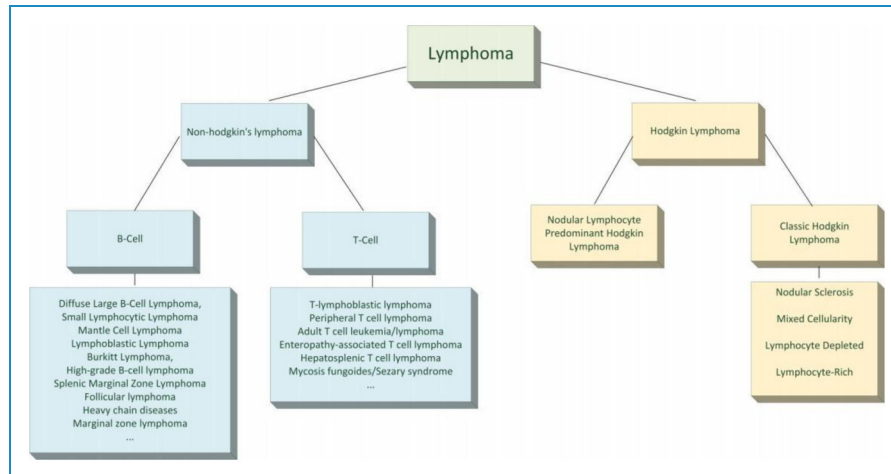


Figure 1. Lymphoma typing and subtype schematic diagram.

Medical image segmentation is a fundamental technique in image analysis that aims to delineate specific anatomical or pathological structures from the background in medical images.⁴ In the context of lymphoma, segmentation plays a pivotal role not just by outlining lesions or tumors, but more importantly by enabling robust and reproducible extraction of clinically meaningful quantitative features. Among these, total metabolic tumor volume (TMTV) has emerged as a key biomarker—crucial for staging, risk stratification, treatment planning, and response monitoring. The ability to automatically and consistently quantify such indicators underscores segmentation’s value as a foundational tool for downstream clinical decision making, rather than an end goal itself. Recent studies, such as Albano et al.,⁵ have further emphasized the clinical relevance of TMTV and the necessity of accurate segmentation in achieving it. Additionally, segmentation supports treatment planning (e.g. defining radiation fields or surgical margins), tracking disease progression, assessing therapeutic outcomes, and informing prognosis.⁶ Earlier approaches to lymphoma segmentation primarily relied on traditional machine learning methods, including thresholding based on pixel intensity, edge detection to identify boundary changes,⁷ region-growing algorithms that expand from seed points based on predefined similarity criteria,⁸ and clustering techniques that group pixels by feature similarity.⁹ These methods offer interpretable models with relatively modest data and computational demands, and clear decision logic that aids understanding.¹⁰ However, they depend heavily on handcrafted features and expert intervention, introducing subjectivity and potential variability across institutions and clinician expertise levels, which may lead to inconsistent segmentation results and limit reproducibility in clinical settings. In contrast, deep learning has increasingly become the dominant paradigm in lymphoma segmentation due to several key advantages beyond those already recognized—such as

higher efficiency and end-to-end automation. Deep learning models are trained on large-scale, high-quality annotated datasets, enabling them to learn complex visual patterns and clinical decision strategies from experienced clinicians. This allows for encapsulation of expert-level knowledge and promotes more consistent, reproducible outcomes in practice.^{7,11} Moreover, unlike traditional approaches that rely on manual feature engineering and require substantial human involvement in multiple stages, deep learning enables direct learning from raw imaging data, supporting fully automated pipelines and achieving superior performance under standardized, operator-independent evaluation settings, as evidenced by results from international benchmarks such as BraTS,¹² HECKTOR,¹³ and AutoPET.¹⁴ Deep learning also facilitates automatic, hierarchical, and task-specific feature learning—a process known as automatic feature learning—which allows models to optimize feature representations in a data-driven manner. Empirical evidence confirms that such learned features are often more informative and discriminative than handcrafted alternatives, especially in complex medical imaging tasks. For instance, Shen et al.¹⁵ reviewed deep learning applications in medical image analysis and highlighted the consistent superiority of Convolutional Neural Network (CNN)-derived features over manually designed ones across classification, detection, and segmentation. Similarly, Litjens et al.¹⁶ concluded that deep features extracted by convolutional networks generally outperform traditional features across diverse imaging modalities and pathological conditions. Notably, Isensee et al.¹⁷ demonstrated that the nnU-Net framework achieved state-of-the-art results in 23 different biomedical segmentation challenges without requiring any manual feature engineering. These findings collectively demonstrate that deep learning-derived features are not only more efficient to obtain—eliminating the need for time-consuming feature design—but also more effective

in capturing high-level semantic information essential for accurate medical image interpretation.

The purpose of lymphoma segmentation research based on medical images is to serve clinical applications. Application-oriented deep learning methods in lymphoma segmentation research aim to integrate deep learning models into clinical workflows and application scenarios. This approach also addresses issues of model interpretability and generalization in clinical settings, thereby compensating for the shortcomings of deep learning methods in this regard.¹⁸ To summarize various studies comprehensively, we conducted searches on Google Scholar and Web of Science using keywords “lymphoma segmentation” and “deep learning” to retrieve recent literature. We screened 45 studies, excluding non-English articles, studies outside the scope of deep learning-based segmentation, studies not related to lymphoma segmentation, and duplicate articles. We included studies published from 2018 to 2024, focusing on methods tested and validated, written in English, specifically addressing lymphoma segmentation using deep learning methods, ensuring all results were verified. Different from existing reviews, this review not only reviews the latest advances, strengths, and weaknesses of lymphoma segmentation from a deep learning perspective but also compares and summarizes relevant methods. Furthermore, it explores this issue from a clinical application perspective. The structure of this article includes an introduction section, a comprehensive overview of deep learning technology in the second section, a review of recent advances in deep learning applied to lymphoma segmentation in the third section, a discussion and outlook on clinical applicability and technical limitations in the fourth section, and finally, a summary of the entire article in the fifth section.

Fundamentals of deep learning

Introduction to deep learning

Deep learning methods are a class of algorithms inspired by the structure of the human brain, which construct multilayer neural networks to learn complex representations of data.¹⁹ These methods have achieved significant successes in various fields such as image processing^{20,21} natural language processing,²² and speech recognition.²³ In the field of medical image segmentation, deep learning methods particularly demonstrate their powerful capabilities. Deep learning models learn high-level features of data through their multilayered structures. Lower layers may learn basic features like edges, while higher layers may capture shapes, textures, and even semantic information. Neurons in each layer perform nonlinear transformations on input data, enabling the network to capture complex patterns.¹⁵ Compared to traditional machine learning methods, deep learning models automate feature extraction, reducing the

need for manual feature engineering. These models use backpropagation algorithms to compute gradients of the loss function with respect to network parameters, and optimization algorithms like gradient descent to update network weights. Deep learning models typically require large amounts of data for training to learn generalized feature representations. The application of deep learning methods in medical image segmentation continues to expand, enhancing the automation level and improving the accuracy and reliability of segmentation results.

Fundamentals of Convolutional Neural Networks

The CNN is a landmark model in deep learning that combines deep learning and image processing techniques, achieving significant advancements in image analysis.²⁴ Similarly, in medical image segmentation, CNN serves as the backbone network for the majority of segmentation models. CNN reduces the number of parameters and enhances training efficiency through weight sharing and leveraging spatial relationships, operating as a supervised learning model.^{24,25} Since the concept of visual receptive fields was introduced by Hubel et al.²⁶ in 1962, followed by Fukushima’s proposal of the Neocognitron based on receptive fields in 1980,²⁷ and LeCun et al.’s²⁸ development of LeNet5 using the backpropagation algorithm in 1998, CNN has undergone a progression from theoretical exploration to experimental development. The introduction of AlexNet²⁹ marked CNN’s breakthrough in the ImageNet competition in 2012, solidifying its core position in computer vision research and continuously driving further advancements in the field.

Advancements of deep learning in lymphoma segmentation research

In the field of medical image segmentation for lymphoma, specific tasks rely predominantly on deep learning models.

Lymphoma imaging encompasses diverse modalities that differ significantly from natural images, based on various physical properties and sources of energy. The resulting images span multiple scales. Therefore, training a unified multiscale base model from these multimodal images is currently impractical. All studies included in this article focus on specialized segmentation models for lymphoma. Next, we will elaborate on dataset characteristics, network structures, supervision methods, and model performance. The relevant content has also been summarized in Tables 1 and 2.

Dataset

In lymphoma segmentation research based on deep learning methods, the critical initial step that determines the direction and content of the study is the construction of the

Table 1. Basic information of lymphoma segmentation research based on deep learning.

Reference	Year	Network	Modalities	Supervision	Dimension	Dataset
Jemaa et al. ³⁰	2020	U-Net and V-Net	PET/CT	Supervision	3D	Private dataset
Alderuccio et al. ³¹	2022	ResU-Net	PET/CT	Supervision	2D	Private dataset
Nam et al. ³²	2024	ResNet	CT	Supervision	3D	Private dataset
Jiang et al. ³³	2022	nnU-Net	PET/CT	Supervision	3D	Private dataset
Yousefirizi et al. ³⁴	2022	3D U-Net	PET/CT	Supervision	3D	Private dataset
Liu et al. ³⁵	2022	3DU-Net	PET/CT	Supervision	3D	Private dataset
Constantino et al. ³⁶	2023	3D U-Net	PET/CT	Supervision	3D	Private dataset
Yousefirizi et al. ³⁷	2024	3D U-Net	PET/CT	Semisupervision	3D	Private dataset
Huang et al. ³⁸	2022	V-Net	PET/CT	Weak-supervision	3D	Private dataset
Jurdi et al. ³⁹	2024	U-Net	MR	Supervision	2D	Private dataset
Luo et al. ⁴⁰	2023	EfficientNet+U-Net	PET/CT	Supervision	2D	Private dataset
Liu et al. ⁴¹	2022	U-Net	PET/CT	Supervision	2D	Private dataset
Zhai et al. ⁴²	2022	CBS-APU-Net	PET/CT	Supervision	2D	Private dataset
Zhu et al. ⁴³	2023	CGBO-Net	PET/CT	Supervision	2D	HECKTOR
Pennig et al. ⁴⁴	2021	DeepMedic	MR	Supervision	3D	Private dataset
Yuan et al. ⁴⁵	2021	Hybrid Learning Component+CNNs	PET+CT	Supervision	2D	Private dataset
Hu et al. ⁴⁶	2020	ResU-Net	PET/CT	Supervision	2D	Private dataset
Diao et al. ⁴⁷	2023	CSAE-Net+PFAS-Net	PET+CT	Supervision	2D	Private dataset
Xu et al. ⁴⁸	2020	DeepLabv3	PET/CT	Supervision	2D	Private dataset
Weisman et al. ⁴⁹	2020	DeepMedic	MR+PET/CT	Supervision	3D	Private dataset
Li et al. ⁵⁰	2019	DenseNet	PET/CT	Supervision	2D	Private dataset
Huang et al. ⁵¹	2021	U-Net	PET/CT	Supervision	2D	Private dataset
Diao et al. ⁵²	2021	EFNet+U-Net	PET+CT	Supervision	3D	Private dataset
Huang et al. ⁵³	2022	EFNet+U-Net	PET+CT	Supervision	3D	Private dataset
Conte et al. ⁵⁴	2021	U-Net+GAN	MR	Unsupervision in GAN	2D	BRATS
Thiery et al. ⁵⁵	2023	U-Net+GNN	PET/CT	Supervision	2D	Private dataset
Chen et al. ⁵⁶	2022	HAFS-Net	PET+CT	Supervision	2D	Private dataset
Wang et al. ⁵⁷	2021	U-Net	PET/CT	Supervision	2D	Private dataset

(continued)

Table 1. Continued.

Reference	Year	Network	Modalities	Supervision	Dimension	Dataset
Tie et al. ⁵⁸	2024	3D swinU-Net	PET/CT	Supervision	3D	AutoPET
Wang et al. ⁵⁹	2023	Memory-HDRDS-UNet	PET/CT	Supervision	2D	Private dataset
Durand et al. ⁶⁰	2021	3D U-Net	PET/CT	Supervision	3D	Private dataset
Klimont et al. ⁶¹	2023	nnU-Net	CT	Supervision	3D	Private dataset
Lu et al. ⁶²	2023	nnU-Net	MR	Supervision	3D	Private dataset
Fu et al. ⁶³	2024	nnU-Net	MR	Supervision	3D	Private dataset
Wang et al. ⁶⁴	2022	PSR-Net	PET/CT	Supervision	2D	Private dataset
Ahamed et al. ⁶⁵	2022	Cascaded U-Net	PET/CT	Supervision	2D	Private dataset
Wang et al. ⁶⁶	2022	U-Net	PET/CT	Supervision	2D	Private dataset
Yousefirizi et al. ⁶⁷	2024	Cascaded U-Net	PET/CT	Supervision	2D	Private dataset
Huang et al. ⁶⁸	2024	MMCAF+CNNs	PET/CT	Supervision	3D	Private dataset
Shi et al. ⁶⁹	2023	U-Net+GAN	PET/CT	Unsupervision in GAN	2D	Private dataset
Zhou et al. ⁷⁰	2021	U-Net	PET/CT	Supervision	2D	Private dataset
Kuker et al. ⁷¹	2022	ResU-Net	PET/CT	Supervision	2D	Private dataset
Pang et al. ⁷²	2024	U-Net	PET/CT	Supervision	2D	Private dataset
Revailler et al. ⁷³	2022	V-Net	PET/CT	Supervision	3D	Private dataset
Zhao et al. ⁷⁴	2024	nnU-Net	MR	Supervision	3D	Private dataset

2D: two-dimensional; 3D: three-dimensional; CT: computed tomography; PET: positron emission tomography.

dataset. A high-quality dataset forms the foundation for training effective models.⁷⁵ Firstly, dataset diversity and balance are crucial. The dataset needs to include images of different types and stages of lymphoma to ensure the model generalizes well across various scenarios. Some types of lymphoma may be more common than others, necessitating balance in the dataset. Secondly, annotation poses a unique challenge. Unlike natural image segmentation datasets, lymphoma segmentation annotations typically involve pixel or voxel-level annotations rather than scribble annotations or region of interest annotations. This is due to the high heterogeneity inherent in lymphoma imaging, which also exists within subtypes of the same lymphoma type. Accurate annotation serves as crucial supervision signal for training deep learning models. Errors or ambiguous annotations can lead to incorrect feature learning, affecting segmentation accuracy.⁷⁶ Thirdly, dataset scale is advantageous compared to traditional machine learning methods. Deep learning methods excel with large-scale datasets. Sufficient data volume helps models learn features better,

enhancing their generalization capability. Insufficient data may lead to overfitting, where the model performs well on training data but poorly on new, unseen data.⁷⁷ Finally, constructing multimodal image datasets and multi-task datasets including different information achieves the goals of segmentation models, such as lymphoma identification, prognosis prediction, and treatment monitoring. These datasets also contribute significantly to interpretable deep learning models.^{7,11}

Among the 45 studies included in this article, the data modalities encompass positron emission tomography/computed tomography (PET/CT) with multimodal fusion and multitemporal data, different sequences of magnetic resonance imaging (MRI) data, and contrast-enhanced CT data. The proportions of the various modes are shown in Figure 2. The selection of different data modalities depends on specific medical issues in lymphoma segmentation research. PET/CT combines functional and structural imaging, providing information on lesion metabolism and anatomical location, playing a crucial role in lymphoma

Table 2. Performance for lymphoma segmentation research based on deep learning.

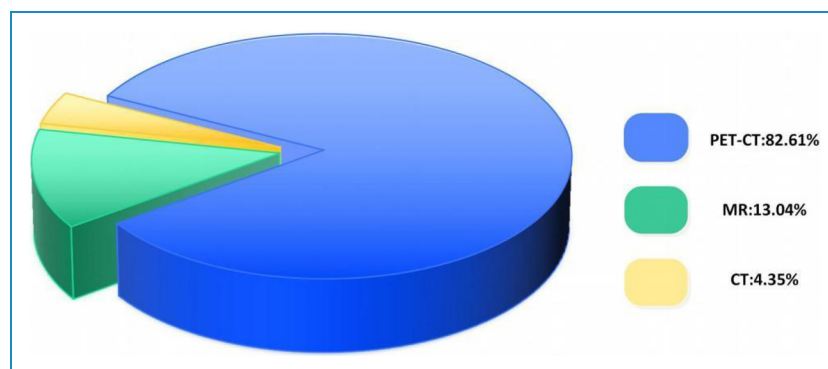
Reference	Performance
Jemaa et al. ³⁰	DSC=88.6%; Sensitivity=93.0%
Alderuccio et al. ³¹	MSE=0.547
Nam et al. ³²	DSC=39.90%; Precision=60.9%
Jiang et al. ³³	DSC=0.78; Jaccard=0.69; Sensitivity=0.81; HD=0.82
Yousefirizi et al. ³⁴	DSC=0.77; HD=0.16; MAD=3.929
Liu et al. ³⁵	DSC =86.8%
Constantino et al. ³⁶	DSC=87.0%
Yousefirizi et al. ³⁷	DSC=69.0%
Huang et al. ³⁸	DSC=75.2%
Jurdi et al. ³⁹	DSC=65.0%
Luo et al. ⁴⁰	DSC=78.1%; Jaccard=64.9%
Liu et al. ⁴¹	DSC=75.4%
Zhai et al. ⁴²	DSC=81.89%; Jaccard=73.71%; Accuracy=91.26%; HD=5.62
Zhu et al. ⁴³	DSC=90.7%; Precision=89.4%; Jaccard=83.1%
Pennig et al. ⁴⁴	Sensitivity=85.0%
Yuan et al. ⁴⁵	DSC=73.03%
Hu et al. ⁴⁶	DSC=66.64%; Precision=71.84%; Jaccard=53.05%
Diao et al. ⁴⁷	DSC=79.81%; Precision=75.1%; Specificity=99.90%
Xu et al. ⁴⁸	DSC=74%; Sensitivity=81%
Weisman et al. ⁴⁹	Sensitivity=85%
Li et al. ⁵⁰	DSC=72.84%
Huang et al. ⁵¹	DSC=72.6%
Diao et al. ⁵²	DSCsts = 78%;Sensitivitysts = 85%;Precisionsts = 75%
Huang et al. ⁵³	DSC=84.6%
Conte et al. ⁵⁴	DSC=82.0%; MSE=0.006
Thiery et al. ⁵⁵	AUC=0.72
Chen et al. ⁵⁶	DSC=87.75%; Sensitivity=88.94%; Specificity=99.93%
Wang et al. ⁵⁷	DSC=85.58%; Sensitivity=94.63%
Tie et al. ⁵⁸	DSC=77.2%

(continued)

Table 2. Continued.

Reference	Performance
Wang et al. ⁵⁹	DSC=86.90%; Sensitivity=96.16%
Durand et al. ⁶⁰	DSC=0.73; Sensitivity=0.75; Specificity=0.79; Jaccard=0.68
Klimont et al. ⁶¹	DSC=0.81
Lu et al. ⁶²	Sensitivity = 95.7%;DSC _{ctv} = 0.76;DSC _c = 0.73
Fu et al. ⁶³	Δ DSC = 0.19; Δ HD = 8
Wang et al. ⁶⁴	DSC=90.94%; Sensitivity=87.18%
Ahamed et al. ⁶⁵	AUC=0.92
Wang et al. ⁶⁶	DSC=70.4%
Yousefirizi et al. ⁶⁷	DSC=0.66
Huang et al. ⁶⁸	DSC=86.90%; HD=2.71; Sensitivity=94.62%; Specificity=99.86%
Shi et al. ⁶⁹	DSC=86.09%; Precision=81.08%; MAE=10.718
Zhou et al. ⁷⁰	Sensitivity _s = 0.71;Sensitivity _b = 0.84
Kuker et al. ⁷¹	MSE=54.7
Pang et al. ⁷²	DSC=86.67%; Jaccard=69.23%
Revailler et al. ⁷³	DSC _{total} = 83.6%;DSC _{chl} = 70%;DSC _{fl} = 68%;DSC _{diblcl} = 93%
Zhao et al. ⁷⁴	DSC=93.49%

AUC: area under the curve; DSC: dice similarity coefficient; HD: Hausdorff distance; MAD: mean absolute deviation; MAE: mean absolute error; MSE: mean squared error.

**Figure 2.** The proportion of different modal data in lymphoma segmentation studies.

diagnosis, staging, treatment evaluation, and prognosis assessment.⁷⁸ Consequently, it is the primary modality type in this type of research dataset, comprising 38 studies. The majority of lymphoma types included are non-Hodgkin lymphomas, with a smaller portion being Hodgkin

lymphomas. The lymphoma subtypes primarily studied include diffuse large B-cell lymphoma (DLBCL), primary mediastinal large B-cell lymphoma (PMBCL), mantle cell lymphoma, and follicular lymphoma, all clinically prevalent. The anatomical coverage in these PET/CT datasets

varies; most cover the entire body excluding the head and neck. Two studies focus on the chest and mediastinum, one on nasopharynx, and one on abdominal organs. Another data modality is different sequences of MRI data. Unlike PET/CT, MRI offers higher soft tissue contrast and multiplanar imaging capabilities, addressing PET/CT's limitations in evaluating central nervous system lymphomas.⁷⁹ Six studies utilize this data modality, focusing mainly on PMBCL and primary central nervous system lymphoma (PCNSL). The third data modality is contrast-enhanced CT, which, due to its limited scanning range, primarily provides anatomical information and has limitations for lymphomas that can manifest throughout the body. It is an unconventional clinical assessment method for lymphoma but is chosen for its low radiation and rapid imaging characteristics, making it suitable for assessing lymphomas in children.⁸⁰ Two studies adopt this data modality, both sourced from pediatric lymphoma patients. Due to the ethical and privacy considerations of medical data, most high-quality datasets used are private. The openly accessible lymphoma datasets include the HECKTOR challenge dataset,¹³ the BraTS dataset,⁸¹ and AutoPET dataset.¹⁴ The lack of high-quality annotated datasets to some extent constrains the development of lymphoma segmentation research.

HECKTOR challenge dataset. The HECKTOR (Head and Neck Tumor Segmentation) dataset provides annotated PET/CT images for the task of segmenting primary tumors and lymph node metastases in patients with head and neck squamous cell carcinoma (HNSCC).¹³ It includes multimodal imaging data from multiple institutions, reflecting real-world heterogeneity in acquisition protocols and patient anatomy. While the dataset mainly targets HNSCC cases, it also includes a limited number of patients with head and neck lymphoma, expanding its relevance to lymphoproliferative diseases in this anatomical region. This diversity enables research not only on carcinoma segmentation but also on modeling lesion appearance in lymphoma, where nodal involvement may differ in morphology and uptake. The challenge promotes the development of robust, generalizable models capable of handling the complex anatomy and variable lesion characteristics typical of head and neck cancers. Annotations are provided by expert radiologists and nuclear medicine physicians, enabling reliable benchmarking for automated segmentation methods.

BraTS dataset. The Brain Tumor Segmentation (BraTS) dataset is a widely used benchmark for evaluating the segmentation of gliomas from multimodal MRI scans, including T1, T1c, T2, and fluid attenuated inversion recovery sequences.¹² It provides manual segmentations of enhancing tumor, peritumoral edema, and necrotic core, supporting detailed subregion-level analysis. In addition to gliomas, the dataset has also been used in studies

involving PCNSL cases, which may exhibit imaging patterns overlapping with high-grade gliomas but differ in biological behavior and treatment response. The inclusion of such cases broadens the dataset's applicability to lymphoid malignancies in the brain and enables comparative evaluation of segmentation methods across tumor types. The dataset spans several editions and includes both preoperative and postoperative scans, with expert annotations curated for consistency. BraTS has driven innovation in brain tumor segmentation by encouraging methodological comparisons under standardized settings.

AutoPET dataset. The AutoPET dataset consists of whole-body fludeoxyglucos (FDG)-PET/CT images with expert-annotated lesion segmentations from a wide range of oncological cases.¹⁴ It is specifically designed to foster research on fully automated, end-to-end segmentation of metabolically active tumors. The dataset includes diverse malignancies, with multiple cases of lymphoma, making it especially relevant for developing models aimed at quantifying tumor burden in systemic cancers. AutoPET presents significant variability in lesion location, shape, and uptake patterns, posing challenges that mimic clinical reality. It supports the development of models aimed at calculating key quantitative biomarkers, such as TMTV, with potential applications in disease monitoring, treatment planning, and prognostic evaluation across cancer types.

Lymphoma segmentation network architecture

The network architecture is the core component of deep learning models, and its importance in lymphoma segmentation research is reflected in its impact on feature extraction, spatial resolution preservation, handling of multiscale and multimodal data, computational efficiency, generalization, and robustness.¹⁹ Proper selection and design of network architecture can significantly enhance segmentation performance and the feasibility of clinical applications. This article specifically discusses the composition of network architecture in terms of backbone network selection and strategies for network structure adjustment.

Classification of backbones. In the network structure of deep learning models, "backbone" refers to the core structure that constitutes the main framework of the network. It is responsible for extracting features from input data, which serve as the basis for subsequent tasks. The backbone also plays a crucial role in achieving computational efficiency, automation, and scalability of the model.⁸² In lymphoma segmentation research, all utilized backbones are based on CNN architectures and their variants. In this section, we will summarize and classify the backbones used in the 45 studies included in our analysis, providing specific explanations for each (Figure 3).

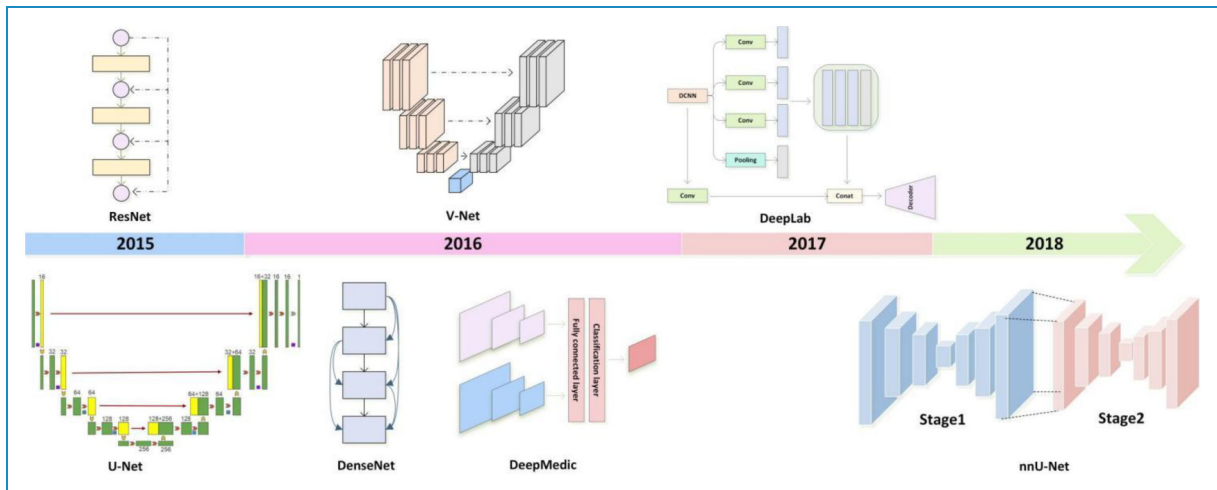


Figure 3. The main backbone species in lymphoma segmentation studies.

ResNet. ResNet, short for Residual Network, was proposed by He et al.⁸³ from Microsoft Research in 2015. It achieved first place in both classification and object detection tasks in the ImageNet competition of that year. The development of ResNet primarily aimed to address the issues of vanishing and exploding gradients that arise with increasing network depth. At the core of ResNet is the residual block, each of which contains two paths: a normal path of convolution layers and an identity shortcut connection that adds the input directly to the output. This design allows the network to learn identity mappings, where the input is propagated directly to the output without undergoing complex transformations. ResNet’s network structure typically consists of stacking multiple residual blocks, enabling the construction of networks of varying depths such as ResNet-50,⁸⁴ ResNet-101,⁸⁵ etc. By using residual connections, gradients can propagate more effectively through the network, thereby mitigating the vanishing gradient problem during training of deep networks. This capability allows for the construction of deeper network architectures, enhancing the model’s learning capacity. The design of residual blocks enables the network to learn more complex feature representations. Due to these characteristics, ResNet has been widely adopted as a backbone network in various applications, including lymphoma segmentation research. At present, ResNet continues to demonstrate strong performance and stability in various tasks due to its modular design, ease of optimization, and compatibility with other architectural innovations such as attention mechanisms and multiscale processing. However, its limitations include relatively high computational cost compared to lightweight alternatives, lack of adaptability to non-Euclidean data or dynamic input structures, and less efficiency in capturing global context compared to newer transformer-based models. Despite these drawbacks, ResNet remains a foundational and versatile architecture, particularly valuable

when combined with other modern techniques in medical image analysis.

U-Net. U-Net is a CNN architecture specifically designed for medical image segmentation tasks in deep learning. Originally proposed by Ronneberger et al. in 2015,⁸⁶ U-Net was developed to address the challenges of biomedical image segmentation where precise localization of anatomical or pathological regions is critical. Its architecture features a characteristic “U” shape composed of a contracting (encoder) path and an expansive (decoder) path. The encoder path captures context by progressively reducing spatial resolution and increasing feature channels, while the decoder path restores resolution through upsampling, enabling accurate delineation of segmentation boundaries. A key innovation in U-Net is the use of skip connections that link corresponding layers in the encoder and decoder, allowing fine-grained spatial information lost during down-sampling to be preserved and reintegrated during reconstruction. This design enhances the model’s ability to handle noisy inputs and maintain fine structural details—an essential capability in medical imaging tasks. Initially implemented as a two-dimensional (2D) architecture, U-Net processed slice-by-slice inputs and was widely applied to planar imaging modalities. However, this approach had limitations in capturing volumetric contextual information inherent in three-dimensional (3D) medical imaging data such as PET/CT scans used in lymphoma studies. To address this, researchers extended U-Net to 3D by replacing 2D convolutions and operations with their 3D counterparts, enabling the model to learn from volumetric input data directly. This transition from 2D U-Net to 3D U-Net⁸⁷ significantly improved performance in volumetric segmentation tasks, particularly for diseases like lymphoma where lesions exhibit irregular shapes, heterogeneous uptake, and complex spatial distributions in PET/CT

volumes. By leveraging the full contextual information across slices, 3D U-Net architectures are better suited to accurately identify metabolically active regions, leading to more consistent and clinically relevant segmentation results. Furthermore, the open and modular design of U-Net makes it highly adaptable—researchers can modify the architecture by incorporating attention mechanisms, residual connections, or multiscale feature fusion modules to enhance task-specific performance. In lymphoma segmentation, such adaptations have been instrumental in improving delineation accuracy, especially in challenging cases involving small or low-contrast lesions. At present, U-Net retains several significant advantages, including architectural simplicity, flexibility for customization, and strong baseline performance across various segmentation tasks. Its extensive use and open-source implementations have fostered reproducibility and accelerated development in the medical imaging community. However, limitations remain: standard U-Net variants may struggle with long-range contextual dependencies due to their inherently local receptive fields, and the 3D version, while more powerful, imposes substantial memory and computational demands, particularly with high-resolution volumetric data. Additionally, U-Net may require further architectural tuning to address issues such as class imbalance and uncertainty estimation in clinical settings. Despite these challenges, U-Net and its evolving variants continue to form the backbone of state-of-the-art segmentation models, offering a robust and extensible foundation for deep learning applications in lymphoma imaging and beyond.

DenseNet. DenseNet, short for Dense Convolutional Network, was introduced by Huang et al.⁸⁸ from the University of Groningen in 2016. The core idea of DenseNet is feature reuse, achieved by connecting each layer to every other layer in a feed-forward fashion, enhancing the information flow and parameter efficiency of the network. In DenseNet, each layer receives the outputs of all preceding layers as inputs and passes its own output to all subsequent layers. This dense connectivity pattern facilitates more effective information flow within the network. Due to the dense connections, gradients can propagate directly from deeper layers to shallower ones, which helps alleviate the vanishing gradient problem and allows the network to be deeper. By reusing features through dense connections, DenseNet reduces the number of parameters compared to other deep networks, making it more parameter-efficient. After each dense block, DenseNet employs transition layers to reduce the size of feature maps and decrease the number of channels, facilitating the subsequent processing of the network. DenseNet typically replaces traditional fully connected layers with global average pooling layers, which further reduces the number of parameters and enhances the network's robustness to variations in input sizes. At present, DenseNet remains

advantageous for its efficient use of parameters, improved gradient flow, and ability to promote feature reuse, making it suitable for tasks where model compactness and deep supervision are essential. It has shown effectiveness in medical image analysis, including segmentation and classification, particularly when computational resources are limited. However, its limitations include increased memory consumption due to the concatenation of feature maps from all preceding layers, which can lead to high graphics processing unit (GPU) usage during training, especially in deeper variants. Additionally, as the network grows, the feature concatenation can become computationally intensive and may introduce redundancy in the learned representations. Despite these challenges, DenseNet continues to serve as a strong backbone in various deep learning applications, offering a good balance between performance and efficiency, especially in scenarios where deeper models are required without significantly increasing the number of parameters.

V-Net. V-Net is another variant of CNN architecture, first proposed by Milletari et al.⁸⁹ in 2016. Similar to U-Net, V-Net's structure consists of an encoder path and a symmetric decoder path. However, V-Net is designed for 3D images, eliminating the need for slicing the input dataset. It introduces residual structures to accelerate convergence, replaces pooling layers with convolutional layers, and proposes a new objective function based on maximizing the Dice similarity coefficient (DSC). The DSC is a statistical metric used to quantify the spatial overlap between two binary segmentation results, typically the predicted and ground truth regions in medical image analysis, with values ranging from 0 (no overlap) to 1 (perfect agreement). In the encoder path, each stage consists of one to three convolutional layers to extract image features. Residual learning is applied at each stage, where the input feature map undergoes several convolutions and nonlinear activations, and the resulting feature map is added pointwise to the input feature map. After each stage, downsampling is performed using an appropriate stride to reduce the image resolution, halving the size of the feature map and doubling the number of feature channels. Similarly, the decoder path is symmetric to the encoder path, with each stage consisting of one to three convolutional layers. Parameter learning is applied in each stage of the decoding process, followed by transpose convolutions to double the size of the feature map. Finally, $1 \times 1 \times 1$ convolutions produce two feature map outputs of the same size as the input volume, followed by softmax to output probabilities for each voxel belonging to foreground and background. V-Net utilizes skip-connection structures similar to U-Net to gather fine-grained details that may be lost in the compression path, thereby improving segmentation accuracy. In lymphoma segmentation research, V-Net's residual learning enables the model to learn more complex structures in tumor regions. Its

capability with 3D convolutions is well-suited for clinical needs in lymphoma segmentation from multiple angles. Moreover, like U-Net, V-Net is an open backbone. Its modular design makes it easy to understand and modify, allowing researchers to adjust the network structure based on specific application scenarios to achieve better performance and meet diverse requirements. Currently, V-Net holds several advantages: its native 3D design allows it to fully utilize volumetric data such as PET/CT, enhancing spatial coherence in segmentation; the residual learning mechanism improves training efficiency and performance in complex lesion structures; and its Dice-based loss is especially suitable for medical segmentation tasks with class imbalance. However, its limitations include high computational and memory demands due to 3D convolutions and deeper architecture, making it challenging to deploy on resource-constrained hardware. Additionally, although residual connections help alleviate vanishing gradients, they may lead to feature redundancy when the number of layers becomes very large. Despite these constraints, V-Net remains a powerful and widely adopted architecture in 3D medical image segmentation, particularly when precise volumetric delineation is critical.

DeepMedic. DeepMedic is an open-source backbone designed specifically for medical image analysis, particularly excelling in neuroimaging-related tasks. Originally proposed by Kamnitsas et al.,⁹⁰ DeepMedic utilizes 3D CNNs to perform segmentation and analysis of volumetric medical images, aiming to enhance diagnostic precision and inform treatment planning. One of its key contributions lies in its early and effective adoption of 3D CNNs over conventional 2D architectures, enabling more accurate modeling of spatial context and structural continuity in medical data, such as in brain tumor segmentation. Like most modern 3D deep learning segmentation frameworks, DeepMedic employs a patch-based training strategy, wherein large medical volumes are decomposed into smaller subvolumes or patches for model training. This approach addresses the substantial computational and memory demands inherent to full-volume 3D image processing, which typically exceed the capacity of contemporary GPUs. Thus, patch-based training has become a widely adopted and essential technique across 3D segmentation models, not limited to DeepMedic alone. The DeepMedic framework encompasses a modular workflow—including preprocessing, training, validation, and testing—that can be flexibly adapted to diverse imaging modalities and segmentation tasks. Network configurations, optimization strategies, and training parameters can be customized according to specific research or clinical requirements. Owing to its strengths in neuroanatomical analysis, DeepMedic has also been extended to central nervous system lymphoma segmentation tasks, with architectural adjustments made to accommodate the particular challenges of each dataset and

clinical scenario. At present, DeepMedic offers several advantages: its deep multiscale architecture is adept at capturing both fine and coarse features through parallel pathways operating at different resolutions, enhancing segmentation accuracy, especially in small or irregularly shaped lesions; its patch-based strategy allows training on large 3D volumes even with limited GPU memory; and its modular, open-source design promotes adaptability and reproducibility in clinical and research settings. However, DeepMedic also has limitations. The patch-based approach, while memory-efficient, can lead to discontinuities at patch boundaries and requires careful tuning of patch size and sampling strategies. Additionally, the architecture is relatively rigid compared to newer transformer-based or attention-enhanced models, which offer more flexible and globally aware representations. Its reliance on hand-crafted preprocessing pipelines can further reduce generalizability across datasets with differing imaging characteristics. Despite these challenges, DeepMedic remains a robust and well-validated choice for 3D medical image segmentation, particularly in neuroimaging and lymphoma-related applications where volumetric context and precision are critical.

DeepLabv3. DeepLabv3, short for “DeepLab: Deep Labelling for Semantic Image Segmentation,” is a deep learning model designed for semantic image segmentation.⁹¹ The DeepLab series models hold a significant position in the field of image segmentation by integrating deep convolutional neural networks and conditional random fields (CRFs) to enhance segmentation accuracy. Key features of DeepLabv3 include: multiscale feature fusion, which leverages deep convolutional networks (such as ResNet) to extract and integrate features across various levels, enabling the model to recognize both global structures and local details; Atrous (dilated) convolution, which expands the receptive field without increasing the number of parameters or reducing resolution, allowing broader contextual understanding while maintaining spatial precision; an Encoder-Decoder structure that facilitates efficient feature extraction and reconstruction for fine-grained segmentation; a Pyramid Pooling Module that captures contextual information at multiple spatial scales to improve segmentation performance in scenes with complex structures; and CRF-based refinement in the final stage to enhance the spatial coherence of the segmentation outputs. Currently, DeepLabv3 exhibits several advantages: it achieves state-of-the-art accuracy on many benchmark datasets due to its powerful multiscale context aggregation and spatial precision mechanisms; its modular design allows integration with various backbone networks, enhancing adaptability to specific tasks such as lymphoma segmentation in medical imaging. Moreover, the use of atrous convolutions and pyramid pooling equips the model to handle objects of varying sizes and shapes—a critical advantage

in medical contexts where lesion morphology is diverse. However, DeepLabv3 also has limitations. Its architecture is computationally intensive, especially when using high-resolution inputs and deep backbones, which may hinder deployment in real-time or resource-constrained environments. Additionally, its performance is heavily reliant on large annotated datasets, limiting its effectiveness in medical scenarios where expert-annotated data are scarce. The reliance on CRFs, while beneficial for refinement, introduces additional computational overhead and complexity in postprocessing. Furthermore, DeepLabv3 was originally developed for natural images, and direct application to 3D medical data requires significant architectural adaptations. Despite these challenges, DeepLabv3 remains a powerful and widely used model in semantic segmentation, and its core components—especially atrous convolutions and multiscale context fusion—continue to influence the development of advanced segmentation networks for both natural and medical image domains.

nnU-Net. nnU-Net is an adaptive deep learning segmentation framework proposed by Isensee et al.⁹² in 2018. It is based on the original 2D and 3D U-Net architectures and specifically designed for medical image segmentation tasks. The design philosophy of nnU-Net is to automate the entire segmentation process, including preprocessing, network structure selection, hyperparameter tuning, model training, and postprocessing, without the need for manual intervention. Essentially, it is a collection of U-Net models. Each U-Net structure remains largely unchanged, inheriting the robustness of the U-Net network. Depending on different data modalities, nnU-Net adapts different processing strategies; the 3D U-Net handles 3D data, while the structure of multiple U-Net cascades enables the model to handle multimodal and multiviewpoint data. This feature aligns well with the requirements of lymphoma data, making nnU-Net a crucial backbone for lymphoma segmentation research. Additionally, nnU-Net is a modular open-source framework. At present, nnU-Net exhibits several notable advantages: it eliminates the need for manual architecture and hyperparameter engineering, enabling researchers to focus on dataset preparation and model evaluation; it provides strong baseline performance across diverse datasets and tasks, making it a reliable first choice for segmentation problems; and it emphasizes reproducibility and fairness in model evaluation, promoting standardized benchmarking in the medical image analysis community. However, despite these strengths, nnU-Net also has limitations. Its performance is highly dependent on the quality and quantity of annotated training data, which remains a bottleneck in many clinical applications such as lymphoma segmentation. Moreover, although it automates much of the pipeline, it is computationally intensive, requiring substantial time and GPU resources for training, especially on large 3D volumes. Additionally, its automated design may not always be

optimal for very specialized tasks, where domain-specific architectural innovations or training strategies might yield better results. Overall, nnU-Net represents a significant advance in medical image segmentation by streamlining model development and establishing a high-performance, generalizable baseline, but it still faces challenges in efficiency, customization, and adaptation to data-scarce or highly specialized clinical scenarios.

Adjustments of network architecture based on research

Objectives. Adjusting and modifying the network architecture of deep learning models based on research objectives refers to making targeted modifications to the model's network structure according to specific research or application needs, aiming to achieve better performance, higher efficiency, or improved adaptability. These adjustments and changes are made with the goal of effectively addressing specific problems or optimizing particular tasks, rather than altering the network structure arbitrarily or without purpose.

For lymphoma segmentation studies, these adjustments to network structure can be categorized into four main types based on research objectives (Figure 4).

Improving segmentation accuracy. In lymphoma segmentation research, enhancing segmentation accuracy remains a central objective, as it underpins image-based diagnostic assessment, treatment planning, and response monitoring—thereby helping to reduce clinical errors and improve patient outcomes. While histopathological biopsy remains the definitive standard for lymphoma diagnosis, the term “diagnosis” in this study is used in a broader clinical context, encompassing radiological evaluation based on imaging modalities such as PET/CT. In this setting, segmentation plays a pivotal role in visualizing and quantifying tumor burden, thus contributing meaningfully to clinical staging and therapeutic decision making. To support this goal, numerous studies have proposed adjustments to network architectures, which can be categorized into five primary types (Figure 5).

Edge detection. Improving tumor boundary information is an effective approach to enhancing lymphoma segmentation accuracy. Lymphoma boundaries are inherently uncertain, making it difficult to discern changes in various information classes in medical imaging by naked eye alone. Zhu et al.⁴³ proposed the use of a cross-shaped structure guidance and boundary optimization strategy to enhance boundary segmentation accuracy. The cross-shaped structure extraction (CCE) method is based on axial context, extracting the cross-shaped structure from annotations of the largest tumor slice, which serves as additional input to

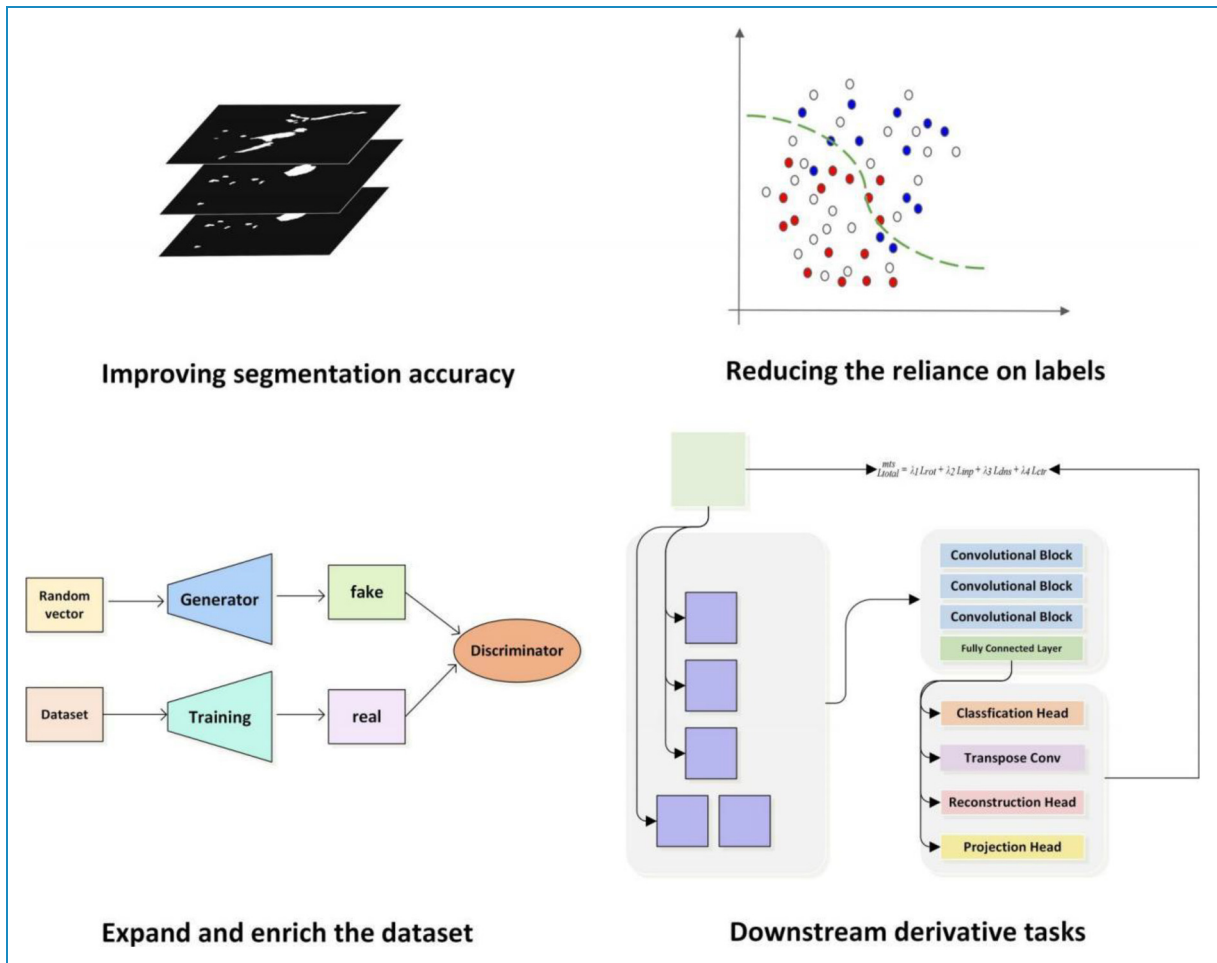


Figure 4. Adjustments of network architecture based on research objectives.

assist in segmenting other tumor slices. The researchers also defined a new boundary gradient change-based loss function. This loss function measures the gradient change between predicted boundary pixels and true boundary pixels, extracting boundary gradients using the Sobel operator and using the mean pixel-wise gradient difference as the loss function value. The Boundary Gradient Change Loss Function collaborates with the true tumor boundaries in the decoder to form a Boundary Optimization Module, supervising boundary segmentation. Luo et al.⁴⁰ integrated the multi-Atlas boundary awareness (MABA) module into the backbone. This module is based on gradient maps, uncertainty maps, and level set maps. The gradient map utilizes gradient information of pixel intensity changes in the image to detect tumor boundaries. In medical images, the junction between tumors and surrounding normal tissue often exhibits significant intensity changes, which can be captured by computing the image gradient. The uncertainty map focuses on uncertain regions between the tumor and surrounding tissues, which may be challenging to classify during segmentation. By analyzing the predicted

probability maps of the network output, the uncertainty map identifies pixels with predicted probabilities close to the classification threshold. The level set map employs a method based on level set evolution to capture the geometric contour and distance information of the tumor. In the level set method, the tumor boundary is viewed as a dynamically evolving curve that gradually approaches the real tumor edge over time. This module focuses on uncertain regions between the tumor and adjacent tissues to obtain potential tumor boundaries. Jurdi et al.³⁹ further optimized their study based on a boundary loss function, the Boundary Irregularity Index (BI). They defined a Dice coefficient-based BI Index to measure dissimilarity between smoothed segmentation maps and true segmentation maps, directly implementing a differentiable and smooth gradient loss function. During training, minimizing the difference between the BI Index and the true BI Index optimizes the network parameters. Thus, the network tends to generate smoother and more regular boundaries during segmentation, thereby improving segmentation accuracy.

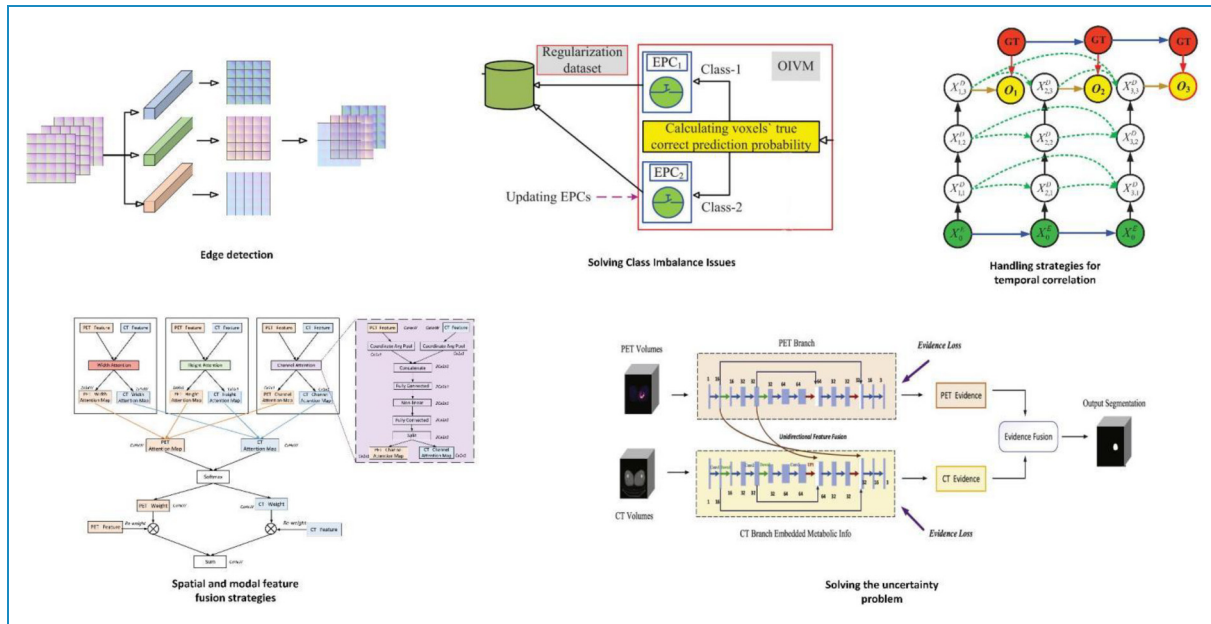


Figure 5. The key components of methods aimed at enhancing segmentation accuracy. In edge detection, Luo et al.⁴⁰ have integrated a Multi-Atlas Boundary Awareness module into the backbone network. This module operates on gradient maps, uncertainty maps, and level set maps. To address class imbalance issues, Wang et al.⁶⁴ introduced the Prior-Shift Regularization module to enhance the model's sensitivity toward minority classes. In the context of time-series problems, Wang et al.⁵⁷ incorporated a Recursive Dense Decoder (RDS-Decoder), which emulates the behavior of a Recurrent Neural Network (RNN) within the decoder. This facilitates the dense reuse of feature information between decoder feature maps at the same scale, capturing temporal dependencies between feature maps. In the process of leveraging spatial and multimodal feature fusion, Yuan et al.⁴⁵ designed dual encoder branches, each processing a different modality of images (positron emission tomography (PET) or computed tomography (CT)). Each encoder branch consists of multiple convolutional layers to extract features at various scales for each modality. A fusion learning component generates spatial fusion maps from the features obtained by the two encoder branches, quantifying the contribution of information from each modality. In tackling issues of uncertainty, the methods of Huang et al.,⁵¹ Diao et al.,⁵² and Huang et al.⁵³ showcase the integration strategy and functionality of the Dempster–Shafer (DS) theory within the model framework.

Solving class imbalance issues. In lymphoma research, the pixels occupied by tumors are a minority. This imbalance in foreground-background pixel ratio may bias model predictions toward the majority class (background), leading to insufficient prediction of tumors and thereby reducing segmentation accuracy. Addressing class imbalance is thus another strategy to improve segmentation accuracy. Liu et al.⁴¹ proposed a novel loss function called Class Balanced Dice Loss (CBDL), which tackles class imbalance by considering the effective number of samples. This method protects loss and gradient computations from being dominated by simple majority negative samples. During training, CBDL is integrated into the backbone as the loss function, calculating the CBDL between network outputs and ground truth labels to measure segmentation accuracy. Additionally, they employed a Class Balanced Coefficient Module (CBCM) to compute balance coefficients for each class based on the effective number of samples. Wang et al.⁶⁴ introduced the Prior-Shift Regularization (PSR) module to address this issue. The PSR module comprises three main parts: firstly, the Prior-Shift Layer (PSL) performs Online Informative Voxel Mining (OIVM)

based on Expected Prediction Confidence (EPC) for each class, extracting informative voxels from the original training dataset for regularization. Secondly, the Regularizer Generation Layer (RGL) generates a Generalized Informative-Voxel-Aware (GIVA) regularizer based on EPC, computed on the regularization dataset created by PSL for model fine-tuning. Lastly, the Expected Prediction Confidence Updating Layer (EPCUL) utilizes L2 loss to update EPC, ensuring adaptive adjustment during training. These steps enable the model to continually adjust attention to informative voxels, thereby enhancing the model's ability to recognize minority classes and addressing class imbalance issues.

Handling strategies for temporal correlation. In lymphoma segmentation research, the datasets predominantly used are PET/CT scans. PET/CT imaging mode does not provide rapid imaging but records metabolic informations of lesions at different time points, fused with plain CT data. Metabolic informations of lesions can vary over time. Normal and diseased tissues may exhibit similar metabolic activities in PET images at a single time point. By comparing images

from different time points, physiological high metabolism areas (e.g., brain and bladder) can be distinguished from pathological ones. Additionally, biological characteristics of lesions like metabolic rate and heterogeneity may manifest different patterns over time. Incorporating temporal correlations from image sequences better reflects the morphological nature of lymphoma. Wang et al.⁵⁷ introduced the Recurrent Dense Siamese Decoder (RDS-Decoder), simulating recurrent neural network (RNN) behavior within the decoder. This allows dense reuse of feature information among decoder feature maps of the same scale, capturing temporal dependencies between feature maps. Wang modified the backbone network to a multibranch GCU-HGCU encoder based on Unet. GCU, a modified Conv-LSTM unit with ReLU activation replacing Tanh, generates high-resolution cell memory maps accumulating uncompressed information from feedforward GCUs, preserving more image details and contextual information. HGCU combines GCU with hyper-dense connections (HDC), selectively aggregating multimodal information and enhancing the model's capability to handle diverse data modalities. Each branch processes specific modal inputs, with cell memory from encoding branches utilized not only within but across all branches, achieving efficient multimodal information fusion. The decoder is a Recurrent Dense Siamese Decoder (RDS-Decoder), making the model sensitive to sequential information and utilizing potential time correlations between data. Pang et al.⁷² introduced concepts of metabolic variance features (MVF) and metabolic heterogeneity features (MHF). MVF quantifies metabolic differences between tissues and organs at different time points using PET images, aiding in evaluating metabolic differences between lymphoma and other tissues. MHF describes metabolic distributions in high metabolism areas, subdividing these areas at 80 seconds and calculating MVF across different time points to assess metabolic distribution heterogeneity. Integrating MVF and MHF with traditional lymphoma features forms a feature statistical analysis module, guiding lymphoma segmentation and eliminating false positives and negatives. In clinical practice, PET and CT images are generally acquired simultaneously during a single imaging session, with CT primarily used for attenuation correction and inherently aligned with PET data. While acquiring multiple PET scans in rapid succession immediately after radiotracer injection is uncommon due to clinical and logistical challenges. Temporal correlations in lymphoma segmentation research most often pertain to images obtained at distinct clinical time points, such as baseline and interim PET/CT scans during the course of treatment. For instance, in the study by Tie et al.,⁵⁸ the dual-branch network processes PET/CT images acquired at baseline (PET1) and at an intermediate treatment stage (PET2), leveraging information from the earlier scan to inform segmentation of the later scan. Their model incorporates a longitudinal cross-attention mechanism, where the longitudinal

attention window attention (LAWA) module in the encoder applies windowed multihead cross-attention to dynamically focus on correlated regions between PET1 and PET2, while the longitudinal attention gate attention (LAAG) module in the decoder integrates attention coefficients from both time points to enhance feature fusion across longitudinal data. This approach respects clinical imaging protocols and effectively utilizes temporal information to improve lymphoma segmentation.

Spatial and modal feature fusion strategies. PET/CT itself, aside from its temporal features, represents a spatio and multimodal dataset combining metabolic and density informations, and it features multiple perspectives such as coronal, sagittal, and axial views. Given these rich characteristics, strategies for feature fusion have been developed. By employing multiple classifier heads, including auxiliary and target heads, synergistic operations capture and integrate information across different scales and modalities within the fused images. This diversity contributes to a more comprehensive understanding of the data, thereby enhancing segmentation accuracy. Hu et al.⁴⁶ proposed a strategy based on multiview and 3D fusion. Initially, three 2D networks based on 2D ResU-Net were trained separately to capture lymphoma image information from different directions. Subsequently, a 3D network based on 3D ResU-Net processed volumetric data. The preliminary results from these approaches were fused using a Conv3D fusion strategy, which not only preserved initial segmentation results from each model but also compensated for spatial information loss in single 2D methods by integrating information from different views. Yuan et al.⁴⁵ designed two encoder branches, each handling a modality image (PET or CT). Each encoder branch consisted of multiple convolutional layers to extract features at different scales for each modality. A hybrid learning component generated spatial fusion maps from features obtained from both encoder branches, quantifying the contribution of each modality's information. Through 3D convolution operations, the hybrid learning unit integrated features from different modalities, producing new fused feature maps. These fused maps were subsequently concatenated with PET- and CT-specific feature maps to obtain final fused feature representations at different scales for lymphoma prediction. Diao et al.⁴⁷ proposed a method based on spatial compression and multimodal feature fusion attention network. Initially, 3D PET/CT volume data underwent coronal plane compression, generating m 2D images. Using these 2D images, a 3D coronal spatial compression attention volume was obtained, leveraging whole-body anatomical information to generate spatial attention maps guiding precise segmentation. In PFAS-Net, input comprised 2D axial PET/CT slices, guided by coronal spatial compression attention maps obtained during the CSAE-Net phase for segmentation. Additionally, a Multi-Modal Fusion Attention (MFA)

module fused PET metabolic information and CT anatomical information across three-dimensions in the network layer. Through attention mechanisms, learned fusion weights weighted the feature maps of different modalities and positions according to learned weights. In the decoding stage's skip connection, the MFA module fused PET features extracted during the encoder phase with CT image features while using spatial compression attention maps obtained during the CSAE-Net phase to guide accurate segmentation.

Wang et al.⁶⁶ introduced a Structured Collaborative Learning (SCL) approach to enhance segmentation accuracy. It employed two main components: the Context-Aware Structured Classifier Population Generation (CA-SCPG) module and the Knowledge-Aware Structured Classifier Population Supervision (KA-SCPS) module. CA-SCPG used a lightweight High-Level Context-Aware Dense Connection (HLCA-DC) mechanism to directly enhance feature propagation in target classifier paths. HLCA-DC strengthened feature propagation by densely reusing previous classifier outputs' prediction maps as inputs for subsequent classifier paths. The KA-SCPS module introduced a novel Knowledge-Aware Dice Loss (KA-DL) to supervise auxiliary classifiers. KA-DL broadcasted knowledge learned by the target classifier to other classifier heads, coordinating the learning process of the classifier group. KA-DL defined positive samples as voxels predicted positively by both the target classifier head and the current classifier head, ensuring optimization of auxiliary classifier heads could explicitly lead to improvements in the target classifier's segmentation performance.

Solving the uncertainty problem. The uncertainty problem in image segmentation encompasses various factors that can affect the accuracy and reliability of segmentation results. In lymphoma segmentation studies, these factors primarily include variability in imaging quality, blurred boundaries between lymphoma and normal tissues, challenges in multi-modal data fusion, subjective nature of annotations, and uncertainties in model predictions. To address these challenges, researchers have introduced the Dempster–Shafer (DS) theory, also known as evidence theory or belief function theory, which is a mathematical framework for handling uncertain information. Proposed by Arthur P. Dempster and Glenn Shafer in the 1970s, DS theory aims to provide a more flexible way than traditional probability theory to represent and combine evidence from different information sources.⁹³ It does not require the additivity condition of probabilities and allows assigning belief to single elements or subsets of the hypothesis space, resembling the evidence gathering process at various levels of abstraction. Moreover, DS theory allows uncertainty to be represented in mass functions and preserves this information during evidence synthesis. The enhancement of segmentation accuracy through DS theory stems from

several aspects: it provides a finer representation of uncertainty, including data uncertainty, model uncertainty, and distribution uncertainty.⁹⁴ This is crucial in medical image segmentation where data often exhibit noise, blur, and incompleteness.⁹⁵ Using Dempster's combination rule, evidence from different modalities (such as PET and CT images) can be effectively fused, addressing information deficiencies or ambiguities that may exist in single-modal data. When conflicting segmentation results arise from different modalities, DS theory offers a mechanism to quantify and resolve these conflicts, rather than relying on simple voting or linear combinations. DS theory also offers a way to quantify the model's confidence in its predictions, thereby improving interpretability and enabling medical professionals and researchers to better understand and trust the model's outputs. Huang et al.,⁵¹ Diao et al.,⁵² and Huang et al.⁵³ have all utilized DS theory to construct evidence fusion layers, using prototypes in feature space to compute belief functions for each voxel, quantifying the uncertainty of whether lymphoma is present or absent at that location. By leveraging the advantages of PET and CT images, these methods are capable of maintaining high resolution while improving recognition and segmentation accuracy of tumor regions.

Reducing the reliance on labels. In lymphoma segmentation, labels refer to pixel-level annotations of lymphoma regions provided by expert clinicians based on imaging data. These annotations are essential for guiding and optimizing image segmentation algorithms to accurately identify and delineate lymphoma tissues. In supervised learning, such labels serve as ground truth, enabling the model to learn distinctions between lymphoma and surrounding tissues. However, due to the clinical complexity of lymphoma diagnosis—which involves over 100 distinct pathological subtypes with varying radiological and metabolic presentations—obtaining accurate and consistent annotations is particularly challenging. High-quality segmentation requires not only careful interpretation of imaging features but also consideration of underlying pathological results, as different lymphoma subtypes may present differently on PET/CT. This heterogeneity underscores the importance of precise lesion segmentation in supporting the diagnostic and monitoring processes, which are inherently complex. In practice, the scarcity of accurately annotated medical imaging data limits the scalability of supervised learning approaches in medical image analysis, especially in lymphoma-related tasks. To address this issue, researchers have increasingly turned to weakly supervised and semisupervised learning methods as promising alternatives that reduce dependence on extensive manual annotations while preserving or even enhancing model performance. Weakly supervised learning refers to scenarios where the supervision signals are incomplete, inexact, or noisy. In the context of lymphoma segmentation, this may involve

using partially labeled images, coarse annotations (e.g. bounding boxes or partial masks), or image-level tags rather than pixel-wise labels. For example, Huang et al.³⁸ proposed a weakly supervised approach where only a portion of the lymphoma volume in each patient was manually annotated, while the rest of the tumor region remained unlabeled. To compensate for the lack of dense annotation, they introduced multiscale feature consistency constraints, enforcing the network to learn consistent representations across predicted tumor regions and the sparse labeled regions. The model also employed cosine similarity to compute feature-level distances between tumor and normal tissue, improving discriminative learning in the absence of dense labels. Their training framework incorporated a combination of supervised loss, deep supervision loss, and a regularization loss—where the latter enforced spatial and semantic consistency between weakly labeled and predicted regions, facilitating robust training from partially annotated data. On the other hand, semisupervised learning leverages both labeled and unlabeled data during training. It assumes that even in the absence of manual labels, meaningful patterns can still be extracted from the data distribution. Yousefirizi et al.³⁷ developed a novel semisupervised segmentation framework tailored for medical imaging by incorporating a loss function derived from the Fuzzy C-Means (FCM) clustering objective. Unlike traditional Dice or cross-entropy loss, which rely on crisp class assignments, FCM captures the inherent fuzziness and uncertainty in medical image classification, making it more suitable for modeling soft boundaries in tumors. Their method enabled the model to use soft cluster memberships of unlabeled voxels in conjunction with labeled samples to better learn decision boundaries, leading to improved generalization. By fusing supervised and unsupervised loss components, the model was able to maintain high segmentation accuracy even when trained on a limited number of annotated cases. In summary, both weakly and semisupervised learning strategies offer valuable solutions to the data annotation bottleneck in lymphoma segmentation. Weak supervision allows models to benefit from coarse or partial annotations by introducing auxiliary constraints and consistency mechanisms. Semisupervised learning, meanwhile, utilizes unlabeled data to refine model representations and decision boundaries through hybrid loss functions and regularization. These approaches reduce reliance on exhaustive manual annotation while enabling the development of accurate, scalable, and clinically meaningful lymphoma segmentation systems.

Expand and enrich the dataset. Characteristics of lymphoma image data and the requirements for training deep learning models exhibit conflicting situations: firstly, medical images face ethical and acquisition challenges, limiting the scale of constructed datasets. Moreover, due to the

difficulty of annotation and the diverse types of lymphoma, datasets often do not cover all types comprehensively, leading to limitations. On the other hand, deep learning models demand large-scale datasets with diverse content to achieve better clinical generalization and robustness. Therefore, expanding and enriching datasets without additional case collection is a research direction. One approach to address this issue is integrating generative adversarial networks (GANs) into the backbone network of models. GANs, introduced by Goodfellow et al.⁹⁶ in 2014, consist of two key components: the generator and the discriminator. These components competitively train during the process, forming an adversarial training mechanism to enhance model performance. The generator aims to produce realistic data (e.g., images, videos, or audio) from random noise input via neural networks. Meanwhile, the discriminator's task is to distinguish between fake data generated by the generator and real data from the dataset, maximizing its ability to correctly classify them. Throughout training, the generator and discriminator compete against each other: the generator attempts to deceive the discriminator, while the discriminator strives not to be deceived. This adversarial training drives the generator to produce higher-quality data. Conte et al.⁵⁴ and Shi et al.⁶⁹ have both employed this approach to enrich limited lymphoma data content and verified the feasibility of generated images.

Downstream DerivativeTasks. In lymphoma segmentation research, analyzing segmentation results for downstream tasks is also a research direction closely related to clinical applications, often involving disease diagnosis, treatment, and prognosis. The metrics formed from such studies are not exclusive to deep learning models; these tasks can also be accomplished using traditional machine learning or statistical methods, providing supervised signals for model training. This training strategy based on clinical practical issues enhances deep learning models' clinical generalization ability. Downstream derivative tasks in these studies include identifying and detecting small lymphoma lesions, calculating TMTV and MTV,⁹⁷ and linking them to disease prognosis. From the perspective of network structure adjustment, there are two main approaches to achieve these goals: multitask learning and cascade learning. Both are common techniques in the fields of deep learning and machine learning, where certain parts of the model can be reused to solve multiple problems or tasks across different stages. Both methods leverage the concept of feature sharing, where features learned for one task can provide useful information for others. By sharing model components or features, both approaches can improve learning efficiency and reduce computational resources. They can be implemented within the framework of deep learning, using neural networks to tackle complex tasks. However, they differ fundamentally, which will be elaborated based on specific research in the following sections.

Multitask learning. Multitask learning (MTL) typically refers to the simultaneous learning of multiple related or unrelated tasks, where these tasks may share underlying feature representations but have separate objective functions optimized jointly to improve overall performance across tasks. In medical image analysis, particularly in lymphoma segmentation, MTL has emerged as a powerful paradigm, allowing a single network to handle tasks such as lesion segmentation and prognosis prediction in a unified framework. A common approach involves using a shared encoder for joint feature extraction, followed by task-specific decoder branches, and training with a composite loss function that balances multiple task-specific objectives. For example, Liu et al.³⁵ proposed a U-Net-based MTL framework where shared features supported both lesion segmentation and the prediction of 2-year event-free survival (2y-EFS), with a weighted sum of segmentation and classification losses facilitating joint optimization. Despite its potential, MTL faces several challenges, including task conflicts and negative transfer due to divergent learning objectives, imbalanced task difficulty and label availability (pixel-level segmentation vs patient-level survival data), and limited interpretability, which can hinder clinical trust. Solutions such as dynamic loss weighting, curriculum learning, task-specific modules, and integration of explainable AI (artificial intelligence) techniques like attention visualization are being explored to mitigate these issues. Nevertheless, the benefits of MTL are significant: it enhances feature representation through shared learning, improves data efficiency when annotated data is limited, and aligns with real-world clinical workflows by enabling end-to-end models capable of both anatomical analysis and prognostic insight. Future directions include incorporating uncertainty-aware loss functions, integrating multi-modal data such as clinical or genomic information, and developing continual learning mechanisms to adapt to evolving clinical knowledge. Overall, MTL offers a promising strategy for advancing lymphoma segmentation and related tasks, though it requires careful design and validation to ensure robustness, fairness, and clinical usability.

Cascade learning. Cascade learning focuses on sequentially addressing a series of tasks where the output of each task serves as the input for the next task, forming a processing pipeline.⁹⁸ Networks in cascade learning consist of multiple independent models, each dedicated to a specific task, connected sequentially. Each model may be trained independently, and the training process can be conducted in stages where the output of each stage is used as the input for the next stage. This approach is also a primary strategy for achieving downstream derivative tasks in lymphoma segmentation research, including tasks such as TMTV and MTV computation, where the segmentation results of the previous part are fed into the next part of the network.

Jemaa et al.³⁰ cascaded Unet and Vnet to sequentially achieve whole-body-region-organ-lesion segmentation and TMTV computation goals. Yousefirizi et al.³⁴ in their study cascaded two 3D U-Nets using a soft voting module, achieving the above tasks under a semisupervised setting. In summary, multitask learning emphasizes solving multiple tasks simultaneously within a single model, whereas cascade learning emphasizes sequentially solving a series of tasks, where each task may be handled by different models or model components. Both approaches are effective methods for enhancing learning efficiency and performance but are applicable in different scenarios and for different requirements.

Model performance and included metrics

In the task of lymphoma segmentation based on deep learning, evaluation metrics are crucial for quantifying segmentation performance. They reflect the algorithm's efficacy and the accuracy of segmentation results. The following section will summarize the evaluation metrics applied in current research and compile them, along with experimental results, in Table 2.

Dice similarity coefficient. DSC is a statistical tool used to evaluate the similarity or overlap between two samples, particularly in medical image analysis for assessing the accuracy of image segmentation. The DSC value ranges between 0 and 1, where a value closer to 1 indicates a higher overlap between the two samples and thus a more accurate segmentation result.

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (1)$$

Here, A represents the ground truth, B the algorithm's predicted segmentation region, $|A \cap B|$ denotes the intersection of A and B , which is the size of the overlapping area between the two regions. $|A|$ and $|B|$ represent the total area or volume of A and B , respectively.

Jaccard Index. The Jaccard Index is a measure of similarity between two sets. In fields such as medical image segmentation, the Jaccard coefficient is commonly used as an evaluation metric to quantify the overlap between segmentation results and ground truth annotations. The expression is given by:

$$\text{Jaccard} = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

Here, A represents the set of ground truth annotations, and B the set of predicted segmentation results by the algorithm. $|A \cap B|$ denotes the size of the intersection of sets A and B , which is the number of elements common to both sets. $|A \cup B|$ denotes the size of the union of sets A and B ,

which is the total number of elements in both sets combined. The Jaccard coefficient ranges between 0 and 1, where a value closer to 1 indicates a higher degree of overlap between the sets, reflecting better consistency between the segmentation result and the ground truth annotation. The Jaccard coefficient provides a simple and intuitive method to assess the performance of segmentation algorithms, particularly suitable for binary image segmentation tasks where sets can be simplified to sets of pixels.

Accuracy. ACC is a commonly used evaluation metric in classification tasks, used to measure the model's correctness in classifying all samples. It is defined as the proportion of correctly classified samples to the total number of samples. The formula for accuracy is given by:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (3)$$

Here, TP (True Positives) represents the number of samples that are correctly identified as positive samples. TN (True Negatives) denotes the number of samples that are correctly identified as negative samples. FP (False Positives) indicates the number of negative samples incorrectly classified as positive, and FN (False Negatives) indicates the number of positive samples incorrectly classified as negative. Accuracy values range from 0 to 1, with higher values typically indicating better classification performance. However, in the context of medical image segmentation—particularly in modalities such as PET, CT, and MRI where there is often a severe class imbalance between background and lesion voxels—accuracy may not serve as a reliable performance metric. Even models with suboptimal segmentation performance can yield deceptively high accuracy scores due to the predominance of background voxels. While accuracy is sometimes reported, it provides limited insight in such scenarios. Therefore, overlap-based metrics like the Dice coefficient and Jaccard index, as well as distance-based metrics such as the Hausdorff distance, are more informative and widely accepted for evaluating segmentation quality, especially in the presence of class imbalance.

Sensitivity. Sensitivity, also known as True Positive Rate (TPR) or Recall, is a crucial metric for evaluating the performance of classification models, particularly in fields like medical image segmentation and disease diagnosis. It measures the model's ability to correctly identify positive samples (e.g. vascular regions). The formula for sensitivity is defined as:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

Sensitivity values range between 0 and 1, where higher values indicate a stronger ability of the model to recognize positive samples. In medical imaging, higher sensitivity

implies that the model can effectively detect relevant features or regions of interest, which is critical for accurate segmentation and disease diagnosis.

Specificity. Specificity, also known as True Negative Rate (TNR), is a critical metric for evaluating the performance of classification models, particularly in fields such as medical image segmentation and disease diagnosis. It measures the model's ability to correctly identify negative samples (such as nonlesion areas). The expression is given by:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (5)$$

Specificity values range from 0 to 1, where a higher value indicates stronger capability of the model to recognize negative samples. For instance, in medical image segmentation, a high specificity means that nearly all true nonlesion areas are correctly identified, thereby reducing the likelihood of misdiagnosis.

Precision. Precision is a crucial evaluation metric in classification tasks, used to measure the proportion of correctly predicted positive samples among all samples predicted as positive by the model. A high precision indicates that a higher proportion of positively predicted samples are indeed true positives, implying fewer false positives (FP). The formula for precision is given by:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

Precision values range from 0 to 1, where a higher value signifies greater accuracy of positive predictions by the model. In practical applications like medical image segmentation, precision helps us understand the model's accuracy in predicting regions of interest, thereby evaluating its performance.

Mean squared error. Mean squared error (MSE), an index measuring the discrepancy between predicted values and actual values, represents the average of the squared prediction errors. The formula for calculating MSE is as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

Herein, y_i denotes the true value, $(\hat{y})_i$ signifies the predicted value, and n indicates the number of samples. The lower the value of MSE, the higher the predictive accuracy of the model.

Mean absolute deviation. Mean absolute deviation (MAD) is a statistical measure that quantifies the amount of variation or dispersion from the mean in a set of values. It represents the average distance of each data point from the mean. MAD is calculated by taking the absolute value of the

difference between each data point and the mean, summing these absolute values, and then dividing by the number of data points. The formula for MAD is:

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^n |x_i - \mu| \quad (8)$$

where x_i represents each individual data point; μ the mean of the data; and n the total number of data points. MAD is particularly useful in statistical analysis because it is robust to outliers and provides a measure of dispersion that is in the same units as the original data. It is often used in the context of medical imaging, where it can help assess the variability in image intensity values or the accuracy of image segmentation algorithms. In the context of lymphoma segmentation, MAD can be used to evaluate how closely the segmented tumor boundaries align with the ground truth, with a lower MAD indicating a better segmentation result.

Hausdorff distance. Hausdorff distance (HD) is used to measure the maximum distance between two sets, and here it is employed to assess the maximum inconsistency between segmentation boundaries and true boundaries.⁹⁹ The formula for HD is:

$$\text{HD}(A, B) = \max \left(\max_{x \in A} d(x, B), \max_{y \in B} d(y, A) \right) \quad (9)$$

where A and B are two sets of points; $d(x, B)$ the distance from point x to the nearest point in set B . $d(y, A)$ is the distance from point y to the nearest point in set A . HD quantifies the maximum deviation between the segmentation boundary and the true boundary, aiding in the evaluation of segmentation algorithm accuracy, particularly in boundary localization.

Area under the curve. Area under the curve (AUC) is a widely used metric to evaluate the performance of binary classification models, particularly in medical image segmentation and diagnostic tasks. It represents the area under the receiver operating characteristic (ROC) curve, which plots the TPR (or sensitivity) against the FPR at various threshold settings. AUC quantifies the model's ability to distinguish between positive and negative classes, where a value of 1.0 indicates perfect discrimination and 0.5 corresponds to random guessing. The ROC curve is defined by the following:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (10)$$

The AUC is then computed as the integral of the ROC curve:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR}) \quad (11)$$

In practice, AUC is often approximated numerically using the trapezoidal rule based on discrete threshold points.

Discussion

Clinical prospects of deep learning strategies for lymphoma segmentation

Research in medical image processing derives its materials from clinical scenarios, driven by specific clinical problems. Throughout this article, most studies focus on evaluation metrics related to natural image segmentation, while some researchers translate clinical issues into technical problems for computer-based research (Figure 6). In this section, we will discuss these two aspects: the application of lymphoma segmentation research based on deep learning strategies in clinical settings, and its potential translational prospects.

Diagnosis and staging. The preliminary evaluation of lymphoma through imaging, particularly PET/CT, plays a pivotal role not only in diagnosis but also in disease staging, which is central to treatment planning. Contemporary research on lymphoma segmentation has decomposed this task into multiple subproblems, such as enhancing segmentation accuracy and refining automated decision-support pipelines. As discussed in previous sections, numerous efforts have been made to improve network architectures to address these objectives. Recent studies have explored challenges including temporal and spatial correlations in imaging data, feature fusion strategies, class imbalance, and uncertainty estimation. While these theoretical advancements hold promise for improving lymphoma-related imaging analysis, large-scale clinical validation remains limited. This limitation is primarily due to the difficulty of acquiring comprehensive lymphoma datasets and the presence of over one hundred histological subtypes,² each exhibiting significant heterogeneity in radiological and metabolic features. Moreover, current public and private datasets fail to represent the full spectrum of lymphoma subtypes, thereby constraining the clinical translation of automated segmentation methods. Accurate imaging-based staging—particularly using PET/CT—is critical for guiding treatment strategies, such as determining therapy intensity or eligibility for clinical trials, highlighting the necessity for robust, generalizable segmentation models.

Disease monitoring and prediction. Recent research has increasingly integrated lymphoma segmentation into multitask and cascade learning frameworks to support downstream tasks such as disease monitoring and prognostication. In multi-task learning settings, segmentation is jointly optimized with related objectives—such as classification, out-come prediction, or survival estimation—

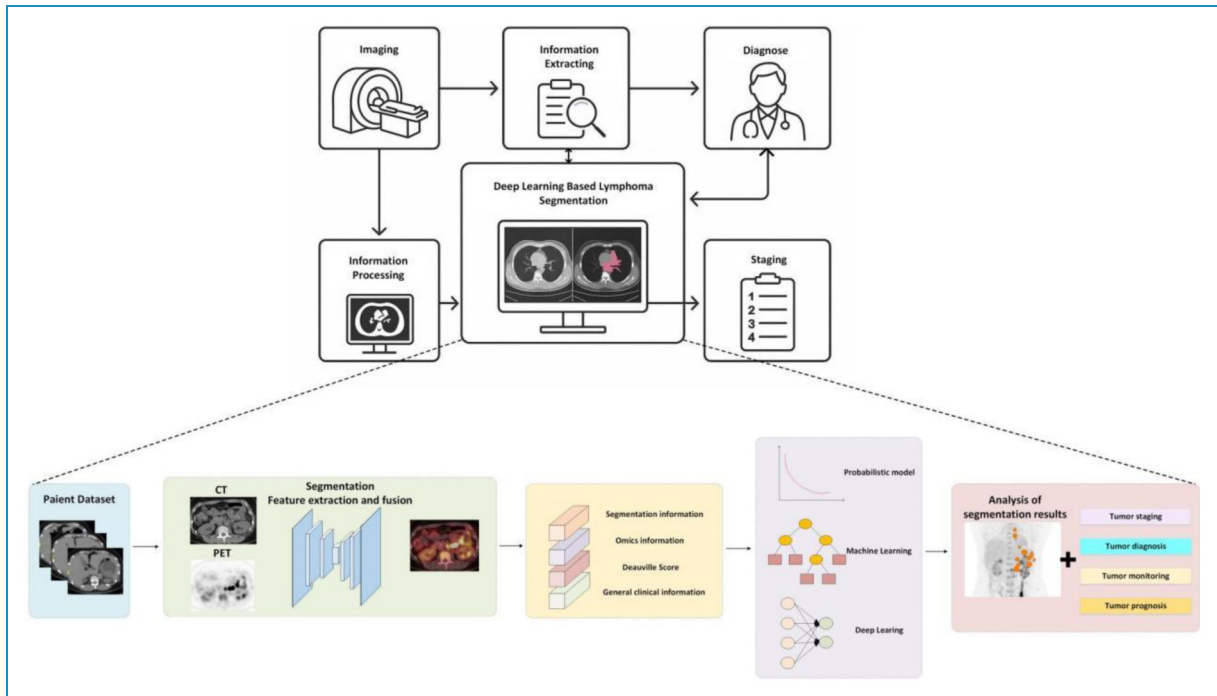


Figure 6. Workflow of a clinically application-oriented integrated lymphoma segmentation method.

allowing the model to benefit from shared representations and improve overall performance. Cascade learning, on the other hand, typically involves using the output of the segmentation network—such as lesion boundaries or volumetric measurements—as input to subsequent modules, including radiomic feature extraction or risk stratification algorithms. A key quantitative parameter in this context is TMTV, derived from PET/CT imaging. TMTV quantifies tumor metabolic activity based on FDG uptake,⁹⁷ and its accurate computation relies on precise segmentation of lymphoma lesions.⁷³ ROC curve analysis can be used to determine optimal TMTV thresholds for patient risk stratification. Baseline PET/CT scans acquired before therapy are used to assess TMTV and predict prognosis, while interim scans help monitor treatment response and guide therapeutic adjustments. TMTV also serves as an independent prognostic factor in multivariate Cox regression models. Additional imaging-derived indicators—such as tumor dissemination patterns,¹⁰⁰ metabolic heterogeneity,¹⁰¹ and total lesion glycolysis¹⁰²—further contribute to patient stratification. The development of segmentation-guided, multi-indicator, and multitask models presents a promising direction for advancing personalized lymphoma management.

Limitations of technology

Despite the promising translational potential of deep learning-based lymphoma segmentation, several technical

challenges must be addressed to enable reliable and widespread clinical deployment. A primary concern is the limited generalizability of deep learning models across different clinical sites. Models trained on data from a single institution or scanner often underperform on external datasets due to domain shifts arising from variations in imaging protocols, scanner manufacturers, and reconstruction parameters. To overcome this, recent research has explored strategies such as domain adaptation,¹⁰³ multicenter collaborative training, and federated learning¹⁰⁴ to enhance cross-site robustness. Another significant challenge is the severe class imbalance inherent in lymphoma segmentation, where lesions often occupy only a small proportion of the total imaging volume. This imbalance can bias models toward the background class, reducing their sensitivity to lesion regions. To mitigate this, approaches such as specialized loss functions¹⁰⁵ and lesion-aware sampling strategies have been proposed and are actively being refined. Furthermore, the high computational demands of deep learning models—particularly for processing high-resolution 3D volumetric PET/CT data—limit their practical deployment in real-time and resource-constrained clinical environments. Although methods such as model compression, quantization, and network pruning offer potential solutions, achieving efficiency without compromising accuracy remains a key bottleneck. Interpretability also poses a major barrier to clinical adoption. Most deep learning models operate as black boxes, providing little insight into their decision-making processes. Since

clinicians require transparent, explainable outputs to validate and trust automated suggestions, recent efforts have focused on integrating attention mechanisms, saliency maps, and feature attribution methods. However, the clinical relevance and robustness of these interpretability techniques remain under investigation. Finally, reproducibility is a longstanding issue in medical image analysis. Variations in preprocessing steps, model architectures, training settings, and evaluation metrics often make it difficult to reproduce and fairly compare results. The adoption of standardized frameworks—such as nnU-Net—and public bench-marks and challenges are crucial for advancing reproducible and comparable research in this field. Together, these limitations highlight the need for continued technical innovation and standardization to enhance the robustness, fairness, and clinical utility of deep learning approaches for lymphoma segmentation.

Conclusion

In summary, deep learning strategies are advancing the progress of lymphoma segmentation research. The clinical motivation underlying this research provides new directions and perspectives for lymphoma segmentation studies. Future research in lymphoma segmentation will increasingly align with medical realities, aiming for the transformation of technology into practical applications.

Contributorship

WL was involved in conceptualization, methodology, investigation, and writing—original draft; FY in writing—review and editing and formal analysis; PT in writing—review and editing, methodology, and investigation; TZ in writing—review and editing and conceptualization; and WS in writing—review and editing, supervision, and project administration.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Jilin Province Health Talent Special Fund (grant no. JLSWSRCZX2023-14) and the Natural Science Foundation of Jilin Province (grant no. YDZJ202301ZYTS065).

Guarantor


Prof. Weizhang Shen takes responsibility for the integrity of the work as a whole, from inception to published article.


Informed consent


Not applicable.

ORCID iDs

Wanru Liang  <https://orcid.org/0009-0006-9710-3250>

Feiyang Yang  <https://orcid.org/0009-0003-2410-6651>

Tianyang Zhang  <https://orcid.org/0009-0002-8928-5921>

Weizhang Shen  <https://orcid.org/0000-0003-0786-0890>

Peer review

This manuscript was externally peer reviewed. All peer review was conducted independently and anonymously.

References

1. Connors JM, Cozen W, Steidl C, et al. Hodgkin lymphoma. *Nat Rev Dis Primers* 2020; 6: 61.
2. Lu P. Staging and classification of lymphoma. *Semin Nucl Med* 2005; 35: 160–164. Elsevier.
3. Lewis RB, Mehrotra AK, Rodríguez P, et al. From the radiologic pathology archives: gastrointestinal lymphoma: radiologic and pathologic findings. *Radiographics* 2014; 34: 1934–1953.
4. Ramesh KKD, Kiran Kumar G, Swapna K, et al. A review of medical image segmentation algorithms. *EAI Endorsed Transact Pervasive Health Technol* 2021; 7: e6–e6.
5. Boellaard R, Buvat I, Nioche C, et al. International benchmark for total metabolic tumor volume measurement in baseline 18f-FDG PET/CT of lymphoma patients: a milestone toward clinical implementation. *J Nucl Med* 2024; 65: 1343–1348.
6. Intlekofer AM and Younes A. Precision therapy for lymphoma—current state and future directions. *Nat Rev Clin Oncol* 2014; 11: 585–596.
7. Veziroglu EM, Farhadi F, Hasani N, et al. Role of artificial intelligence in pet/ct imaging for management of lymphoma. *Semin Nucl Med* 2023; 53: 426–448. Elsevier.
8. Shrivastava N and Bharti J. Automatic seeded region growing image segmentation for medical image segmentation: a brief review. *Int J Image Graph* 2020; 20: 2050018.
9. Sharma P and Suji J. A review on image segmentation with its clustering techniques. *Int J Signal Proc Image Proc Pattern Recognit* 2016; 9: 209–218.
10. Seo H, Khuzani MB, Vasudevan V, et al. Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications. *Med Phys* 2020; 47: e148–e167.
11. Yuan J, Zhang Y and Wang X. Application of machine learning in the management of lymphoma: current practice and future prospects. *Digital Health* 2024; 10: 20552076241247963.
12. Ghaffari M, Sowmya A and Oliver R. Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the brats 2012–2018 challenges. *IEEE Rev Biomed Eng* 2019; 13: 156–168.
13. Oreiller V, Andrearczyk V, Jreige M, et al. Head and neck tumor segmentation in PET/CT: the Hecktor challenge. *Med Image Anal* 2022; 77: 102336.

14. Gatidis S, Früh M, Fabritius M, et al. The AutoPET challenge: towards fully automated lesion segmentation in oncologic PET/CT imaging. 2023.
15. Shen D, Wu G and Suk H-I. Deep learning in medical image analysis. *Annu Rev Biomed Eng* 2017; 19: 221–248.
16. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017; 42: 60–88.
17. Isensee F, Jaeger PF, Kohl SA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 2021; 18: 203–211.
18. Isavand P, Aghamiri SS and Amin R. Applications of multimodal artificial intelligence in non-Hodgkin lymphoma B cells. *Biomedicines* 2024; 12: 2024.
19. LeCun Y, Bengio Y and Hinton G. Deep learning. *Nature* 2015; 521: 436–444.
20. Tian C, Fei L, Zheng W, et al. Deep learning on image denoising: an overview. *Neural Netw* 2020; 131: 251–275.
21. Zhang H, Xu H, Tian X, et al. Image fusion meets deep learning: a survey and perspective. *Inf Fusion* 2021; 76: 323–336.
22. Bouraoui A, Jamoussi S and Hamadou AB. A comprehensive review of deep learning for natural language processing. *Int J Data Mining Model Manag* 2022; 14: 149–182.
23. Lee W, Seong JJ, Ozlu B, et al. Biosignal sensors and deep learning-based speech recognition: a review. *Sensors* 2021; 21: 1399.
24. Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks. *Pattern Recognit* 2018; 77: 354–377.
25. Dhillon A and Verma GK. Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress Artif Intell* 2020; 9: 85–112.
26. Barlow HB. David Hubel and Torsten Wiesel: their contributions towards understanding the primary visual cortex. *Trends Neurosci* 1982; 5: 145–152.
27. Suzuki K. Overview of deep learning in medical imaging. *Radiol Phys Technol* 2017; 10: 257–273.
28. LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proc IEEE* 1998; 86: 2278–2324.
29. Alom MZ, Taha TM, Yakopcic C, et al. The history began from Alexnet: a comprehensive survey on deep learning approaches. *arXiv preprint arXiv:1803.01164*, 2018.
30. Jemaa S, Fredrickson J, Carano RA, et al. Tumor segmentation and feature extraction from whole-body FDG-PET/CT using cascaded 2d and 3d convolutional neural networks. *J Digit Imaging* 2020; 33: 888–894.
31. Alderuccio JP, Kuker RA, Lehmkuhl D, et al. Calculation of metabolic tumor volume in diffuse large B-cell lymphoma using a deep learning-aided automated method. *Blood* 2022; 140: 4972–4973.
32. Nam Y, Kim S-Y, Kim K-A, et al. Development and validation of deep learning-based automated detection of cervical lymphadenopathy in patients with lymphoma for treatment response assessment: a bi-institutional feasibility study. *J Imag Inform Med* 2024; 37: 734–743.
33. Jiang C, Chen K, Teng Y, et al. Deep learning-based tumour segmentation and total metabolic tumour volume prediction in the prognosis of diffuse large B-cell lymphoma patients in 3D FDG-PET images. *Eur Radiol* 2022; 32: 4801–4812.
34. Yousefirizi F, Dubljevic N, Ahamed S, et al. Convolutional neural network with a hybrid loss function for fully automated segmentation of lymphoma lesions in FDG PET images. *Medical Imag 2022: Image Proc* 2022; 12032: 214–220. SPIE.
35. Liu P, Zhang M, Gao X, et al. Joint lymphoma lesion segmentation and prognosis prediction from baseline FDG-PET images via multitask convolutional neural networks. *IEEE Access* 2022; 10: 81612–81623.
36. Constantino CS, Leocádio S, Oliveira FPM, et al. Evaluation of semiautomatic and deep learning-based fully automatic segmentation methods on [18f] FDG PET/CT images from patients with lymphoma: influence on tumor characterization. *J Digit Imaging* 2023; 36: 1864–1876.
37. Yousefirizi F, Shiri I, Hyun OJ, et al. Semi-supervised learning towards automated segmentation of pet images with limited annotations: application to lymphoma patients. *Phys Eng Sci Med* 2024; 47: 1–17.
38. Huang Z, Guo Y, Zhang N, et al. Multi-scale feature similarity-based weakly supervised lymphoma segmentation in pet/ct images. *Comput Biol Med* 2022; 151: 106230.
39. El Jurdi R, Nichelli L, Alentorn A, et al. Border irregularity loss for automated segmentation of primary brain lymphomas on post-contrast MRI. *Medical Imag 2024: Image Proc* 2024; 12926: 430–437. SPIE.
40. Luo S, Jiang H and Wang M. C2BA-UNet: a context-coordination multi-atlas boundary-aware UNet-like method for PET/CT images based tumor segmentation. *Comput Med Imaging Graph* 2023; 103: 102159.
41. Liu R, Wang M and Jiang H. CBDL-UNet: UNet with class balanced dice loss for foreground-background class-imbalance lymphoma segmentation in pet images. In: 2022 4th International Conference on Intelligent Medicine and Image Processing, pages 43–48, 2022.
42. Zhai W, Chen J, Min D, et al. Multi-scale/attention convolutional network for diffuse large B lymphoma segmentation in PET image. In: 2022 14th International Conference on Wireless Communications and Signal Processing (WCSP), pages 138–142. IEEE, 2022.
43. Zhu X, Jiang H and Diao Z. CGBO-Net: cruciform structure guided and boundary-optimized lymphoma segmentation network. *Comput Biol Med* 2023; 153: 106534.
44. Pennig L, Hoyer UCI, Goertz L, et al. Primary central nervous system lymphoma: clinical evaluation of automated segmentation on multiparametric MRI using deep learning. *J Magn Reson Imaging* 2021; 53: 259–268.
45. Yuan C, Zhang M, Huang X, et al. Diffuse large B-cell lymphoma segmentation in PET-CT images via hybrid learning for feature fusion. *Med Phys* 2021; 48: 3665–3678.
46. Hu H, Shen L, Zhou T, et al. Lymphoma segmentation in pet images based on multi-view and Conv3D fusion

- strategy. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pages 1197–1200. IEEE, 2020.
47. Diao Z, Jiang H and Shi T. A spatial squeeze and multi-modal feature fusion attention network for multiple tumor segmentation from PET–CT volumes. *Eng Appl Artif Intell* 2023; 121: 105955.
 48. Xu G, Cao H, Dong Y, et al. Focal loss function based DeepLabv3+ for pathological lymph node segmentation on PET/CT. In: Proceedings of the 2020 2nd International Conference on Intelligent Medicine and Image Processing, pages 24–28, 2020.
 49. Weisman AJ, Kieler MW, Perlman SB, et al. Convolutional neural networks for automated PET/CT detection of diseased lymph node burden in patients with lymphoma. *Radiol Artif Intell* 2020; 2: 5.
 50. Li H, Jiang H, Li S, et al. DenseX-Net: an end-to-end model for lymphoma segmentation in whole-body PET/CT images. *IEEE Access* 2019; 8: 8004–8018.
 51. Huang L, Denœux T, Tonnelet D, et al. Deep PET/CT fusion with Dempster-Shafer theory for lymphoma segmentation. In: Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12, pages 30–39. Springer, 2021.
 52. Diao Z, Jiang H, Han X-H, et al. EFnet: evidence fusion network for tumor segmentation from PET-CT volumes. *Phys Med Biol* 2021; 66: 205005.
 53. Huang L, Ruan S, Decazes P, et al. Lymphoma segmentation from 3D PET-CT images using a deep evidential network. *Int J Approx Reason* 2022; 149: 39–60.
 54. Conte GM, Weston AD, Vogelsang DC, et al. Generative adversarial networks to synthesize missing t1 and FLAIR MRI sequences for use in a multisequence brain tumor segmentation model. *Radiology* 2021; 299: 313–323.
 55. Thiery O, Rizkallah M, Bailly C, et al. Graph-based multi-modal multi-lesion DLBCL treatment response prediction from PET images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 103–112. Springer, 2023.
 56. Chen S, Li A, Chen J, et al. Hybrid attention fusion segmentation network for diffuse large B-cell lymphoma in PET-CT. In: 2022 14th International Conference on Wireless Communications and Signal Processing (WCSP), pages 72–76. IEEE, 2022.
 57. Wang M, Jiang H, Shi T, et al. HD-RDS-UNet: leveraging spatial-temporal correlation between the decoder feature maps for lymphoma segmentation. *IEEE J Biomed Health Inform* 2021; 26: 1116–1127.
 58. Tie X, Shin M, Lee C, et al. Automatic quantification of serial PET/CT images for pediatric Hodgkin lymphoma patients using a longitudinally—aware segmentation network. *ArXiv*, 2024.
 59. Wang M and Jiang H. Memory-net: coupling feature maps extraction and hierarchical feature maps reuse for efficient and effective PET/CT multi-modality image-based tumor segmentation. *Knowl Based Syst* 2023; 265: 110399.
 60. Blanc-Durand P, Jégou S, Kanoun S, et al. Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network. *Eur J Nucl Med Mol Imaging* 2021; 48: 1362–1370.
 61. Klimont M, Oronowicz-Jaśkowiak A, Flieger M, et al. Deep learning-based segmentation and volume calculation of pediatric lymphoma on contrast-enhanced computed tomographies. *J Pers Med* 2023; 13: 184.
 62. Lu G, Zhou W, Zhao K, et al. Automatic segmentation model of primary central nervous system lymphoma based on multiple sequences of magnetic resonance images using deep learning. 2023.
 63. Fu G, Nichelli L, Herran D, et al. Comparing foundation models and nnU-Net for segmentation of primary brain lymphoma on clinical routine post-contrast T1-weighted MRI. 2024.
 64. Wang M, Jiang H, Shi T, et al. PSR-Nets: deep neural networks with prior shift regularization for PET/CT based automatic, accurate, and calibrated whole-body lymphoma segmentation. *Comput Biol Med* 2022; 151: 106215.
 65. Ahamed S, Dubljevic N, Bloise I, et al. A cascaded deep network for automated tumor detection and segmentation in clinical pet imaging of diffuse large B-cell lymphoma. *Medical Imag 2022: Image Proc* 2022; 12032: 934–941. SPIE.
 66. Wang M, Jiang H, Shi T, et al. SCL-Net: structured collaborative learning for PET/CT based tumor segmentation. *IEEE J Biomed Health Inform* 2022; 27: 1048–1059.
 67. Yousefirizi F, Klyuzhin IS, Hyun OJ, et al. TMTV-Net: fully automated total metabolic tumor volume segmentation in lymphoma PET/CT images—a multi-center generalizability analysis. *Eur J Nucl Med Mol Imaging* 2024; 51: 1–18.
 68. Huang H, Qiu L, Yang S, et al. Vision transformer-based multimodal feature fusion network for lymphoma segmentation on PET/CT images. *arXiv preprint arXiv: 2402.02349*, 2024.
 69. Shi T, Jiang H, Wang M, et al. Metabolic anomaly appearance aware U-Net for automatic lymphoma segmentation in whole-body PET/CT scans. *IEEE J Biomed Health Inform* 2023; 27: 2465–2476.
 70. Zhou Z, Jain P, Lu Y, et al. Computer-aided detection of mantle cell lymphoma on 18F-FDG PET/CT using a deep learning convolutional neural network. *Am J Nucl Med Mol Imaging* 2021; 11: 260.
 71. Kuker RA, Lehmkuhl D, Kwon D, et al. A deep learning-aided automated method for calculating metabolic tumor volume in diffuse large B-cell lymphoma. *Cancers (Basel)* 2022; 14: 5221.
 72. Pang W, Li S, Jiang H, et al. MTR-PET: multi-temporal resolution PET images for lymphoma segmentation. *Biomed Signal Process Control* 2024; 87: 105529.
 73. Revailier W, Cottureau AS, Rossi C, et al. Deep learning approach to automatize TMTV calculations regardless of

- segmentation methodology for major FDG-avid lymphomas. *Diagnostics* 2022; 12: 417.
74. Zhao C, Song J, Yuan Y, et al. An improved attention module based on nnU-Net for segmenting primary central nervous system lymphoma (PCNSL) in MRI images 1. *J Xray Sci Technol* 2024; 32: 1–17.
 75. Liang W, Tadesse GA, Ho D, et al. Advances, challenges and opportunities in creating data for trustworthy AI. *Nature Machine Intell* 2022; 4: 669–677.
 76. Galbusera F and Cina A. Image annotation and curation in radiology: an overview for machine learning practitioners. *Eur Radiol Exp* 2024; 8: 11.
 77. Ying X. An overview of overfitting and its solutions. *J Phys: Conf Ser* 2019; 1168: 022022. IOP Publishing.
 78. Cheson BD. PET/CT in lymphoma: current overview and future directions. *Semin Nucl Med* 2018; 48: 76–81. Elsevier.
 79. Wayne Slone H, Blake JJ, Shah R, et al. CT and MRI findings of intracranial lymphoma. *Am J Roentgenol* 2005; 184: 1679–1685.
 80. Fishman EK, Kuhlman J and Jones RJ. CT of lymphoma: spectrum of disease. *Radiographics* 1991; 11: 647–669.
 81. Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging* 2014; 34: 1993–2024.
 82. Elharrouss O, Akbari Y, Almaadeed N, et al. Backbones-review: feature extraction networks for deep learning and deep reinforcement learning approaches. *arXiv preprint arXiv:2206.08016*, 2022.
 83. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.
 84. Koonce B and Koonce B. Resnet 50. Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization 2021; 1: 63–72.
 85. Ghosal P, Nandanwar L, Kanchan S, et al. Brain tumor classification using ResNet-101 based squeeze and excitation deep neural network. In: 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), pages 1–6. IEEE, 2019.
 86. Siddique N, Paheding S, Elkin CP, et al. U-Net and its variants for medical image segmentation: a review of theory and applications. *IEEE Access* 2021; 9: 82031–82057.
 87. Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention– MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II 19, pages 424–432. Springer, 2016.
 88. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4700–4708, 2017.
 89. Milletari F, Navab N and Ahmadi S-A. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pages 565–571. IEEE, 2016.
 90. Kamnitsas K, Ferrante E, Parisot S, et al. DeepMedic for brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Second International Workshop, BrainLes 2016, with the Challenges on BRATS, ISLES and mTOP 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 2, pages 138–149. Springer, 2016.
 91. Chen LC, Zhu Y, Papandreou G, et al. DeepLab: Deep labelling for semantic image segmentation, 2018.
 92. Isensee F, Petersen J, Klein A, et al. nnU-Net: self-adapting framework for U-Net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.
 93. Shafer G. Dempster-Shafer theory. *Encyclopedia of Artificial Intelligence* 1992; 1: 330–331.
 94. Rombaut M and Zhu YM. Study of Dempster-Shafer theory for image segmentation applications. *Image Vis Comput* 2002; 20: 15–23.
 95. Fidon L, Aertsen M, Kofler F, et al. A Dempster-Shafer approach to trustworthy AI with application to fetal brain MRI segmentation. *IEEE Trans Pattern Anal Mach Intell* 2024; 46: 3784–3795.
 96. Creswell A, White T, Dumoulin V, et al. Generative adversarial networks: an overview. *IEEE Signal Process Mag* 2018; 35: 53–65.
 97. Feres CCP, Nunes RF, Teixeira LLC, et al. Baseline total metabolic tumor volume (TMTV) application in Hodgkin lymphoma: a review article. *Clin Transl Imaging* 2022; 10: 273–284.
 98. Marquez ES, Hare JS and Niranjana M. Deep cascade learning. *IEEE Trans Neural Networks Learning Syst* 2018; 29: 5475–5485.
 99. Huttenlocher DP, Klanderman GA and Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell* 1993; 15: 850–863.
 100. Pals ST, de Gorter DJ and Spaargaren M. Lymphoma dissemination: the other face of lymphocyte homing. *Blood, J Am Soc Hematol* 2007; 110: 3102–3111.
 101. Schürch CM, Federmann B, Quintanilla-Martinez L, et al. Tumor heterogeneity in lymphomas: a different breed. *Pathobiology* 2018; 85: 130–145.
 102. Guo B, Tan X, Ke Q, et al. Prognostic value of baseline metabolic tumor volume and total lesion glycolysis in patients with lymphoma: a meta-analysis. *PLoS ONE* 2019; 14: e0210224.
 103. Singhal P, Walambe R, Ramanna S, et al. Domain adaptation: challenges, methods, datasets, and applications. *IEEE Access* 2023; 11: 6973–7020.
 104. Wen J, Zhang Z, Lan Y, et al. A survey on federated learning: challenges and applications. *Int J Machine Learning Cybernetics* 2023; 14: 513–535.
 105. Li X, Lv C, Wang W, et al. Generalized focal loss: towards efficient representation learning for dense object detection. *IEEE Trans Pattern Anal Mach Intell* 2022; 45: 3139–3153.