3D-based Convolutional Neural Networks for Medical Image Segmentation: A Review

Review Paper

Siti Raihanah Abdani
College of Computing, Informatics, and Mathematics,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia
2022236678@student.uitm.edu.my

Nursuriati Jamil
College of Computing, Informatics, and Mathematics,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia
lizajamil@computer.org

Syed Mohd Zahid Syed Zainal Ariffin*
College of Computing, Informatics, and Mathematics,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia
zahidzainal@uitm.edu.my

Shafaf Ibrahim
College of Computing, Informatics, and Mathematics,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia
shafaf2429@uitm.edu.my

*Corresponding author

Abstract *– Medical image segmentation is essential for disease screening and diagnosis, particularly through techniques like anatomical and lesion segmentation that can be used to isolate critical regions of interest. However, manual segmentation is labor-intensive, costly, and susceptible to subjective bias, underscoring the need for automation. Deep learning, particularly convolutional neural networks (CNNs), has significantly advanced segmentation accuracy and efficiency. With the introduction of 3D imaging, research has evolved from 2D CNNs to 3D CNNs, which leverage inter-slice information to improve segmentation precision. This paper aims to provide a literature review of studies published between 2018 and 2024 on platforms such as Google Scholar and ScienceDirect, where the identified relevant research are "3D segmentation" and "3D medical imaging". This study outlines the key stages of 3D CNN segmentation that include preprocessing, region-of-interest extraction, and post-processing. Furthermore, this study emphasizes the application of 3D CNN architectures to complex lung imaging scenarios, such as lung cancer and COVID-19. Although 3D CNNs outperform 2D CNNs in preserving spatial continuity across slices, they present notable limitations. Key challenges include heavy computational and high memory demands, as well as a dependency on large annotated datasets, which are often scarce in medical imaging. Additionally, effective multiscale feature learning remains a challenging issue, with current architectures struggling to generalize the features of interest across several usage variations. To further improve the segmentation performance, future research should prioritize developing adaptive algorithms and fostering interdisciplinary collaboration between computer scientists and medical professionals to design efficient and scalable models, designed specifically for clinical applications. This future research direction will enhance diagnostic accuracy and segmentation quality in 3D medical imaging.*

Keywords: *Medical Imaging, Semantic Segmentation, Artificial Intelligence, Deep Learning, Diagnosis Tools*

Received: November 18, 2024; Received in revised form: December 17, 2024; Accepted: January 6, 2025

1. INTRODUCTION

Image segmentation, particularly in medical applications, is essential for accurately distinguishing and isolating regions of interest within medical images, thereby aiding in diagnosis and treatment planning. Although manual segmentation is often more precise, it is time-intensive and susceptible to subjectivity issues, prompting the need for automated approaches. A commonly used conventional technique is thresholding, where lung regions are identified as the largest blob, as described by Manikandan [1]. However, this method lacks robustness when applied to lung disease cases, such as those affected by pneumonia or COVID-19, due to substantial variability in lung image data. To address this limitation, adaptive techniques like watershed segmentation have been explored, as demonstrated by Navya and Pradeep [2]. Nonetheless, these methods are too dependent on basic assumptions, such as the use of Sobel edge operators, which may not effectively handle intensity variations across different CT scans. Similarly, preprocessing filters such as Wiener, mean, and median filters increase the computational load and do not ensure segmentation robustness [3]. The application of these filters is often followed by morphological masking as the post-processing step, which can inadvertently erode critical lung regions, thereby reducing segmentation accuracy.

Alternative methods, including those that combine super pixels and fuzzy clustering [4], have been proposed to enhance segmentation outcomes. However, super pixel-based clustering can be computationally demanding and frequently fails to achieve good pixel-level accuracy. This pixel grouping into super pixel blobs may obscure minor variations within these regions, potentially leading to the misclassification of lung areas. Such inaccuracies are particularly problematic for disease staging identification, where even slight errors can have substantial diagnostic implications. A review by Ker *et al.* [5] explores the application of machine learning, particularly convolutional neural networks (CNNs), in medical image analysis. It highlights the advantages of machine learning in handling large medical data by analyzing the data's hierarchical relationships without the extensive use of feature engineering.

Deep learning, particularly convolutional neural networks (CNNs), has emerged as a powerful approach for medical image segmentation [6-8] CNNs are capable of automatically learning and extracting features from large datasets, yielding more accurate and reliable segmentation results. The growing availability of large datasets and enhanced computational resources has led to the increasing adoption of deep learning methods in healthcare, where this approach provides robust solutions to the limitations of conventional machine learning techniques. Additionally, the deep learning approach can also be used to facilitate the generation of synthetic datasets through the generative adversarial network (GAN) approach, which helps mitigate privacy concerns that are associated with sensitive medical data [9-12].

Alongside advancements in segmentation techniques, improvements in 3D imaging technologies, such as computed tomography (CT) and magnetic resonance imaging (MRI), have also significantly enhanced medical imaging capabilities. These technologies enable a more detailed and precise assessment of pathologies, particularly small-scale anomalies like cancerous nodules, which may not be discernible in 2D imaging [13, 14]. By providing a comprehensive view of the anatomical structures, 3D imaging reduces the likelihood of mistaking small pathologies for noise and offers a more accurate representation of tissues that might otherwise appear as artifacts in 2D images [15].

Despite the potential of deep learning and advanced imaging technologies, several challenges persist in medical image segmentation. Deep learning models rely on large volumes of annotated data, and the quality of these ground truth data is dependent on the expertise of the annotators, which may lead to inconsistencies. Standardizing annotation practices and improving training for annotators are crucial for enhancing segmentation accuracy. Furthermore, medical segmentation tasks can be divided into anatomical and lesion segmentation. Anatomical segmentation involves delineating organs or structures, which can be complex due to visual similarities between different structures. In contrast, lesion segmentation focuses on identifying abnormal regions, which vary greatly in size, shape, and location across patients, adding to the difficulty of accurate segmentation.

Research on 3D CNNs for medical segmentation has also made substantial progress, with studies exploring both 2D slice-based methods and full 3D volumetric approaches. While 2D methods often overlook crucial inter-slice information, 3D approaches utilize the entire data volume to produce improved segmentation outcomes. Although existing reviews on 3D CNNs [16-18] discuss various facets of 3D segmentation, our work focuses specifically on the application and methodological workflows of 3D CNN segmentation in medical imaging. Additionally, this study examines the use of 3D deep learning methods in lung imaging, where the modification or the improvement of the backbone networks will be categorized and discussed. Further discussions are also added to address the network's limitations and challenges. Finally, future directions in this field are proposed, highlighting areas for continued research and development.

2. METHODOLOGY

This review was conducted by searching the Google Scholar and ScienceDirect databases for peer-reviewed journal articles and conference proceedings. Only English-language articles published between 2018 and 2024 were selected. The search terms are set to "3D segmentation" and "medical image," while exclusion criteria are set to omit books, newspapers, non-peer-reviewed articles, and any study that is not specifically focused on 3D image segmentation within medical applications. Initially, 114 articles were identified, but the selection was refined to focus on papers discussing 3D deep learning algorithms (specifically 3D CNNs) applied to medical image segmentation (see Fig. 2). Fig. 1 summarizes the literature review methodology.

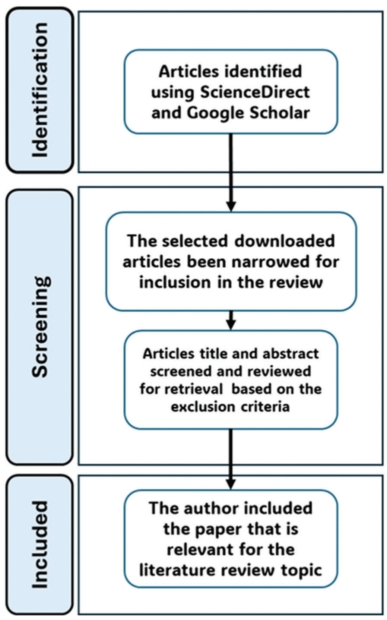

Fig. 1. Literature Review Methodology

3. MEDICAL IMAGE SEGMENTATION

This section begins with a general overview of the importance of medical image segmentation, followed by a discussion comparing the basic 2D CNN and 3D CNN methodologies. Figure 3 outlines the general steps in medical image segmentation using 3D CNNs, which include image pre-processing, region of interest (ROI) identification, 3D CNN segmentation, binary mask generation, and image post-processing. Based on Figure 2, each stage of the research methodology employed in the selected papers is discussed, except for image segmentation, which is later analyzed in greater detail specifically for the lung imaging. This process provides a comprehensive understanding of the overall approach to 3D CNN segmentation in medical imaging, with particular emphasis on lung imaging.

The development of automated segmentation algorithms has been extensively researched in various applications. Recent advancements in the field of medical image processing have led to the emergence of several segmentation models that can be categorized broadly into three classes: 1) conventional image processing-based algorithms, 2) machine learning-based algorithms, and 3) deep learning-based algorithms.

Conventional semantic segmentation algorithms, such as edge-based methods, are commonly employed to identify borders within an image. These methods rely on gradient-based edge detection operators, including Prewitt, Canny, Sobel, Roberts, and Laplacian filters. Despite its limitations, edge-based segmentation can be integrated with more advanced techniques to enhance its performance further. Besides that, local shape analysis has also been applied to segment lung pathologies [19]. In this approach, a set of predefined generic shapes representing local pathologies is compared with the tested input data using a geodesic distance metric. Another method proposed by Cui *et al.* [20] employs a more sophisticated technique involving predefined features through a boundary expansion approach. In this method, an initial seed representing the pathological region is defined, and color information is utilized to expand the region based on a fixed 20% threshold.

For machine learning-based category, it can be further divided into two approaches which are supervised and unsupervised learning. Unsupervised learning, particularly clustering methods, partitions data into distinct groups based on inherent feature similarities. Among these, the K-means algorithm is one of the most widely utilized clustering techniques. In contrast, supervised learning through classification tasks relies on a labeled training dataset, where each data point is associated with a specific target ground truth. One of the most used supervised algorithms is the K-nearest neighbor (K-NN) classifier.

For the third category, the deep learning-based approach mainly leverages Convolutional Neural Network (CNN), which is known for its robust feature extraction capabilities that have demonstrated exceptional performance in tasks such as natural image classification, object detection, and segmentation. As for the segmentation task, Fully Convolutional Neural Network (FCNN) is one of the earliest semantic segmentation models that is particularly well-suited for medical image segmentation tasks. These deep learning-based methods surpass traditional techniques in terms of robustness and accuracy, establishing themselves as the dominant approach in automatic medical image segmentation. Many medical image segmentation tasks have utilized the enhanced versions of the U-Net architecture, which is a symmetric network with skip connections between the encoder and decoder paths [21].

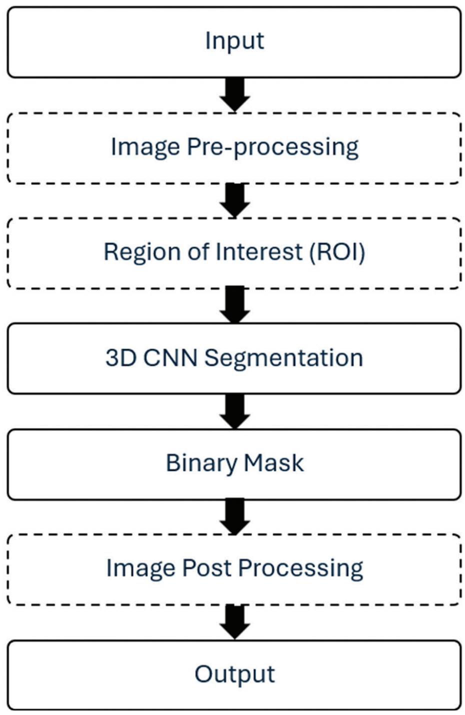

Fig. 2. The general flow of a 3D CNN segmentation. The dotted line boxes are not compulsory steps, while solid line boxes are compulsory steps.

4. 2D VS 3D CNN SEGMENTATION FOR 3D MEDICAL IMAGES

The application of deep learning-based segmentation to 3D medical images can be approached in two distinct ways. The first approach involves directly feeding 3D imaging data into a 3D CNN architecture. The second approach entails slicing the 3D imaging data into a series of 2D slices and inputting these individual 2D slices into a 2D CNN architecture. While considerable research has focused on the second approach due to its lower computational requirements, the 3D CNN approach holds particular advantages for segmentation tasks, especially when dealing with boundaries and edges. This is because 3D CNNs retain more spatial information by maximizing the interslice context, as compared to 2D methods, which may fail to capture important volumetric relationships across slices [22]. In

3D medical imaging, ROI often extends across multiple slices, making the interslice information critical. Additionally, 3D convolutional kernels can process data in all three spatial dimensions, as opposed to 2D convolutional kernels which can only analyze data in two-dimensional format [23].

The 2D U-Net architecture [24] which takes inspiration from the fully convolutional network consists of 23 layers in its symmetric encoder-decoder network. This architecture is typically divided into two segments: the down-sampling path (encoder) and the up-sampling path (decoder). During the down-sampling phase, convolutional and pooling layers are applied to the input image, generating feature maps at varying levels of abstraction. The up-sampling phase gradually restores the size of the feature maps by using deconvolutional layers. To recover the detailed information lost during the down-sampling process, the feature maps are merged with corresponding higher-resolution feature maps from the encoder side. These up-sampling procedures have been implemented in [25] to facilitate the reconstruction of 3D models.

However, since much of medical imaging data is inherently three-dimensional, the application of a 2D U-Net network can lead to the loss of critical spatial information. Moreover, the two-dimensional structure of the network results in the loss of contextual information during the down-sampling process [26] This limitation can reduce the network's sensitivity to fine border details, as they are usually not effectively restored during the up-sampling phase. Consequently, there is a need to employ a three-dimensional network for further optimization that may also contribute to the information loss. Additionally, the input data must undergo slicing, where the 3D data is divided into multiple 2D slices. This process may reduce the network's accuracy, as the correlation between adjacent slices will be lost.

To address these challenges, Çiçek *et al.* [27] proposed the 3D U-Net, which is an extension of the original U-Net architecture by incorporating 3D convolutional and pooling layers in the encoder side and 3D deconvolutional layers in the decoder side. However, the 3D U-Net only utilized three down-sampling layers due to the high computational cost, which limits its ability to extract deep-layer image features. This restriction has resulted in reduced accuracy for certain medical image segmentation tasks. To overcome this challenge, Milletari *et al.* [28] introduced another model, the V-Net, which incorporates residual connections to enable deeper network architectures. Subsequently, over the years, numerous modifications and enhancements have been proposed that significantly improve segmentation accuracy. 3D CNN-based segmentation methods have been successfully applied across a range of medical imaging applications, including head and neck [29], heart [30], lung [31], kidney [32], liver [33], brain [34, 35], and multi-organ segmentation [36], as illustrated in Fig. 3.

The main contribution of this review paper is the discussion of several categories of these model modifications, highlighting their contributions to performance improvements. In subsequent sections, we discuss the limitations of these modifications and propose future research directions. These insights aim to guide future researchers in medical image processing and provide valuable perspectives for healthcare professionals or clinicians.

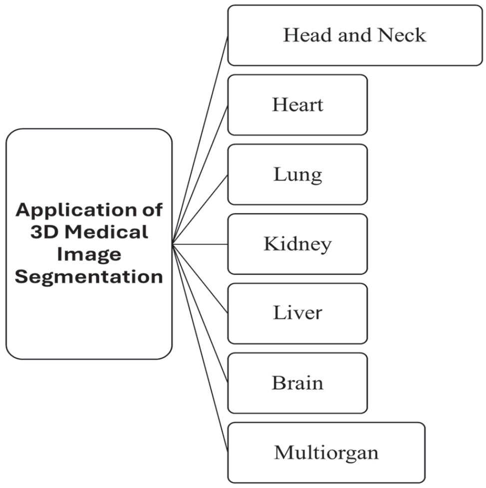

Fig. 3. 3D CNN segmentation applications in the medical imaging field

5. THE SUMMARY OF THE INCLUDED PAPERS

Based on the findings from the included studies, it was observed that not all studies incorporated pre-processing steps as part of their methodology as can be seen in Figure 2, despite its potential to enhance image quality and facilitate better feature extraction. This variability in methodology highlights differing approaches, with some studies relying entirely on the robustness of their 3D CNN models for effective segmentation. Generally, many studies employed single-stage segmentation pipelines where the 3D CNN directly processes the input images. In contrast, there are also studies that utilized a two-stage pipeline, which includes a preliminary Region of Interest (ROI) extraction step. This additional ROI extraction stage allows the model to focus on specific areas of the image, potentially improving segmentation performance by reducing irrelevant and noisy data.

Studies that incorporated pre-processing techniques alongside 3D CNNs will be discussed in detail in Section 5.1, with an emphasis on how these techniques contributed to improved model performance and segmentation accuracy. In contrast, studies that utilized ROI extraction as part of their two-stage pipeline are analyzed in Section 5.2, highlighting the role of this additional step in optimizing the segmentation process.

Moreover, Section 5.3 explores the contribution of post-processing techniques, which are integral steps in refining the segmentation outputs of 3D CNNs. This respective section details how post-processing methods, such as smoothing, morphological operations, or filtering, are employed to enhance the quality of segmentation results. By structuring the review analysis in this manner, this paper aims to present a comprehensive evaluation of the segmentation pipeline, specifically that pertains to 3D CNN methodologies.

There are several significant trends that have been identified with regard to the design of 3D CNN architectures to address specific challenges of medical imaging dataset characteristics, such as low contrast, noise, or irregular anatomical structures. The adaptability of these architectures reflects their targeted approach to overcoming image-related limitations, emphasizing the importance of architectural customization in achieving effective segmentation outcomes.

In Table 1, we also categorize the segmentation methods based on their backbone networks. The backbone network serves as a reference framework for understanding how these approaches are classified. The identified backbone networks can be generally grouped into 3D U-Net, 3D FCN, 3D CNN, V-Net, and others. It was observed that most of the segmentation models are based on 3D U-Net, highlighting its popularity in medical image segmentation tasks. This preference is likely due to its symmetric encoder-decoder structure, which is particularly effective for capturing multi-scale contextual information and maintaining spatial precision which is crucial for medical imaging applications.

In addition to 3D U-Net, Fully Convolutional Network (FCN) architectures are also frequently employed. FCN eliminates the fully connected layers on the decoder side, enabling pixel-wise predictions and making them suitable for dense segmentation tasks. Meanwhile, vanilla Convolutional Neural Networks (CNNs), which are the foundational architecture for image analysis, have been adapted to 3D applications for volumetric segmentation but often lack the multi-scale feature aggregation of U-Net variants. Furthermore, V-Net is another prominent backbone utilized in medical segmentation. It is a 3D extension of the U-Net design, incorporating residual connections to enhance gradient flow during training, which is particularly beneficial for deeper networks.

In summary, while 3D U-Net remains the dominant choice due to its proven effectiveness, architectures like FCN, vanilla CNN, and V-Net provide additional options, catering to specific requirements of segmentation tasks. Given the focus of this review on 3D CNN-based segmentation methods, a substantial portion of the discussion is dedicated to lung imaging applications. These methods serve as a representative example of the capabilities and variations inherent in 3D CNN-based approaches, making them an ideal case for an in-depth analysis of semantic segmentation strategies. This section will explore the categories and limitations of 3D CNN architectures, as well as highlight the available public lung imaging databases that are commonly used in this field.

Table 1. Summary of the included papers

Reference	Pre-Processing	ROI	3D Backbone Network	Post-Processing
Zhang *et al.* [21]	x	✓	UNET	x
Xu *et al.* [37]	x	x	VNET	x
Shi *et al.* [38]	x	✓	UNET	x
Li *et al.* [39]	✓	x	UNET	x
Jin *et al.* [40]	✓	x	UNET	✓
González Sánchez *et al.* [41]	✓	x	UNET	x
Dalvit Carvalho da Silva *et al.* [42]	x	✓	UNET	x
Ren *et al.* [29]	✓	✓	CNN	x
Nikan *et al.* [43]	✓	x	FCN	x
Gao *et al.* [44]	✓	x	UNET	x
López-Linares Román *et al.* [30]	✓	x	VNET + FCN	x
Chen *et al.* [45]	✓	x	UNET	x
Brahim *et al.* [46]	✓	✓	UNET	x
Zhang *et al.* [47]	✓	x	UNET	x
Yang *et al.* [48]	x	x	UNET	x
Xiao *et al.* [49]	✓	✓	UNET	x
Wang *et al.* [50]	x	✓	VNET	x
Wang *et al.* [51]	x	x	UNET	x
Hussain *et al.* [52]	✓	x	UNET	x
Hossain *et al.* [31]	✓	x	CNN	✓
Zhao *et al.* [32]	✓	✓	UNET	✓
Yang *et al.* [53]	✓	x	CNN	x
Kang *et al.* [54]	✓	✓	UNET	✓
Yang *et al.* [55], [55]	x	✓	FCN	✓
Zheng *et al.* [56]	✓	✓	UNET	x
Xu *et al.* [57]	✓	x	CNN	x
Meng *et al.* [58]	✓	x	CNN	✓
Hu *et al.* [59]	✓	x	CNN	✓

Reference	Pre-Processing	ROI	3D Backbone Network	Post-Processing
Deng *et al.* [60]	✓	x	CNN	x
Alalwan *et al.* [33]	✓	x	UNET	x
Qayyyum *et al.* [61]	x	x	CNN	x
Subramaniam *et al.* [62]	x	x	UNET	x
Sharrock *et al.* [63]	✓	x	VNET	x
Saleem *et al.* [64]	✓	x	UNET	x
Niyas *et al.* [65]	✓	x	UNET	x
Liang *et al.* [66]	✓	x	UNET	x
Li *et al.* [35]	✓	x	UNET	x
Radiuk *et al.* [67]	✓	x	UNET	x
Lin *et al.* [68]	x	x	UNET	x
Feng *et al.* [36]	✓	✓	CNN	✓
Yousefi *et al.* [69]	✓	x	UNET	x
Souadih *et al.* [70]	x	✓	CNN	✓
Dai *et al.* [71]	✓	x	CNN	x
Chen *et al.* [72]	✓	x	CNN	x
Chao *et al.* [73]	✓	x	CNN	x
Baldeon *et al.* [74]	✓	x	CNN	✓
Liu *et al.* [75]	✓	x	CNN	x
Hua *et al.* [76]	x	x	UNET	x
Wang *et al.* [77]	✓	✓	UNET	x
Ao *et al.* [78]	✓	x	CNN	x
Ding *et al.* [79]	✓	x	UNET	x
Xiao *et al.* [80]	x	✓	CNN	x
Yang *et al.* [81]	✓	x	CNN	x
Yang *et al.* [82]	✓	x	UNET VNET	x
Chen et al. [83]	x	x	UNET INCEPTION RESNET	x
Bose et al. [84]	✓	x	UNET	x
Singh et al. [85]	✓	✓	CNN	x

5.1. IMAGE PRE-PROCESSING

According to Table 1, it appears that most studies for semantic segmentation of medical imaging do include pre-processing steps as part of their methodology. To enhance the effectiveness of the training process, 3D medical images are typically preprocessed before being fed into a CNN model. This preprocessing step helps in improving the input data quality due to the presence of unknown noise within the patient's body, which may introduce artifacts. These artifacts can result in unnatural intensity variations, significantly affecting image quality. The outlier voxels generated by these artifacts can negatively impact the performance of deep-learning models during the training process [32]. As a result, several preprocessing techniques have been proposed, including voxel intensity normalization and data augmentation.

Voxel value normalization is commonly applied to CT scan images, as each type of tissue in the scan corresponds to a distinct Hounsfield unit (HU) value. Normalizing the HU scale or applying window clipping enhances the features of the target organ, thereby improving the quality of the training process [29], [30]. Each organ would return different HU scale clipping, for example, head and neck values are in the range of [-200 200] [29], while a lung CT scan would be in the range of [-1000 400] and a kidney CT scan would be in the range of [-100 30] [49].

Data augmentation is another widely used technique to address the challenge of limited training data, a common problem in medical image research [33, 41, 43, 45, 54]. Image augmentation involves generating synthetic data to supplement the existing real dataset, which can be achieved through both simple and complex data generation methods. Simple augmentation techniques include basic image processing operations such as translation, rotation, zooming, and flipping [86]. In contrast, more complex augmentation methods may involve the use of Generative Adversarial Networks (GANs) to generate new data based on specific conditions [87]. Additionally, for brain imaging, skull stripping techniques have been employed to improve segmentation accuracy, as demonstrated in studies by [63], [64], [65], and [66].

5.2. REGIONS OF INTEREST EXTRACTION

Instead of feeding raw input data directly into the 3D CNN architecture, some researchers have chosen to apply ROI extraction approach before the 3D CNN segmentation as you can see in the figure 3. In this method, only a subset of the raw data, specifically the extracted ROI, is input into the 3D CNN model [32].

The primary goal of this approach is to reduce the complexity of the segmentation process and lower computational costs. ROI extraction also serves as an initialization step for subsequent segmentation stages. A notable research trend involves the use of a two-stage CNN segmentation process, where ROI extraction typically focuses on anatomical lesions.

Table 2. ROI extraction methodologies

Type of ROI	Purpose of ROI	Reference
Automatic	Region Selection/ organ localization	[21, 36, 38, 40, 44, 46, 50, 55, 56, 70, 77]
	Organ segmentation	[42]
Manual	Region selection	[29, 85]
	Fixed region selection	[31]
	Statistical calculation	[80]

For instance, Zhang et al. [21] implemented a two-stage segmentation approach, where the first stage involves a coarse ROI extraction, followed by a refinement stage that produces the binary output maps. In their study, automated ROI extraction is performed using a 3D-DMFNet, which detects the femur region and removes irrelevant areas, thus reducing memory usage for the latter refinement stage, which is carried out by the 3D ResUNet model. Similarly, Jin et al. [40] performed both localization and segmentation of the frontal vertebrae slices, utilizing the intensity patterns of the vertebrae for RoI extraction via the U-Net architecture. The concept of employing organ localization methods is commonly applied as a coarse-to-fine segmentation approach, where the organ is first localized, followed by lesion segmentation using a series of CNN networks. A limitation of this approach is that the accuracy of lesion segmentation is heavily dependent on the input from the automated ROI extraction process.

In contrast, Ren et al. [26, 29] employed manual annotation for ROI extraction, utilizing multi-atlas-based segmentation methods. While the studies in [28, 31] relied on the researchers' prior knowledge of the lung's location, opting for fixed region selection on each slice. This also applied in [85], the authors proposed a method to manually enhance sharp edges and shapes around the anomalous region of CT scans before inputting them into the 3D CNN. Additionally, due to the small size of the hippocampus, other research has focused on statistical location-based methods, performing cropping based on calculated regions [80].

5.3. IMAGE POST-PROCESSING

Although, in theory, post-processing should not be required for the CNN model since they are designed to leverage all relevant information to generate optimal results, current network architectures are unable to explicitly enforce certain output constraints, such as 3D connectivity and shape conformity. Therefore, further research is needed to integrate such constraints into the design of network structures. Additionally, overfitting remains a concern, which makes post-processing steps essential for rule-based methods. In this study, a simple 3D connectivity analysis was employed to remove small, isolated regions. Gaussian smoothing was also applied to improve specific cases, while the probability output from the network was utilized to assess the reliability of the segmentation maps, enabling case-specific post-processing adjustments [36].

Jin et al. [40] proposed methods to reduce false positives by excluding small predictions (i.e., those under 200 voxels) and refining the segmentation through binary conversion and connected component analysis. Their approach incorporated mask padding and applied an optimal threshold of 0.75. They also used morphological operations, such as dilation and erosion, to eliminate noise and small patches, resulting in a more than 50% reduction in false positives across various models [31].

Zhao et al. [32] employed a post-processing technique based on kidney anatomy, retaining only those tumor components connected to the kidneys, which significantly enhanced the segmentation performance. In a related study [28], segmentation results were binarized, with a focus on the two largest connected components. Morphological operations were applied to improve accuracy, particularly for small tumors, leading to an improvement of 1.77% and 2.82% in renal tumor segmentation for different training models.

In a study by Yang et al. [55], input volumes were limited to 64 slices, requiring the division of regions into smaller sub-volumes. The segmentation process was refined using majority voting and a 3D conditional random field (CRF) algorithm to correct misclassifications. Similarly, Meng et al. [58] employed fully connected CRFs (FC-CRFs) to refine segmentation boundaries, utilizing CT values and category labels for improved accuracy. In the work of Hu et al. [59], morphological operations were applied to align segmented liver tissues with manual annotations. However, challenges remained in distinguishing organs with similar intensity values, as highlighted by Souadih et al. [70]. In their approach, prior anatomical knowledge combined with mathematical morphology was used to accurately locate the sphenoid sinus, with final segmentations confirmed through largest connected component analysis [74].

6. 3D CNN SEGMENTATION FOR LUNG IMAGING

This section explores various 3D deep-learning techniques applied to lung imaging. Segmenting lung regions is a critical step in the screening and diagnosis of lung-related diseases, such as COVID-19, pneumonia, lung cancer nodules, and other medical conditions [82]. The analysis highlights the unique segmentation challenges posed by each lung disease and how 3D CNN-based algorithms are designed to address these issues, as depicted in Fig. 4 and Table 1.

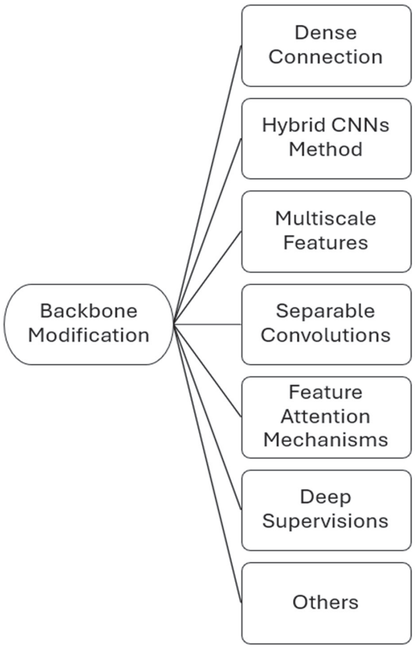

Fig. 4. 3D CNN modifications of segmentation models applied to lung imaging

The reviewed studies categorized 3D CNN segmentation approaches based on their architectural backbones, including U-Net, FCN, CNN, and V-Net, with U-Net being the most frequently employed. The analysis of these models will focus on the modification steps applied to these backbone architectures that aim to enhance segmentation performance and overcome inherent model limitations, often referred to as algorithmic advantages. This section emphasizes lung imaging as a representative use case, as techniques applied to this application are applicable to other medical imaging scenarios. The backbone modifications discussed include dense connections, hybrid CNN methods, multiscale features, separable convolutions, feature attention mechanisms, deep supervisions, and others. Some studies fall into multiple modification categories, as researchers often combined and tailored their models to meet specific objectives or segmentation goals.

6.1. BACKBONE MODIFICATIONS

Many lung lesions are small in size, presenting challenges for segmentation models like U-Net, which is known to be less effective for fine-grained cases. As a result, several studies proposed significant backbone modifications to address these issues. For instance, in [49], a 3D-UNet architecture was enhanced with a 3D-Res2Net module. This hierarchical connection network improves multi-scale feature extraction, capturing finer details and reducing the likelihood of vanishing or exploding gradient problems. The inclusion of 3D-SE blocks recalibrates channel weights, which further optimizes feature representation. The modified architecture, termed as 3D-Res2UNet, achieved a Dice coefficient of 95.30% on the LUNA16 dataset, surpassing the baseline 3D-UNet (89.12%) and 3D-UNet+CRF (93.25%).

In [83], a multiscale block called MSCblock replaced 3D convolution blocks within U-Net. Inspired by the Inception-ResNet architecture, this approach combined parallel convolutional layers of different kernel sizes and identity mappings, enhancing the multi-scale feature capability of the model with a more efficient training process. The MSDS-Unet [48] integrated ResNet modules at each block of a 3D U-Net, enabling the network to capture inter-slice continuity and learn richer feature representations. A two-pathway deep supervision mechanism improved gradient flow, leading to better segmentation performance. These enhancements addressed key challenges like vanishing gradients and insufficient feature representation, making the network robust for complex tasks such as lung tumor segmentation.

Other notable modifications include the SegSEUNet architecture [47], which incorporated Recombination and Recalibration Modules (RRM) with SegSE blocks. This embedding enhanced both spatial and channel recalibration, focusing more on tumor-relevant regions while suppressing irrelevant features. SegSEUNet achieved a Dice coefficient of 0.806 ± 0.120, outperforming traditional SE blocks (0.740 ± 0.144).

The study in [50] proposed an adaptation of V-Net using Parametric ReLU (PReLU) activations and CoordConv layers, which incorporated positional awareness that is critical for pulmonary lobe segmentation. The model achieved an average Dice coefficient of 0.947, significantly surpassing the baseline V-Net model (0.795).

Finally, in [52], a modified 3D U-Net with residual connections was employed for volumetric segmentation. This approach stabilized gradient flow and effectively learned from sparse expert-annotated data, improving the model's Dice scores from 0.730 ± 0.066 (baseline) to 0.763 ± 0.069, when it is combined with gradient-based active learning strategies.

6.1.1. DENSE CONNECTIONS

The integration of a dense Conditional Random Field (CRF) framework significantly improved the segmentation model's ability to delineate precise tumor boundaries. For instance, in [47], the CRF refined segmentation probability maps across scales, mitigating boundary inaccuracies and enhancing spatial consistency. The Dice coefficient improved from 0.842 ± 0.082 to 0.851 ± 0.071, and the Positive Predictive Value (PPV) increased from 0.900 ± 0.107 to 0.917 ± 0.101. Dense connections within the 3D-Res2Net module also ensured efficient gradient flow, enabling superior performance for small and irregular nodules.

6.1.2. HYBRID CNN METHODS

The hybrid CNN modifications come in various strategies such as cascading more than one CNN and combining multiple parallel CNNs, which have been proposed to address the main limitations of a single CNN model. One common approach is coarse-to-fine segmentation, where a coarse segmentation model provides input for a fine segmentation network. For example, in [83], the authors employed a lightweight 3D CNN to capture long-range contextual information and a 2D CNN for fine-grained semantic details. The two networks were fused using a hybrid feature fusion module, which improved computational efficiency and segmentation accuracy. The proposed Hybrid Segmentation Network (HSN) achieved a mean Dice score of 0.898, outperforming standalone 3D CNNs (0.844) and 2D CNNs (0.751).

Another coarse-to-fine approach was proposed in [49], where lung parenchyma was first segmented to isolate the region of interest, followed by a detailed segmentation of lung nodules using a 3D-Res2UNet. This method reduced the influence of surrounding tissues, leading to improved segmentation accuracy for small lesions.

Another hybrid method is the pseudo-3D approach, where 2D feature maps are stacked and processed using 3D convolutions. For instance, in [31], the LungNet framework used stacked 2D slices fused via 3D convolutions, achieving a Dice coefficient of 70.39, outperforming traditional U-Net and LungNet models while maintaining computational efficiency.

6.1.3. MULTISCALE FEATURES

The variation in object sizes and shapes in medical images necessitates the implementation of multi-scale feature extraction. In [84], the D3MSU-Net architecture employed dense dilated convolutions with varying dilation rates to expand the receptive field without increasing the size of the parameters. This design effectively captured multi-scale spatial features, enhancing segmentation accuracy for diverse biomedical datasets. Similarly, MSDS-Unet [48] used multi-scale deeply supervised learning, combining features at different scales to handle heterogeneous tumor characteristics, particularly for small and big-sized tumors.

A multi-scale strategy was also employed in [41], where image cubes of varying dimensions were processed through separate SegSEUNet models. The resultant output maps were further refined using a dense CRF method, resulting in improved segmentation performance. Ablation studies revealed that removing the multi-scale strategy reduced the Dice coefficient from 0.851 ± 0.071 to 0.820 ± 0.115, highlighting its effectiveness.

6.1.4. SEPARABLE CONVOLUTIONS

Deep learning architectures often face challenges due to high computational costs, requiring the development of efficient methods to mitigate these issues. Separable convolution has emerged as one of the main techniques used to reduce computational cost and the number of parameters. For instance, the S3D method proposed in [83] replaces a full 3D convolution with two consecutive layers: a 2D convolution to capture spatial features and a 1D convolution to extract temporal features. This approach effectively decouples the learning process into spatial (inter-slice) and temporal (intra-slice) components. Compared to models utilizing full 3D convolutions, the S3D approach demonstrates superior performance, achieving a 1.1% improvement in Dice evalution.

6.1.5. FEATURE ATTENTION MECHANISMS

Attention mechanisms play a critical role in the segmentation model, particularly in recalibrating feature maps for tumor regions. The SegSE block [47], which is a novel extension of SE blocks, adds spatial recalibration for voxel-specific attention. This mechanism's performance surpasses conventional channel-only recalibration in SE blocks, making it more suitable for segmentation tasks. Comparative studies in the paper demonstrate that SegSE blocks yield better performance than CBAM and SE mechanisms, with a Dice coefficient improvement from 0.740 ± 0.144 (SE) and 0.751 ± 0.179 (CBAM) to 0.806 ± 0.120.

Another approach used the 3D-SE blocks, which are integrated into the Res2Net modules that act as attention mechanisms, enhancing feature map focus by reassigning channel-wise weights. This mechanism improves the model's sensitivity to small or irregular lung nodules, resulting in better segmentation accuracy even for edge features. For example, the proposed network accurately segments smooth ellipse-like and jagged edges, contributing to its high Dice score as shown in [49].

6.1.6. DEEP SUPERVISIONS

Deep supervision is a core innovation in the MSDS-Unet [48] architecture. By integrating multi-level supervision mechanisms, the network incorporates direct side outputs from hidden layers alongside auxiliary tasks. This approach ensures an effective learning process across different stages of the network. The use of hard fusion and soft fusion strategies combines local and global losses, resulting in more accurate segmentation labeling. Furthermore, the deep supervision mechanism allows the network to better handle multiscale features and provides consistent improvements in segmentation accuracy.

6.1.7. OTHERS

A unique contribution in [50] is the use of CoordConv layers, which enhance the conventional convolution operator by adding three additional channels that correspond to the x, y, and z coordinates. These added channels enable the model to leverage spatial location as a "soft constraint," significantly reducing errors in segmenting lobes with overlapping or indistinct boundaries. The inclusion of CoordConv layers improved the overall Dice coefficient from 0.795 (baseline) to 0.916. For example, the left-upper lobe Dice coefficient increased from 0.859 to 0.958 with this modification.

A key innovation of another study in [52] is the introduction of gradient-based sample weighting mechanisms to address the noise in machine-generated pseudo-annotations. The first mechanism evaluates gradient similarity, which reflects the alignment of gradients between pseudo-labeled data and expert-annotated validation data, emphasizing sample trustworthiness. The second mechanism assesses gradient magnitude to measure the informativeness of training samples by identifying those that provide new information to the model. By combining these strategies, the model dynamically prioritizes the most reliable and informative samples during the training process. This approach increased the Dice score from 0.607 (using only gradient similarity) to 0.616 when both strategies were employed on a challenging dataset.

To reduce dependency on extensive expert annotation, the method in [52] incorporates a noisy teacher-based active learning strategy. Machine-generated pseudo-labels from the noisy teacher are used to annotate unlabeled data, while a query function adaptively selects the most informative samples for training. By combining gradient similarity and magnitude weights, the model eliminates less trustworthy samples, ensuring a more accurate optimization process. This strategy significantly enhanced segmentation performance, with Dice scores improving from 0.590 (semi-supervised learning alone) to 0.621 when active learning was applied to the Challenge data. While the model's performance on the Benchmark dataset with the active learning strategy managed to further improve the Dice score from 0.756 ± 0.085 to 0.763 ± 0.069.

6.2. CURRENT RESEARCH LIMITATION / CHALLENGES IN LUNG IMAGING

Based on the previous discussion, it is evident that various deep learning-based 3D CNN segmentation methods have demonstrated promising outcomes in generating medical imaging segmentation maps. At the same time, it can be concluded that researchers have introduced diverse approaches to enhance the performance of basic algorithms. Additionally, it should be noted that there are several limitations observed in the field of medical image segmentation, particularly when dealing with challenges such as small lesion size that often causes class imbalance, and poor image quality, which is normally encountered in certain modalities like CT scans.

Firstly, architectural constraints within these algorithms pose a huge challenge. The absence of self-adaptive mechanisms often restricts the model's ability to achieve optimal performance across diverse datasets. Additionally, certain 3D CNN architectures, such as those proposed in [47], exhibit deficiencies in capturing fine contour details, leading to inaccuracies in segmenting complex anatomical structures.

A significant challenge also lies in multiscale feature learning. Many diseases exhibit multiscale characteristics, requiring the models to effectively capture features across varying scales. Despite efforts to integrate multiscale modules, current methods often struggle to accurately learn details across scales, particularly in detecting small tumors, where features may be subtle and highly variable. While other multiscale techniques such as waterfall connections have also been explored [49], their utility remains largely confined to specialized applications, such as small tumor detection, rather than providing generalizable solutions applicable across a broad range of clinical scenarios. Another prominent limitation of this 3D network is the high computational demands to efficiently execute the deep CNN models. As these models grow increasingly complex with many layers, coupled with advanced modules such as squeeze blocks and multiscale pathways, the computational burden and training time of this 3D model has increased significantly. The requirement for extensive computational resources may render effective deployment in clinical environments impractical due to the limited access to high-performance computing infrastructure [28, 31, 47, 49].

Data scarcity in the medical field also presents a significant challenge. High-performing deep learning models require large and well-annotated datasets for optimal training processes. However, several factors such as privacy concerns, the labor-intensive nature of annotation, and the limited availability of publicly accessible datasets often impede the development of robust models. This data shortage issue can lead to overfitting and diminished generalizability problems, thereby reducing the algorithm's effectiveness across diverse patient populations [28, 88].

In conclusion, despite notable advances in the use of 3D CNNs for medical image segmentation, the field still faces several challenges, including architectural limitations, difficulties in multiscale feature extraction, high computational demands, and constrained data availability. Overcoming these obstacles will necessitate continued research into adaptive and resource-efficient algorithms, potentially benefiting from increased collaboration between the fields of computer science and medicine. This topic will be discussed in the next section.

6.3 FUTURE RESEARCH RECOMMENDATIONS IN LUNG IMAGING

From the findings in section 6.2, there are several future research directions that should be explored, which are further discussed in the following subsections.

6.3.1. Challenges in Medical Image Segmentation Dataset

A primary challenge in medical image segmentation lies in the availability of data. Due to strict privacy concerns surrounding patient information, open access to medical datasets remains limited. It is imperative to revisit and refine protocols for data protection to facilitate the use of anonymized datasets without compromising patient confidentiality. Addressing this issue could significantly benefit the research community. Additionally, there is a notable scarcity of volumetric data necessary for training robust deep-learning models. Collaborative efforts between healthcare institutions, domain experts, and image-processing researchers are essential to expand the availability of such data. Establishing training programs for postgraduate students under the guidance of clinical investigators, who are experts in specific diseases, may also support data collection efforts.

Another pressing issue is the labor-intensive and time-consuming process of creating annotated ground truth data. Semi-supervised learning techniques and transfer learning can be leveraged to mitigate this limitation. Pretrained deep learning models, for instance, can effectively reduce the demand for large annotated datasets by utilizing knowledge transfer across related domains.

6.3.2. Advancements in Network Architecture

Currently, most network architectures for medical image segmentation are heavily based on U-Net, which has demonstrated excellent performance in various applications. However, exploring alternative backbones, such as HRNet, could reveal additional potential. Moreover, reconsidering the parameter size within these architectures is also crucial. Increasing complexity by simply adding more parameters is not always efficient. Strategies like pyramid pooling and dilated (atrous) convolutions have emerged as promising alternatives. Dilated convolutions, in particular, help address multiscale challenges, as diseases often present lesions of varying sizes and shapes depending on their stage. However, careful investigation of dilation rates is necessary to avoid the "gridding" effect that arises when large dilation rates are used.

As networks grow more complex, the associated increase in computational cost must also be considered. Depthwise separable convolutions offer a potential solution by significantly reducing the number of parameters, which is particularly advantageous for 3D medical imaging applications. While reduced parameters will lower computational demands, researchers must ensure that model performance and accuracy are not compromised. Balancing these trade-offs may involve integrating techniques like attention mechanisms or deep supervision to maintain existing model performance.

6.3.3. Generalization of Deep Learning Models

A significant limitation of current deep learning models is their generalizability across various conditions. Most models are developed and tested using data from a single source, which limits their ability to generalize across different conditions that may be encountered when varying imaging instruments and configurations are used to capture the images. Expanding studies to include multicenter datasets could greatly enhance model robustness and applicability.

Additionally, many current research often focuses on the segmentation of a single organ or modality. Broadening this scope to include multiple organs or multimodal imaging data for specific diseases could yield more versatile and generalized models. Encouraging healthcare institutions to collect multimodal datasets for particular organs or diseases would further support this research direction and open new avenues for automated medical screening and diagnosis.

6.4 PUBLIC LUNG CLINICAL DATASETS

Most of the research studies utilized public datasets and a few of them mixed with private datasets. Usually, public datasets are the preferred dataset for comparison purposes so that the generalizability capability of the tested algorithms can be compared fairly [47, 48, 50, 88, 89] shown in table 2. It is also important to consider privacy concerns in the medical field, which limit the availability of certain datasets. However, recent trends show that the use of private datasets has become increasingly important. Most of the public datasets are not too big in numbers, highlighting the need for a hybrid approach of combining public and private datasets to support a more effective training process of deep learning models. This strategic combination approach also helps address the challenges of overfitting and class imbalance, ultimately enabling the models to produce better generalization capability in medical research and applications.

Table 2. Public Lung Clinical Datasets

Public Datasets	Segmentation Tasks	Studies that utilize the dataset
NSCLC-Radiomics	Lung Tumor	[31, 47]
LIDC	Lung Tumor	[47]
LUNA 16	Lung Nodule	[36, 48, 88]
COVID-19 – Ma et al.	Covid-19 Lesion	[89]
COVID-19 Challenge	Covid-19 Lesion	[89]

7. CONCLUSION

This review provides a valuable foundation for those new to the application of 3D CNNs in medical image segmentation. It offers the public health communities and computer science researchers, a clearer understanding of both the advantages and limitations of automated segmentation, particularly within the context of lung disease segmentation tasks. While no single "optimal" method currently exists for segmenting medical images, this paper presents a comprehensive overview of recent advancements in 3D CNN research, serving as a basis for future progress in the field. However, it is crucial to recognize that the deployment of 3D CNN models on real-world datasets remains a significant challenge. To address this, there is an urgent need to amass larger datasets for model training and to explore the potential of synthetically generated data. Furthermore, the development of more robust algorithms that are capable of effectively addressing the multiscale problem is very crucial, given that the variations in lesion and organ sizes across different disease stages differ significantly. This underscores the importance of collaboration between image processing researchers and medical professionals to refine the developed 3D CNN models, ensuring they are aligned with the objectives of having effective and efficient support tools for screening and diagnosis purposes. By fostering such interdisciplinary collaboration, significant strides can be made in improving the accuracy and efficacy of medical image analysis in three dimensions.

8. ACKNOWLEDGEMENT

This work was supported by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme under Grant FRGS/1/2022/ICT02/UITM/02/8.

9. REFERENCES

[1] K. Manikandan, "Blob based segmentation for lung CT image to improving CAD performance", Proceedings of the International Conference on Recent Trends in Information Technology, Chennai, India, 10-12 April 2014, pp. 1-6.

[2] K. T. Navya, G. Pradeep, "Lung nodule segmentation using adaptive thresholding and watershed transform", Proceedings of the 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology, Bangalore, India, 18-19 May 2018, pp. 630-633.

[3] B. B. Verma, S. Saraswat, V. Saraswat, R. Misra, "Interstitial Lung Disease Patterns Classification using Hybrid Features Set and Multi Level Segmentation Implemented by Machine Learning Algorithm", Proceedings of the 8th International Conference on Communication and Electronics Systems, Coimbatore, India, 1-3 June 2023, pp. 1811-1815.

[4] O. W. Herng, A. S. A. Nasir, F. N. Shaari, A. S. A. Sukor, "Superpixels-Based Automatic Density Peaks and Fuzzy Clustering Approach in COVID-19 Lung Segmentation", Proceedings of the IEEE 2nd National Biomedical Engineering Conference, Melaka, Malaysia, 5-7 September 2023, pp. 152-157.

[5] J. Ker, L. Wang, J. Rao, T. Lim, "Deep learning applications in medical image analysis", IEEE Access, Vol. 6, 2017, pp. 9375-9389.

[6] P. Blanc-Durand 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", European Journal of Nuclear Medicine and Molecular Imaging, Vol. 48, 2021, pp. 1362-1370.

[7] S. Shabani, M. Homayounfar, V. Vardhanabhuti, M. A. N. Mahani, M. Koohi-Moghadam, "Self-supervised region-aware segmentation of COVID-19 CT images using 3D GAN and contrastive learning", Computers in Biology and Medicine, Vol. 149, 2022, p. 106033.

[8] M. A. Thanoon, M. A. Zulkifley, M. A. A. Mohd Zainuri, S. R. Abdani, "A review of deep learning techniques for lung cancer screening and diagnosis based on CT images", Diagnostics, Vol. 13, No. 16, 2023, p. 2617.

[9] S. Ebrahimkhani, M. H. Jaward, F. M. Cicuttini, A. Dharmaratne, Y. Wang, A. G. S. de Herrera, "A review on segmentation of knee articular cartilage: from conventional methods towards deep learning", Artificial Intelligence in Medicine, Vol. 106, 2020, p. 101851.

[10] S. Modak, E. Abdel-Raheem, L. Rueda, "Applications of deep learning in disease diagnosis of chest radiographs: A survey on materials and methods", Biomedical Engineering Advances, Vol. 5, 2023, p. 100076.

[11] C. W. Wang, M. A. Khalil, N. P. Firdi, "A survey on deep learning for precision oncology", Diagnostics, Vol. 12, No. 6, 2022, 1489.

[12] S. R. Abdani, M. A. Zulkifley, M. I. Shahrimin, N. H. Zulkifley, "Computer-assisted pterygium screening system: A review", Diagnostics, Vol. 12, No. 3, 2022, p. 639.

[13] Z. Zhou, F. Gou, Y. Tan, J. Wu, "A cascaded multi-stage framework for automatic detection and segmentation of pulmonary nodules in devel-

oping countries", IEEE Journal of Biomedical and Health Informatics, Vol. 26, No. 11, 2022, pp. 5619-5630.

[14] Y. Liu, L. W. L. Yip, Y. Zheng, L. Wang, "Glaucoma screening using an attention-guided stereo ensemble network", Methods, Vol. 202, 2022, pp. 14-21.

[15] X. Xiong, D. H. Steel, B. Obara, "Evaluation of 2D and 3D deep learning approaches for automatic segmentation of the retinal external limiting membrane in spectral domain optical coherence tomography images", Proceedings of the 7th International Conference on Intelligent Informatics and Biomedical Science, Vol. 7, Nara, Japan, 24-26 November 2022, pp. 341-345.

[16] A. Dhillon, G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection", Progress in Artificial Intelligence, Vol. 9, No. 2, 2020, pp. 85-112.

[17] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, B. Gulyás, "3D deep learning on medical images: a review", Sensors, Vol. 20, No. 18, 2020, p. 5097.

[18] Y. He, H. Yu, X. Liu, Z. Yang, W. Sun, S. Anwar, A. Mian, "Deep learning based 3D segmentation: A survey", arXiv:2103.05423, 2021.

[19] S. Diciotti, S. Lombardo, M. Falchini, G. Picozzi, M. Mascalchi, "Automated segmentation refinement of small lung nodules in CT scans by local shape analysis", IEEE Transactions on Biomedical Engineering, Vol. 58, No. 12, 2011, pp. 3418-3428.

[20] H. Cui, X. Wang, D. Feng, "Automated localization and segmentation of lung tumor from PET-CT thorax volumes based on image feature analysis", Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 28 August - 1 September 2012, pp. 5384-5387.

[21] X. Zhang, Y. Zheng, X. Bai, L. Cai, L. Wang, S. Wu, J. Huang, "Femoral image segmentation based on two-stage convolutional network using 3D-DMFNet and 3D-ResUnet", Computer Methods and Programs in Biomedicine, Vol. 226, 2022, p. 107110.

[22] B. Asadi, J. H. Choi, P. Greer, J. Welsh, S. Chalup, J. Simpson, "PO-1686 MRI-Based Synthetic CT Generation for Head and Neck: 2D vs 3D Deep Learning", Radiotherapy and Oncology, Vol. 170, 2022, pp. S1485-S1486.

[23] S. Niyas, S. J. Pawan, M. A. Kumar, J. Rajan, "Medical image segmentation with 3D convolutional neural networks: A survey", Neurocomputing, Vol. 493, 2022, pp. 397-413.

[24] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation", Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, October 5-9, 2015, Part III, pp. 234-241.

[25] N. Elloumi, A. Ben Makhlouf, A. Afli, B. Louhichi, M. Jaidane, J. M. R. Tavares, "CT Images Segmentation Using a Deep Learning-Based Approach for Preoperative Projection of Human Organ Model Using Augmented Reality Technology", International Journal of Computational Intelligence and Applications, Vol. 22, No. 02, 2023, p. 2350006.

[26] M. Tabassum, A. Al Suman, C. Russo, A. Di Ieva, S. Liu, "A Deep Learning Framework for Skull Stripping in Brain MRI", Proceedings of the 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, Sydney, Australia, 24-27 July 2023, pp. 1-4.

[27] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation", Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17-21 October 2016, Part II, pp. 424-432.

[28] F. Milletari, N. Navab, S. A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation", Proceedings of the Fourth International Conference on 3D Vision, Stanford, CA, USA, 25-28 October 2016, pp. 565-571.

[29] X. Ren, L. Xiang, D. Nie, Y. Shao, H. Zhang, D. Shen, Q. Wang, "Interleaved 3D-CNNs for joint segmentation of small-volume structures in head and neck CT images", Medical Physics, Vol. 45, No. 5, 2018, pp. 2063-2075.

[30] K. L.-L. Román et al. "3D pulmonary artery segmentation from CTA scans using deep learning with realistic data augmentation", Lecture Notes in Computer Science, Vol. 11040, 2018, Springer, pp. 225-237.

[31] S. Hossain, S. Najeeb, A. Shahriyar, Z. R. Abdullah, M. A. Haque, "A pipeline for lung tumor detection and segmentation from CT scans using dilated convolutional neural networks", Proceedings of ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing, Brighton, UK, 12-17 May 2019, pp. 1348-1352.

[32] W. Zhao, D. Jiang, J. Pena Queralta, T. Westerlund, "Multi-scale supervised 3D U-Net for kidneys and kidney tumor segmentation", arXiv:2004.08108, 2020.

[33] N. Alalwan, A. Abozeid, A. A. ElHabshy, A. Alzahrani, "Efficient 3D deep learning model for medical image semantic segmentation", Alexandria Engineering Journal, Vol. 60, No. 1, 2021, pp. 1231-1239.

[34] J. Ker, S. P. Singh, Y. Bai, J. Rao, T. Lim, L. Wang, "Image thresholding improves 3-dimensional convolutional neural network diagnosis of different acute brain hemorrhages on computed tomography scans", Sensors, Vol. 19, No. 9, 2019, p. 2167.

[35] P. Li, W. Wu, L. Liu, F. M. Serry, J. Wang, H. Han, "Automatic brain tumor segmentation from Multiparametric MRI based on cascaded 3D U-Net and 3D U-Net++", Biomedical Signal Processing and Control, Vol. 78, 2022, p. 103979.

[36] X. Feng, K. Qing, N. J. Tustison, C. H. Meyer, Q. Chen, "Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images", Medical Physics, Vol. 46, No. 5, 2019, pp. 2169-2180.

[37] J. Xu, J. Liu, D. Zhang, Z. Zhou, C. Zhang, X. Chen, "A 3D segmentation network of mandible from CT scan with combination of multiple convolutional modules and edge supervision in mandibular reconstruction", Computers in Biology and Medicine, Vol. 138, 2021, p. 104925.

[38] D. Shi, Y. Pan, C. Liu, Y. Wang, D. Cui, Y. Lu, "Automatic localization and segmentation of vertebral bodies in 3D CT volumes with deep learning", Proceedings of the 2nd International Symposium on Image Computing and Digital Medicine, Chengdu, China, 13-14 October 2018, pp. 42-46.

[39] W. Li, Y. M. Tang, Z. Wang, K. M. Yu, S. To, "Atrous residual interconnected encoder to attention decoder framework for vertebrae segmentation via 3D volumetric CT images", Engineering Applications of Artificial Intelligence, Vol. 114, 2022, p. 105102.

[40] L. Jin et al. "Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet", EBioMedicine, Vol. 62, 2020.

[41] J. C. G. Sánchez, M. Magnusson, M. Sandborg, Å. C. Tedgren, A. Malusek, "Segmentation of bones in medical dual-energy computed tomography volumes using the 3D U-Net", Physica Medica, Vol. 69, 2020, pp. 241-247.

[42] R. Dalvit Carvalho da Silva, T. R. Jenkyn, V. A. Carranza, "Development of a convolutional neural network based skull segmentation in MRI using standard tesselation language models", Journal of Personalized Medicine, Vol. 11, No. 4, 2021, pp. 310.

[43] S. Nikan, K. Van Osch, M. Bartling, D. G. Allen, S. A. Rohani, B. Connors, S. K. Agrawal, H. M. Ladak, "PWD-3DNet: a deep learning-based fully-automated segmentation of multiple structures on temporal bone CT scans", IEEE Transactions on Image Processing, Vol. 30, 2020, pp. 739-753.

[44] Y. Gao, R. Huang, Y. Yang, J. Zhang, K. Shao, C. Tao, Y. Chen, D. N. Metaxas, H. Li, M. Chen, "FocusNetv2: Imbalanced large and small organ segmentation with adversarial shape constraint for head and neck CT images", Medical Image Analysis, Vol. 67, 2021, p. 101831.

[45] J. Chen, H. Li, G. He, F. Yao, L. Lai, J. Yao, L. Xie, "Automatic 3D mitral valve leaflet segmentation and validation of quantitative measurement", Biomedical Signal Processing and Control, Vol. 79, 2023, p. 104166.

[46] K. Brahim, A. Qayyum, A. Lalande, A. Boucher, A. Sakly, F. Meriaudeau, "A 3D network based shape prior for automatic myocardial disease segmentation in delayed-enhancement MRI", IRBM, Vol. 42, No. 6, 2021, pp. 424-434.

[47] B. Zhang, S. Qi, Y. Wu, X. Pan, Y. Yao, W. Qian, Y. Guan, "Multi-scale segmentation squeeze-and-excitation UNet with conditional random field for segmenting lung tumor from CT images", Computer Methods and Programs in Biomedicine, Vol. 222, 2022, p. 106946.

[48] J. Yang, B. Wu, L. Li, P. Cao, O. Zaiane, "MSDS-UNet: A multi-scale deeply supervised 3D U-Net for automatic segmentation of lung tumor in CT", Computerized Medical Imaging and Graphics, Vol. 92, 2021, p. 101957.

[49] Z. Xiao, B. Liu, L. Geng, F. Zhang, Y. Liu, "Segmentation of lung nodules using improved 3D-UNet neural network", Symmetry, Vol. 12, No. 11, 2020, p. 1787.

[50] W. Wang, J. Chen, J. Zhao, Y. Chi, X. Xie, L. Zhang, X. Hua, "Automated segmentation of pulmonary lobes using coordination-guided deep neural networks", Proceedings of the IEEE 16th International Symposium on Biomedical Imaging, Venice, Italy, 8-11 April 2019, pp. 1353-1357.

[51] L. Wang, X. Ye, D. Zhang, W. He, L. Ju, Y. Luo, Z. Ge, "3D matting: A benchmark study on soft segmentation method for pulmonary nodules applied in computed tomography", Computers in Biology and Medicine, Vol. 150, 2022, p. 106153.

[52] M. A. Hussain, Z. Mirikharaji, M. Momeny, M. Marhamati, A. A. Neshat, R. Garbi, G. Hamarneh, "Active deep learning from a noisy teacher for semi-supervised 3D image segmentation: Application to COVID-19 pneumonia infection in CT", Computerized Medical Imaging and Graphics, Vol. 102, 2022, p. 102127.

[53] E. Yang, C. K. Kim, Y. Guan, B. B. Koo, J. H. Kim, "3D multi-scale residual fully convolutional neural network for segmentation of extremely large-sized kidney tumor", Computer Methods and Programs in Biomedicine, Vol. 215, 2022, p. 106616.

[54] L. Kang, Z. Zhou, J. Huang, W. Han, "Renal tumors segmentation in abdomen CT images using 3D-CNN and ConvLSTM", Biomedical Signal Processing and Control, Vol. 72, 2022, p. 103334.

[55] G. Yang, G. Li, T. Pan, Y. Kong, J. Wu, H. Shu, L. Luo, J. L. Dillenseger, J. L. Coatrieux, L. Tang, X. Zhu, "Automatic segmentation of kidney and renal tumor in CT images based on 3D fully convolutional neural network with pyramid pooling module", Proceedings of the 24th International Conference on Pattern Recognition, Beijing, China, 20-24 August 2018, pp. 3790-3795.

[56] R. Zheng, Q. Wang, S. Lv, C. Li, C. Wang, W. Chen, H. Wang, "Automatic liver tumor segmentation on dynamic contrast enhanced MRI using 4D information: deep learning model based on 3D convolution and convolutional LSTM", IEEE Transactions on Medical Imaging, Vol. 41, No. 10, 2022, pp. 2965-2976.

[57] M. Xu, Y. Wang, Y. Chi, X. Hua, "Training liver vessel segmentation deep neural networks on noisy labels from contrast CT imaging", Proceedings of the IEEE 17th International Symposium on Biomedical Imaging, Iowa City, IA, USA, 3-7 April 2020, pp. 1552-1555.

[58] L. Meng, Y. Tian, S. Bu, "Liver tumor segmentation based on 3D convolutional neural network with dual scale", Journal of Applied Clinical Medical Physics, Vol. 21, No. 1, 2020, pp. 144-157.

[59] P. Hu, F. Wu, J. Peng, P. Liang, D. Kong, "Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution", Physics in Medicine & Biology, Vol. 61, No. 24, 2016, p. 8676.

[60] Z. Deng, Q. Guo, Z. Zhu, "Dynamic regulation of level set parameters using 3D convolutional neural network for liver tumor segmentation", Journal of Healthcare Engineering, Vol. 2019, No. 1, 2019, p. 4321645.

[61] A. Qayyum, A. Lalande, F. Meriaudeau, "Automatic segmentation of tumors and affected organs in the abdomen using a 3D hybrid model for computed tomography imaging", Computers in Biology and Medicine, Vol. 127, 2020, p. 104097.

[62] P. Subramaniam, T. Kossen, K. Ritter, A. Hennemuth, K. Hildebrand, A. Hilbert, J. Sobesky, M. Livne, I. Galinovic, A. A. Khalil, J. B. Fiebach, "Generating 3D TOF-MRA volumes and segmentation labels using generative adversarial networks", Medical Image Analysis, Vol. 78, 2022, p. 102396.

[63] M. F. Sharrock, W. A. Mould, H. Ali, M. Hildreth, I. A. Awad, D. F. Hanley, J. Muschelli, "3D deep neural

network segmentation of intracerebral hemorrhage: development and validation for clinical trials", Neuroinformatics, Vol. 19, 2021, pp. 403-415.

[64] H. Saleem, A. R. Shahid, B. Raza, "Visual interpretability in 3D brain tumor segmentation network", Computers in Biology and Medicine, Vol. 133, 2021, p. 104410.

[65] S. Niyas, S. C. Vaisali, I. Show, T. G. Chandrika, S. Vinayagamani, C. Kesavadas, J. Rajan, "Segmentation of focal cortical dysplasia lesions from magnetic resonance images using 3D convolutional neural networks", Biomedical Signal Processing and Control, Vol. 70, 2021, p. 102951.

[66] J. Liang, C. Yang, L. Zeng, "3D PSwinBTS: an efficient transformer-based Unet using 3D parallel shifted windows for brain tumor segmentation", Digital Signal Processing, Vol. 131, 2022, p. 103784.

[67] P. Radiuk, "Applying 3D U-Net architecture to the task of multi-organ segmentation in computed tomography", Applied Computer Systems, Vol. 25, No. 1, 2020, pp. 43-50.

[68] M. Lin, Q. Cai, J. Zhou, "3D Md-Unet: A novel model of multi-dataset collaboration for medical image segmentation", Neurocomputing, Vol. 492, 2022, pp. 530-544.

[69] S. Yousefi, H. Sokooti, M. S. Elmahdy, F. P. Peters, M. T. M. Shalmani, R. T. Zinkstok, M. Staring, "Esophageal gross tumor volume segmentation using a 3D convolutional neural network", Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16-20 September 2018, pp. 343-351.

[70] K. Souadih, A. Belaid, D. B. Salem, "Automatic segmentation of the sphenoid sinus in CT-scans volume with DeepMedics 3D CNN architecture", Medical Technologies Journal, Vol. 3, No. 1, 2019, pp. 334-346.

[71] W. Dai, B. Woo, S. Liu, M. Marques, C. Engstrom, P. B. Greer, S. Crozier, J. A. Dowling, S. S. Chandra, "CAN3D: Fast 3D medical image segmentation via compact context aggregation", Medical Image Analysis, Vol. 82, 2022, p. 102562.

[72] J. Chen, Z. Wan, J. Zhang, W. Li, Y. Chen, Y. Li, Y. Duan, "Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet", Computer Methods and Programs in Biomedicine, Vol. 200, 2021, p. 105878.

[73] X. Cao, H. Chen, Y. Li, Y. Peng, Y. Zhou, L. Cheng, T. Liu, D. Shen, "Auto-DenseUNet: Searchable neural network architecture for mass segmentation in 3D automated breast ultrasound", Medical Image Analysis, Vol. 82, 2022, p. 102589.

[74] M. B. Calisto, S. K. Lai-Yuen, "EMONAS-Net: Efficient multiobjective neural architecture search using surrogate-assisted evolutionary algorithm for 3D medical image segmentation", Artificial Intelligence in Medicine, Vol. 119, 2021, p. 102154.

[75] Y. Liu, Z. Zhang, J. Yue, W. Guo, "SCANeXt: Enhancing 3D medical image segmentation with dual attention network and depth-wise convolution", Heliyon, Vol. 10, No. 5, 2024.

[76] X. Hua et al. "A lightweight multi-scale multi-angle dynamic interactive transformer-CNN fusion model for 3D medical image segmentation", Neurocomputing, Vol. 608, 2024, p. 128417.

[77] S. Wang, Y. Wang, Y. Peng, X. Chen, "MSA-Net: Multi-scale feature fusion network with enhanced attention module for 3D medical image segmentation", Computers and Electrical Engineering, Vol. 120, 2024, p. 109654.

[78] Y. Ao, W. Shi, B. Ji, Y. Miao, W. He, Z. Jiang, "MS-TCNet: An effective Transformer–CNN combined network using multi-scale feature learning for 3D medical image segmentation", Computers in Biology and Medicine, Vol. 170, 2024, p. 108057.

[79] W. Ding, S. Geng, H. Wang, J. Huang, T. Zhou, "FDiff-Fusion: Denoising diffusion fusion network based on fuzzy learning for 3D medical image segmentation", Information Fusion, Vol. 112, 2024, p. 102540.

[80] Z. Xiao, Y. Zhang, Z. Deng, F. Liu, "Light3DHS: A lightweight 3D hippocampus segmentation method using multiscale convolution attention and vision transformer", NeuroImage, Vol. 292, 2024, p. 120608.

[81] F. Yang, F. Wang, P. Dong, B. Wang, "HCA-former: Hybrid Convolution Attention Transformer for 3D Medical Image Segmentation", Biomedical Signal Processing and Control, Vol. 90, 2024, p. 105834.

[82] J. Yang, H. Li, H. Wang, M. Han, “3D medical image segmentation based on semi-supervised learning using deep co-training”, Applied Soft Computing, Vol. 159, 2024, p. 111641.

[83] W. Chen, H. Wei, S. Peng, J. Sun, X. Qiao, B. Liu, “HSN: hybrid segmentation network for small cell lung cancer segmentation”, IEEE Access, Vol. 7, 2019, pp. 75591-75603.

[84] S. Bose, R. S. Chowdhury, R. Das, U. Maulik, “Dense dilated deep multiscale supervised U-network for biomedical image segmentation”, Computers in Biology and Medicine, Vol. 143, 2022, p. 105274.

[85] S. P. Singh, L. Wang, S. Gupta, B. Gulyas, P. Padmanabhan, “Shallow 3D CNN for detecting acute brain hemorrhage from medical imaging sensors”, IEEE Sensors Journal, Vol. 21, No. 13, 2020, pp. 14290-14299.

[86] E. Elizar, M. A. Zulkifley, R. Muharar, “Scaling and cutout data augmentation for cardiac segmentation”, Proceedings of the International Conference on Data Science and Applications, Vol. 2, 2023, pp. 599-609.

[87] M. A. Zulkifley, S. R. Abdani, N. H. Zulkifley, “COVID-19 screening using a lightweight convolutional neural network with generative adversarial network data augmentation”, Symmetry, Vol. 12, No. 9, 2020, p. 1530.

[88] Y. Feng, P. Hao, P. Zhang, X. Liu, F. Wu, H. Wang, “Supervoxel based weakly-supervised multi-level 3D CNNs for lung nodule detection and segmentation”, Journal of Ambient Intelligence and Humanized Computing, 2023, pp. 1-11.

[89] X. Wang, P. Teng, P. Lo, A. Banola, G. Kim, F. Abtin, J. Goldin, M. Brown, "High throughput lung and lobar segmentation by 2D and 3D CNN on chest CT with diffuse lung disease", Proceedings of Image Analysis for Moving Organ, Breast, and Thoracic Images: Third International Workshop, Fourth International Workshop, and First International Workshop, Granada, Spain, 16-20 September 2018, Proceedings 3, pp. 202-214.