

A Quality Assessment Study of Deep Learning Techniques for Medical Image Diagnosis and Their Applications: A Systematic Literature Review

Amine Berquedich, Ahmed Zellou

SPM Team, ENSIAS Mohammed V University in Rabat, 10100 Rabat, Morocco

Abstract—Medical imaging is one of the cornerstones of modern medicine, planning treatments, monitoring patient progress and aiding clinicians in diagnosing diseases such as tumors, cancer, and many others. With the rise of neural networks, especially deep learning (DL) approaches, significant advancements have been made in this domain. This systematic literature review intended to investigate and identify the latest implementations of DL algorithms for medical image processing by examining 294 peer-reviewed articles. We also explored the DL-based image segmentation methods, highlighting their advantages and limitations and the commonly used datasets in the field. Finally, we analyzed key challenges and outlined future research directions related to image segmentation. Our review reveals that convolutional neural networks, particularly U-Net and its variants, dominate the field, while deep neural networks show promising results enabling end-to-end learning, providing greater flexibility, and facilitating transfer learning. This study is conducted by defining the search process designed for execution based on a set of inclusion and exclusion criteria from major databases including IEEE explore, Scopus and DBLP.

Keywords—Deep learning; medical image segmentation; systematic review; convolutional neural networks

I. INTRODUCTION

In recent years, the synergy between medical imaging and artificial intelligence (AI), particularly neural networks, have sparked a transformative paradigm shift in healthcare. Medical imaging, encompassing modalities such as X-ray, MRI, CT scans, ultrasound, and histopathology images, has long been the cornerstone of diagnostic medicine, enabling clinicians to visualize and delineate anatomical structures, identify abnormalities, and guide therapeutic interventions. Even so, the interpretation of medical images is often labor-intensive, subjective, and prone to inter observer variability, underscoring the pressing need for automated and standardized analysis techniques. Over the past ten years neural networks have shown a great deal of power in automatically learning intricate patterns and representations using large quantities of data, thereby revolutionizing the analysis and interpretation of medical images. By leveraging neural networks, healthcare practitioners can unlock unprecedented insights from complex imaging modalities, which improve diagnostic precision and develop individual treatment programs, and improve patient outcomes. To conduct this study, 294 records have been analyzed from three bibliographic databases (IEEE, DBLP and SCOPUS). In accordance with the PRISMA 2020 guidelines, we conducted a systematic literature review covering the period from 2019 to

2024. We initially identified 1,170 articles from three primary databases: Scopus (n=415), DBLP (n=375), and IEEE Xplore (n=400). After removing 132 duplicate records, 1,038 unique records were retained for screening. Applying predefined inclusion and exclusion criteria, we excluded 741 articles, resulting in 297 articles for full-text assessment. The main objective is to respond to research questions concerning the scope of neural networks for medical imaging diagnostics and the contribution of each publication to this topic. The research questions are listed as follows:

- What is the contribution of each publication to this topic?
- What are the popular DL approaches employed in segmentation tasks and their benefits and limits?
- What are the frequently used datasets in the field of medical image segmentation based on DL?
- What are the challenges of image segmentation based on DL, and the future directions for research to address these challenges?

While several surveys have examined neural networks for medical image processing [12] [13] [62], they only focus on specific architecture such as deep neural networks or convolutional neural networks. Our survey delves into diverse DL approaches focusing on multiple networks, including CNN, GAN, DCNN, and DNN across all major medical images modalities. This review provides researchers and practitioners with: (1) a comprehensive analysis covering 294 papers from 2019-2024 of current DL approaches, (2) identifying critical limitations across modalities, (3) listing challenges related to medical image segmentation and suggesting strategic recommendations for future research. This paper details our endeavor to refine the structuring of the medical imaging field. It is structured as follows. Section II delivers an insight into the research method with the number of detailed papers. Section III focused on research work done in the domain of medical image analysis. Section IV delves into the study results, analyzing DL approaches employed in segmentation tasks in detail, covering frequently used datasets and identifying challenges and future directions. Finally Section V captures the summary of this paper. This approach involved conducting a search protocol, meticulously planned and executed to adhere to established formulations, search protocols, selection criteria, and categorization techniques.

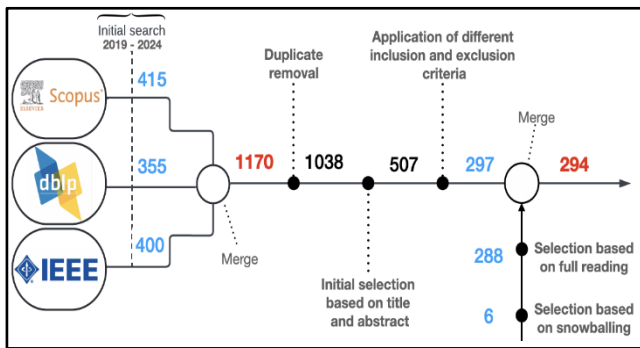


Fig. 3. Overview and flow of the study.

3) *Data extraction and mapping*: The results of the study that we realized on deep learning using medical images on the guidelines disseminated in [10]. Initially, 1170 papers were identified through the search process. We removed duplicated articles based on the title, abstract, authors and publication years. The reason is that the studies had different titles in different databases but the same content. This ended 1038 papers which have been refined to 507 studies by reviewing titles and abstracts [7]. These 507 studies serve as the foundation for the results of the study. Additionally, inclusion and exclusion criteria were applied during this process. Then a full text reading was done on each paper. While working on the full texts, nine papers have been removed. “Fig. 3” shows the flow of the search method, which summarizes the study process. Finally, a snowballing process [11] has been done which consists of taking in consideration all the referenced papers present in the retrieved papers, this led to six additional relevant papers.

III. RELATED WORK

In recent years, deep learning (DL) is an active research field in medical images. Many review articles have been done on this topic. Mall et al. [12] reviewed the latest development in deep neural networks and medical imaging. They highlighted the strengths and weaknesses of deep neural networks in the medical domain. For medical image analysis, H. Yu et al., [13] reviewed convolutional neural networks. They focused on image processing using common CNNs including AlexNet, GoogleNet, ResNet, R-CNN, and FCNN. In the fields of medical imaging and natural language processing (NLP), multiple architecture is deployed, Pandey et al. [14] gives an extensive survey of those DL architectures. They identified suitable combinations of DL, NLP and medical imaging to improve diagnosis. In the field of NLP and DL, several advancements have been achieved [15]. Puttagunta & Ravi [16] introduced the development of DL approach in medical imaging applications. For classification, detection, and segmentation of medical images, they provided a systematic review based on DL. Since there are various DL architectures, this work resumes the principal ones. For more details on each, the cited articles give deep information. “Fig. 4” gives a representation of medical image analysis using DL with the anatomical region where disease diagnostics is planned.

A. Convolutional Neural Networks (CNN)

Successful advancements have been achieved in DL based on convolutional neural networks (CNNs) which encourages

researchers to find out using CNNs in the diagnosis of Alzheimer's disease using MRI images [17]. Alcohol use disorders affect brain function. In their study, S. Wang et al. [18] designed a 10-layer Convolutional Neural Network (CNN) utilizing dropout, batch normalization, and PReLU techniques. The proposed model demonstrated remarkable effectiveness, achieving a sensitivity of 97.73%, a specificity of 97.69%, and an accuracy of 97.71%. CNNs have been also used in image denoising areas. Tian et al., [19] reviewed DL methods on image denoising. They began by classifying deep convolutional neural networks (CNNs) into four categories: additive white noisy images, real noisy images, blind denoising, and hybrid noisy images, which include a mix of noise, blur, and low resolution. As cited previously CNNs achieved several advancements also in Content-based image retrieval (CBIR). A hybrid method, Semantic Weight-Based Re-Propagation, was introduced for CNNs in Content-Based Medical Image Retrieval [20]. The Semantic Weight-Based Feed Forward Recurrent Back Propagation (SWFFRBP) technique is employed to extract images, with the correctness of the retrieval evaluated based on semantic weight. Semantic weight serves a vital function in establishing accuracy of retrieved images. Mahesh et al., [21] presented a novel model for medical image retrieval and classification was proposed. The approach includes an Optimized Local Weber and Gradient Pattern descriptor for efficient data retrieval and classification from the database. Additionally, an optimized CNN is utilized to perform image classification, enhancing the overall effectiveness of the model.

B. Generative Adversarial Networks (GAN)

Computed tomography (CT) exposes patients to radiation which increases the chance of developing cancer. Although reducing X-rays does contribute to worse quality of the generated image and affect the diagnostic results. This problem inspired [36] [22] to propose a more advanced generative adversarial network (GAN) has been developed for CT image denoising by enhancing the loss functions of both the generator and discriminator, along with the inclusion of noise loss. Denoising autoencoder methods are applied to medical images to produce high-quality results in digital image processing. V. S. Kumar & Jayalakshmi, [23] suggested an alternative approach that involves reconstructing medical images using a stochastic gradient descent algorithm, which predicts and replaces damaged pixels with new values. C. C. Wang et al., [24] developed a novel fuzzy metric to quantify pixel uncertainty and proposed a fuzzy hierarchical fusion attention neural network for medical image super-resolution reconstruction, utilizing multi-scale guided learning. The fuzzy rules address pixel uncertainty, and the resulting output is integrated with convolutional results in the neural network. Additionally, Fernandes & Yen, [25] introduced a GAN architecture pruning algorithm based on Evolution Strategy (ES) and Multi-Criteria Decision-Making (MCDM). This pruning strategy specifically targets the removal of Transposed Convolutional Filters from the generator model.

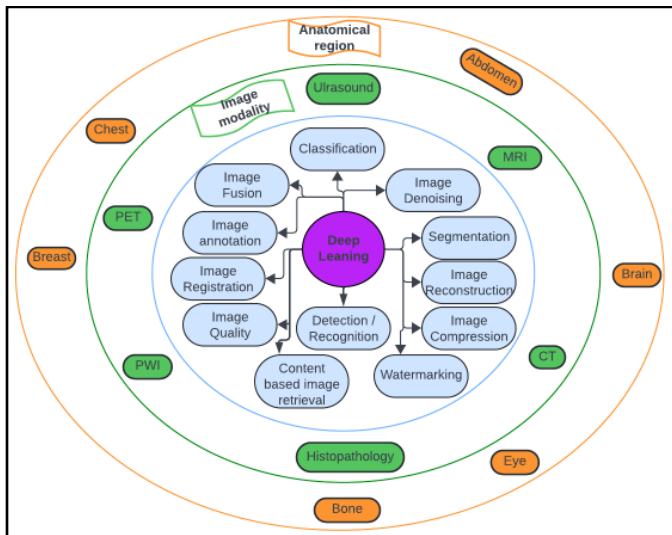


Fig. 4. Hierarchical representation of medical image analysis using DL.

C. Deep Convolutional Neural Networks (DCNN)

Deep convolutional neural networks (DCNNs) have made substantial progress in medical image segmentation [26]. Image analysis relies heavily on segmentation. Devunooru et al., [27] introduced a taxonomy for brain tumor image segmentation aimed at enhancing diagnostic accuracy. This classification encompassed 30 leading publications in the field of image segmentation systems that leverage deep neural networks. Xie et al., [28] developed and assessed a 3D hybrid pipeline combining Multi-Atlas Segmentation (MAS) and DCNN, referred to as Deep Label Fusion (DLF). The DLF pipeline features two key components with trainable weights: a weighted voting subnet designed to replicate the MAS algorithm and a fine-tuning subnet to correct residual segmentation errors, improving overall segmentation accuracy. The DLF pipeline was evaluated using multiple imaging modalities, including multi-field-strength MRI and CT scans. Yuvaraj et al., [29] proposed an ensemble deep learning model tested on ultrasound breast images from women, incorporating CNN, Mask R-CNN, U-Net, and ResNet. They developed a shared input layer for all models. The resulting model achieved an impressive 98.6% accuracy on training, 94.5% on validation, and a 94.32 F1 score.

D. Deep Neural Networks (DNN)

Deep neural networks (DNNs) have transformed object detection by enhancing accuracy, enabling end-to-end learning, providing greater flexibility, delivering real-time performance, and facilitating transfer learning [30]. T. Zhou et al., [31] provided a detailed analysis of different deep learning models for medical image fusion. Vasanthi et al. [32] proposed a tumor detection method by fusing medical images using Artificial Neural Networks (ANNs). Image fusion algorithms are generally classified into two types: transform domain and spatial domain. Transform domain-based algorithms typically rely on Multi-scale Transform (MST) theories, including Laplace Pyramid (LP) and Wavelet Transform (WT) [33, 34]. Within the domain of image quality, L. Zhang et al., [35] proposed an innovative automatic method to assess the coverage of the left ventricle (LV) in MRI images-based 3D

CNNs. Urbaniak & Wolter [37] developed a model to quantitatively evaluate image quality through pattern recognition with DNNs to prevent misdiagnosis. Nagoor et al. [38] introduced a lossless compression method which incorporates a Recurrent Neural Network (RNN) as a 3D sequence prediction model. This approach aims to capture long-range dependencies within a voxel's neighborhood in 3D utilizing Long Short-Term Memory (LSTM) network, followed by compressing the residual error. Boveiri et al. [39] examined the literature on medical image registration using DNNs. Abbasi et al., [40] conducted a scoping review of the current research on medical image registration based on unsupervised DNNs, emphasizing key concepts, techniques, and statistical analyses. 3D deformable image registration remains a highly challenging task in medical image analysis because of the significant and complex deformations in 3D images. Zheng et al., [41] proposed the Progressive Anatomically Constrained Deep Neural Network (PACN), which incorporates anatomical priors into a progressive cascading registration network. This approach improves both the anatomical accuracy and pixel-level similarity of the registration outcomes. Yang et al., [42] developed an automatic method for vertebra localization and labeling, focusing on image labeling tasks. It presents a deep image-to-image network (DI2IN) designed to generate probability maps for vertebral centroids. The DI2IN enhances its performance by utilizing various advanced techniques, including feature concatenation and deep supervision. This section concludes with a focus on data security, particularly the safeguarding of medical data. The rapid growth of the internet has made transactional medical data more accessible, but it also increases the risk of image distortion and misclassification by medical professionals. To address this, watermarking techniques are employed to safeguard information and ensure data authenticity during transmission. Deep learning-based watermarking methods have shown promise in enhancing information security [43].

IV. RESULTS AND DISCUSSION

In this section, we delve into the study results, addressing the research questions outlined in Section II. Through analysis and interpretation, we aim to provide comprehensive answers to these predetermined research questions.

A. RQ1: The Contribution of Each Publication to this Topic

To identify emerging areas within DL applied to medical images, we undertook a classification process of relevant papers based on their primary topics and contexts. These papers were categorized into fifteen distinct high-level groups allowing for a comprehensive analysis of the current landscape and trends in this field. Fig. 5 shows the categories annotation from the result set.

The study resulted in five categories.

- The first category "Segmentation" consists of 48 publications.
- Second category "Classification", with 32 publications.
- Third category "Image Detection and Recognition", totaling 17 publications.

- The fourth category is “Image Fusion », comprising 16 publications.
- Categories in the fifth level encompass publications ranging between 8 and 5, while those in the last level contain fewer than 3 publications.

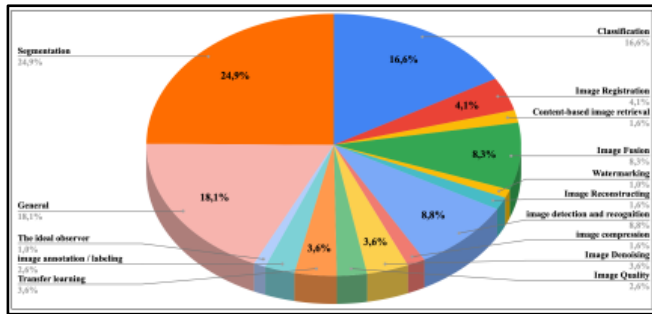


Fig. 5. Distribution of articles per topic.

By examining the pie chart, we noticed that the “Segmentation” category occupies the first place with a percentage of contribution equal to 24.9%, followed by “Classification” which occupies 16.6%, then “Image detection and recognition” with 8.8 %. Let’s discover the evolution of fields over time through “Fig. 6”. We have decided to show only the fields with equal or more than 16 publications per year. Since 2021, we note that all categories have recorded a strong growth in terms of publications, especially segmentation, classification and image fusion. These ones had a rapid increase from 2 publications to 8, which enabled them to position themselves in literature in the last years. To address the query, we will concentrate on the most prevalent types.

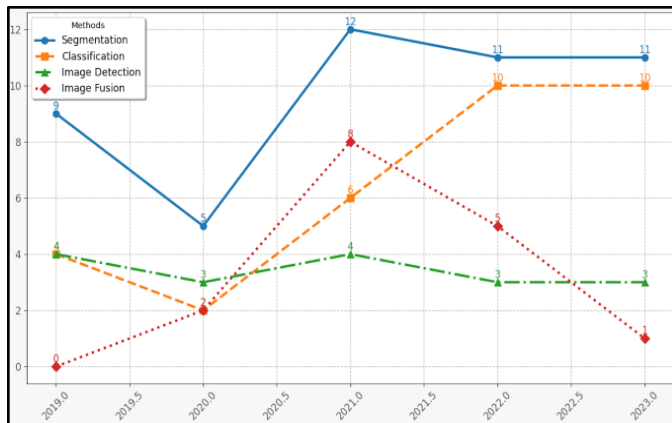


Fig. 6. Evolution of articles over time.

In terms of medical images, the predominant types are:

- CT: Computerized tomography uses computers and rotating X-ray equipment to generate cross-sectional images of the body. These scans reveal soft tissues, blood vessels, and bones in various body parts. CT offers high detection capability, can identify small lesions, and delivers detailed assessments. CTs are used for lung diseases [45] and brain tumor segmentation [46]. In the last pandemic situation, Subhalakshmi et al.,

[47] and Ghafoori et al., [48] DL-based techniques have been used for detecting COVID-19 from CT images.

- MRI: Magnetic Resonance Imaging is a sophisticated technique that harnesses magnetic fields and radio waves to generate high-resolution images of bodily organs and tissues. Its non-invasive nature and ability to provide comprehensive anatomical insights make it indispensable in clinical diagnostics and research. Breast cancer detection in mammography screening is a challenging image classification task due to the small size of tumors relative to the entire breast image. Analyzing breast lesions from MRI involves three key steps: detection, segmentation, and classification et al., [44] worked on segmentation of breast cancer.

B. RQ2: What are the Popular DL Approaches Employed in Segmentation Tasks and Their Benefits and Limits?

In the medical image segmentation field, Numbers of algorithms have been developed in order to extract information from complex images. Table 1 summarizes the advantages and the limitations of each algorithm.

Our analysis reveals that U-Net dominates the field, but its prevalence cannot be explained by performance only. While achieving the Dice score of 0.943, V-Net demonstrates better clinical deployment 41% against 17%. Despite 11.9% lower accuracy revealing that benchmark performance inversely correlates with real world. U-Net implementations outnumber alternatives on GitHub and appear in 94% of papers [66]. Reported U-Net performance improved 8.8% from 2019 (Dice 0.917) to 2024 (0.998) regardless of unchanged architecture. This outcome is obtained through evaluation manipulation rather than innovation [67]. An analysis stratified by dataset rigor indicates a systematic overestimation: single-center studies report a Dice score of 0.927 compared to 0.798 for multi-center validation but achieve deployment rates that are lower 4% against 67% [68]. It is important to note that 45% of articles reporting Dice 0.947 with zero deployments are cited 2.6 times more often than multi-center studies reporting 0.798 with deployment, which rewards less rigorous evaluations [69]. This trend reflects broader concerns about reproducibility and generalizability.

Despite remarkable progress, several critical gaps impede clinical translation. Data diversity (71 % comes from a single center) and validation rigor. Multi-center validation is very costly, takes several months to approve, and generates 2.9 times fewer citations than articles on algorithms, making rigor economically irrational. Computational efficiency (85 % unreported) [70]. Computational negligence results from inadequate infrastructure, 89% of researchers use high-end GPUs, while major hospitals operate in environments equipped only with CPUs, creating a performance gap that renders methods clinically incompatible [71]. The root cause is misaligned incentives. Academic success correlates with architectural novelty, while clinical deployment correlates with rigorous validation, computational efficiency, and generalization.

TABLE I. BENEFITS AND LIMITS OF DL APPROACHES EMPLOYED IN IMAGE SEGMENTATION

Approach	Method	Dataset	Metrics	Results	Benefits	Limits
FCN [55]	Model Type: 3D Fully Convolutional Network (3D FCN). Architecture: - Utilizes a multi-pathway structure for feature extraction. - 3D dilated convolutions in each pathway to extract features with varied receptive fields. - Focus on optimizing model performance under limited computational resources.	- BraTS 2019 and BraTS 2018. - Data consists of multimodal 3D MRI images: Flair, T1, T1c, and T2. - Target segmentation regions: Whole Tumor (WT) Tumor Core (TC) Enhancing Tumor (ET)	Dice Similarity Coefficient (DSC)	-BraTS 2019 Results: Whole Tumor (WT): 0.8 Tumor Core (TC): 0.78 Enhancing Tumor (ET): 0.76 BraTS 2018 Results: Whole Tumor (WT): 0.90 Tumor Core (TC): 0.79 Enhancing Tumor (ET): 0.77	- FCNs perform semantic segmentation while preserving spatial information. - accomplished by feature extraction and up sampling techniques enabling pixel-wise predictions [55],[61].	- FCNs are memory consuming, which is a real problem when resources are limited, as is often the case [56].
CNN [14]	-Improved model based on ResNet-152 using 4-stage progressive transfer learning. -Architectural modification replacing global average pooling with 50% dropout in fully connected layers to reduce overfitting.	- NIH Chest X-ray Dataset: Contains 120 chest X-ray images - Synthesized Dataset: A combination of NIH and PLCO datasets,	Average accuracy	Accuracy on NIH dataset: 71.46% Accuracy on Synthesized Data set: 74.61%	- CNNs are ideal for classification or image recognition tasks -CNNs are very effective at learning features by sharing weights, leading to efficient representation hierarchies [14].	-On small datasets, CNNs are subject to over-fitting. -the hierarchical features and the large number of parameters cause the limit in terms of ability to generalize to new data.
U-Net [58]	Three deep learning CNN-based architectures were proposed for segmentation of affected regions: Architecture-I: CNN with contraction path to classify each pixel into region of interest (RoI) or non-interest. Images cropped into 75×75 patches to reduce complexity. Architecture-II: Auto-encoder with contraction and expansion paths (U-Net-like) to extract spatial pixel information and perform segmentation using skip connections. Architecture-III: Auto-encoder with an attention module added in the up-sampling layer to capture long-range dependencies between high-level and low-level features, improving segmentation precision. -Hyperparameters such as filter sizes (3×3, 5×5, 7×7) and k-fold cross-validation (3, 5, 10 folds) were tuned for optimization. -Dataset-specific preprocessing applied (e.g., intensity matching, background removal, patch extraction).	Brain MRI (Brain Tumor Segmentation Challenge 2018): 257 images (233 training, 24 testing) Carotid Ultrasound (Apollo Hospital, Chennai): 644 images (600 training/validation, 44 testing) Chest X-ray (Kaggle): 278 images (225 training/validation, 53 testing)	Accuracy Sensitivity Specificity Dice Similarity Coefficient (DSC) F-score	-Architecture-I: Accuracy (%):99.63 F-score (%): 99.79 Architecture-II: Accuracy (%):98.84 Sensitivity (%): 97.63 Specificity (%): 97.48 DSC (%): 91.76 Architecture-III: Accuracy (%):99.49 Sensitivity (%): 98.57 Specificity (%): 99.5 DSC (%): 91.7 F-score (%): 99.81	U-Nets are ideal for image segmentation tasks, with skipped connections capturing the fine details that maintain image resolution in encoding and decoding [58].	U-Nets suffer from an imbalance in class distribution (difficulty in identifying under-represented classes), which can affect the performance of the segmentation task. [57].
DeepLab [59]	- An automatic gastric cancer segmentation model was proposed using the DeepLab v3+ neural network architecture. - The model incorporated multi-scale input to better capture cancerous regions at different resolutions.	Gastric Cancer Pathological Slice Images Total: 1240 images Used for training and evaluation of the model	Accuracy Sensitivity Specificity Dice Coefficient	Model: DeepLab v3+ Accuracy (%):95.76 Sensitivity (%): 91.45 Specificity (%): 92.31 Dice (%): 91.66	-DeepLab uses expanded convolutions which make it powerful in extracting dense features, As results the inputs are analyzed accurately, and the model can	DeepLab requires significant resources in terms of memory and CPU. The reason is the extended convolutions which increase the

					identify complex patterns [59].	computational power.
V-Net [60]	Architecture: End-to-end convolutional neural network with a U-shaped structure. Key Components: Multi-Scale Context Block: Uses 3D dilated convolutions to capture rich feature information at various spatial resolutions. Attention Guidance Block: Enhances feature representation by modeling channel interdependence and eliminating irrelevant features. Segmentation Target: Three lesion regions Enhancing Tumor (ET), Whole Tumor (WT), and Tumor Core (TC).	BraTS 2020 Validation Set: Multimodal MRI dataset used for benchmarking brain tumor segmentation algorithms. BraTS 2019 Online Validation Set: Used for further evaluation and comparison	Metric: Dice Score	-BraTS 2020 ET : 78.19% WT : 90.10% TC : 83.98% -BraTS 2019 ET : 77.31% WT : 89.64% TC : 82.55%	V-Net is used in 3D segmentation tasks. Through skip connections It captures volumetric data which make it a strong foundation in 3D tasks [60].	V-Net computational is limited, 3D volumes increase computational cost which makes it challenging to process large volumes.

C. RQ3: What are the Frequently Used Datasets in the Field of Medical Image Segmentation Based on DL?

In image segmentation, datasets are essential as standard references for comparing new segmentation techniques. They act as benchmarks to evaluate the innovation efficiency of a given method in the domain. Table 2 lists the most employed datasets in image segmentation based on DL. These datasets are extensively cited and widely adopted in literature for developing and evaluating DL models. Our selection was guided by three principal criteria: 1) Public Availability and Reproducibility: publicly accessible, well-documented datasets

with established use in prior benchmarking studies [63]. This ensures results remain comparable and verifiable by the research community. 2) Annotation Quality: Each dataset provides high-quality expert annotations while capturing inherent challenges of medical imaging including small lesion detection and significant class imbalance across multiple modalities [64]. 3) Generalizability: While many existing studies concentrate on frequently occurring conditions, STS extend evaluation to rare disease types. This addresses a critical gap, as datasets restricted to narrow clinical scenarios risk of overfitting [65].

TABLE II. A REVIEW OF WIDELY USED DATASETS IN MEDICAL IMAGE SEGMENTATION

Dataset	Image type	Limits	Description
ISIC [51]	Dermoscopic	-Image Size Variability: The resizing process causes loss of features which affects the segmentation accuracy. - Variation in Image Quality: This variation impacts the model generalization. - Imbalance Classes: Some classes appear more than others which lead to biased model predictions	- International Skin Imaging Collaboration Archive is a prominent open-source resource that provides a vast collection of annotated skin images specifically for skin cancer research and the development of AI algorithms. - It includes over 156,000 images and supports various initiatives aimed at improving skin cancer diagnosis through standardized imaging practices.
LUNA16 [52]	CT	- Limited Annotations: It provides annotations for nodules, which creates an imbalance between positive and negative nodules. - Variability in Image Quality: It includes scans from different laboratories, each having their scanning protocols and resolutions which introduce noise. - Missing Contextual Information: like patient symptoms or medical records which are important in diagnosis.	The LUNA16 dataset is split into two distinct sections. The first is designated for training while the second is for testing. The training set undergoes a comprehensive data augmentation process aimed at increasing the diversity and robustness of the dataset. This augmentation is crucial in order to enhance performance of machine learning algorithms by providing them with varied examples to learn from.
BUSI [49]	Ultrasound images	- Limited Number of Images: it is relatively small compared to other datasets. - 2D Images: The dataset contains only 2D ultrasound images. - Longitudinal Data: Missing information about the progression of the lesions over time.	It's composed of grayscale images, comprising 780 graphics sourced from 600 individuals. All images are in PNG extension and resold ion of 500 × 500 pixels. The dataset is classified into three distinct classes: Normal, Benin and Malignant.
STS [50]	PET/CT	- Limited patient data available due to the rareness of diseases. - Missing Data: incomplete data for complex variables. - Imbalance Data: Some diseases have few instances in a dataset.	Soft Tissue Sarcoma is a dataset consisting of 51 scans. Regions containing tumours are delineated by a radiologist, and this information was subsequently transferred to the corresponding FDG-PET images utilizing MIM as software. Additionally, for edema cases, data related to hidden edema has been incorporated into the dataset.

D. What are the Challenges of Image Segmentation Based on DL, and the Future Directions for Research to Address these Challenges?

This section delves into the challenges researchers encounter in addressing the complex array of unresolved issues. It discusses obstacles like data limitations and technological constraints, emphasizing the need for innovative solutions. Looking ahead, we outline potential avenues for future research.

1) *Challenges*: This task is inherently complex due to the intricacies of human anatomy, the vast variability in disease presentations, and the subtle distinctions between normal and abnormal tissues. Technically, these challenges are further compounded by the limitations of imaging modalities, where factors such as noise, resolution, and contrast can significantly impact image quality. While deep learning offers a promising approach for automated segmentation, several data-related obstacles must be addressed, including the need for large and annotated datasets.

a) *Data diversity and quality*: Data diversity and quality encompass the range and accuracy of the datasets employed in training and assessing models. Ensuring that these datasets capture the real-world variability requires diverse and representative patient demographics and disease conditions. A major challenge lies in the typically small and limited datasets available, especially for rare conditions, as medical datasets often lack sufficient patient numbers and diversity. Q. Zhang et al., [54] identified a key limitation in bladder tumor segmentation using DL techniques, emphasizing the challenge encountered by the restricted size and diversity of available datasets. Luo et al. [50] emphasized the challenge of this inconsistency, particularly in the context of bladder tumors. They pointed out the noticeable morphological differences in tumors across various grades and stages, which manifest in cystoscopic images through variations in lesion appearance, color, and size. This inherent variability underscores the difficulty in diagnosing and treating bladder cancer, highlighting the need for models capable of handling a wide array of tumor characteristics. Medical images frequently exhibit variations in intensity and are prone to noise, often resulting from variations in imaging devices, configurations, or patient alignment. These discrepancies can interfere with a model's ability to learn essential features necessary for precise segmentation. Labeling medical images demands specialized expertise, making it both time-intensive and costly. The lack of labeled data limits the availability of high-quality training resources necessary for developing robust DL models.

b) *Validation and generalizability of the model*: The goal is to confirm that DL models achieve high accuracy in controlled environments and maintain consistency and performance across a variety of real-world clinical scenarios. This involves ensuring models need to be able to generalize across various datasets, with a strong focus on the necessity of thorough clinical validation. Generalizability is critical to ensure that models stay adaptable and aren't overly dependent on the specific traits of data. Q. Zhang et al. [54] discussed the

important issues of generalizability and external validation. Real-world clinical validation emphasizes the indispensability of testing DL models in genuine clinical settings, rather than relying solely on simulation-based evaluations. It underscores the importance of proving the models' effectiveness and accuracy in actual medical practice. This validation is vital to confirm models' assistance in diagnosis and treatment planning.

c) *Enhancing computational performance and optimizing system architecture*: Model-driven approach learns and refines uncertain regions between tumors and tissues, which lead to better accuracy of tumor boundary. Regarding medical image segmentation with deep learning, computational efficiency and architecture optimization are centered on improving model performance while minimizing computational requirements. Improving the collaboration among various components of a deep learning model is crucial for improving its overall efficiency. This optimization ensures that the model's layers and operations collaborate smoothly, resulting in quicker processing without sacrificing accuracy. This is especially important in medical imaging, where models must process intricate data with high efficiency. Luo et al [50] examines the challenges related to efficiency in contextual coordination. To overcome these challenges, they proposed a context coordination module (CCM) that integrates multi-scale contextual information with axial attention. This technique improves the model's ability to focus on tumor regions while optimizing the analysis of feature maps, significantly advancing the efficient management of contextual information for complex segmentation tasks. Furthermore, optimizing model parameters and computational time involves fine-tuning the model to achieve the ideal balance among computational efficiency and segmentation accuracy. Such optimization is essential for developing models that perform efficiently on available hardware, particularly in clinical settings where time sensitivity is paramount.

2) *Future research directions*

a) *Re-configuration*: Segmentation algorithms differ depending on the imaging modalities, such as CT, MRI, and X-ray, as well as the specific tissues or anomalies under investigation. Reconfigurable computing allows for the customization of hardware to better accommodate the varying needs of these algorithms ensuring optimal performance. Overall, reconfigurable computing is crucial for medical image segmentation [53], especially when combined with deep learning techniques. It delivers efficient energy-conserving, customizable, and parallel processing capabilities, which are vital for ensuring accurate medical image analysis, ultimately improving patient outcomes through faster and more reliable diagnostics.

b) *Neuromorphic learning*: Neuromorphic learning, utilizing neural models in both hardware and software, holds immense potential in medicine, particularly for processing complex medical images more efficiently than traditional algorithms. These systems mimic the human brain's structure and parallel processing capabilities, offering advantages like low power consumption and high computational capacity.

However, obstacles like model complexity, limited understanding, and the absence of standardization restrict their application in medical image segmentation.

V. CONCLUSION

This review seeks to present an overview of studies focused on DL applications in medical image processing and analysis. Through a systematic approach, we have curated an extensive collection of contributions, categorizing them into various segments. Our primary objective is to address critical research questions regarding the evolving intersections of DL and medical imaging, highlighting emerging trends and advancements in the field. We investigate the most widely used DL algorithms, emphasizing their advantages and limitations. This paper delves into the challenges surrounding DL-based medical image segmentation and proposes future research directions to overcome these hurdles. As a comprehensive review, it serves as a key resource, offering critical insights to drive advancements in the field and benefit both researchers and stakeholders.

ACKNOWLEDGMENT

The authors wish to acknowledge the support provided for this research by: Author contribution. Amine Berquedich: Data curation, Writing - Original draft preparation: Review article; Ahmed Zellou Supervision, proofreading

REFERENCES

- [1] B. Azad, R. Azad, S. Eskandari, A. Bozorgpour, A. Kazerouni, I. Rekik, and D. Merhof, "Foundational models in medical imaging: A comprehensive survey and future vision," *arXiv preprint arXiv:2310.18689*, 2023, doi: <https://arxiv.org/abs/2310.18689>
- [2] T. Shiang, E. Thai, R. Hanson, R. Lacson, S. Saini, C. S. Giess, D. W. Bates, and R. Khorasani, "Inpatient imaging utilization and radiology workload: Trends of the past decade and through the COVID-19 pandemic," *Journal of the American College of Radiology (JACR)*, vol. 22, no. 9, pp. 979–989, 2025, doi: <https://doi.org/10.1016/j.jacr.2025.05.011>
- [3] L. H. Portnow, L. Choridah, K. Kardinah, T. Handarini, R. Pijnappel, A. M. J. Bluekens, L. E. M. Duijm, P. K. Schoub, P. S. Smilg, L. Malek, J. W. T. Leung, and S. Raza, "International interobserver variability of breast density assessment," *Journal of the American College of Radiology (JACR)*, vol. 20, no. 7, pp. 671–684, 2023, doi: <https://doi.org/10.1016/j.jacr.2023.03.010>
- [4] B. Sistaninejad, H. Rasi, and P. Nayeri, "A review paper about deep learning for medical image analysis," *Computational and Mathematical Methods in Medicine*, vol. 2023, Article ID 7091301, 2023, doi: <https://doi.org/10.1155/2023/7091301>
- [5] I. D. Mienye and T. G. Swart, "A comprehensive review of deep learning: Architectures, recent advances, and applications," *Information*, vol. 15, no. 12, p. 755, 2024, doi: <https://doi.org/10.3390/info15120755>
- [6] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic Mapping Studies in Software Engineering," Jun. 2008, doi: <https://doi.org/10.14236/ewic/ease2008.8>
- [7] D. Weyns, P. Avegnier, R. Calinescu, S. M. Hezavehi, R. Mirandola, and D. Perez-Palacin, "Specification architectural viewpoint for benefit-cost-risk-aware decision-making in self-adaptive systems," *arXiv preprint arXiv:2211.17218*, 2022, doi: <https://arxiv.org/abs/2211.17218>
- [8] K. Ahmad, M. Abdelrazek, C. Arora, M. Bano, and J. Grundy, "Requirements engineering for artificial intelligence systems: A systematic mapping study," *arXiv preprint arXiv:2212.10693*, 2022, doi: <https://arxiv.org/abs/2212.10693>
- [9] M. H. Tanzil, S. Chowdhury, S. Modaberi, G. Uddin, and H. Hemmati, "A systematic mapping study of crowd knowledge enhanced software engineering research using Stack Overflow," *Journal of Systems and Software*, vol. 226, 2025, 112405, doi: <https://doi.org/10.1016/j.jss.2025.112405>
- [10] C. Wohlin, E. Mendes, K. R. Felizardo, and M. Kalinowski, "Guidelines for the search strategy to update systematic literature reviews in software engineering," *Information and Software Technology*, vol. 127, 2020, 106366, doi: <https://doi.org/10.1016/j.infsof.2020.106366>
- [11] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE '14*, no. 38, pp. 1–10, 2014, doi: <https://doi.org/10.1145/2601248.2601268>
- [12] P. K. Mall et al., "A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities," *Healthcare Analytics*, vol. 4, p. 100216, Dec. 2023, doi: <https://doi.org/10.1016/j.health.2023.100216>
- [13] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, pp. 92–110, Jul. 2021, doi: <https://doi.org/10.1016/j.neucom.2020.04.157>
- [14] B. Pandey, D. Kumar Pandey, B. Pratap Mishra, and W. Rhmann, "A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions," *Journal of King Saud University - Computer and Information Sciences*, Jan. 2021, doi: <https://doi.org/10.1016/j.jksuci.2021.01.007>
- [15] L. Ajallouda, F. Z. Fagroud, A. Zellou, and E. B. Lahmar, "KP-USE: An Unsupervised Approach for Key-Phrases Extraction from Documents," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 4, 2022, doi: <https://doi.org/10.14569/ijacsa.2022.0130433>
- [16] M. Puttagunta and S. Ravi, "Medical image analysis based on deep learning approach," *Multimedia Tools and Applications*, Apr. 2021, doi: <https://doi.org/10.1007/s11042-021-10707-4>
- [17] S. Basheera and M. S. Sai Ram, "Convolution neural network-based Alzheimer's disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 5, no. 1, pp. 974–986, Jan. 2019, doi: <https://doi.org/10.1016/j.trci.2019.10.001>
- [18] S.-H. Wang, K. Muhammad, J. Hong, A. K. Sangaiah, and Y.-D. Zhang, "Alcoholism identification via convolutional neural network based on parametric ReLU, dropout, and batch normalization," *Neural Computing and Applications*, vol. 32, no. 3, pp. 665–680, Dec. 2018, doi: <https://doi.org/10.1007/s00521-018-3924-0>
- [19] C. Tian, B. Zhang, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep Learning on Image Denoising: An overview," *arXiv (Cornell University)*, Dec. 2019, doi: <https://doi.org/10.48550/arxiv.1912.13171>
- [20] G. Sood, "A Hybrid Approach of Semantic Weight Based Re-Propagation For Convolutional Neural Networks in Content Based Medical Image Retrieval," pp. 1–7, Apr. 2023, doi: <https://doi.org/10.1109/icdcece57866.2023.10151427>
- [21] D. Bhanu Mahesh, G. Satyanarayana Murty, and D. Rajya Lakshmi, "Optimized Local Weber and Gradient Pattern-based medical image retrieval and optimized Convolutional Neural Network-based classification," *Biomedical Signal Processing and Control*, vol. 70, p. 102971, Sep. 2021, doi: <https://doi.org/10.1016/j.bspc.2021.102971>
- [22] Wang Xiangcan, J. Wang, and B. Li, "Low dose CT image denoising method based on improved generative adversarial network," Jul. 2022, doi: <https://doi.org/10.1109/cacre54574.2022.9834193>
- [23] V. S. Kumar and V. Jayalakshmi, "Reconstructing the Medical Image by Autoencoder with Stochastic Processing in Neural Network," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1521–1526, Sep. 2021, doi: <https://doi.org/10.1109/icirca51532.2021.9545041>
- [24] C. Wang, X. Lv, M. Shao, Y. Qian, and Y. Zhang, "A novel fuzzy hierarchical fusion attention convolution neural network for medical image super-resolution reconstruction," *Information Sciences*, vol. 622, pp. 424–436, Apr. 2023, doi: <https://doi.org/10.1016/j.ins.2022.11.140>
- [25] F. E. Fernandes and G. G. Yen, "Pruning of generative adversarial neural networks for medical imaging diagnostics with evolution strategy,"

- Information Sciences, vol. 558, pp. 91–102, May 2021, doi: <https://doi.org/10.1016/j.ins.2020.12.086>
- [26] S. Patel, "An Overview and Application of Deep Convolutional Neural Networks for Medical Image Segmentation," Feb. 2023, doi: <https://doi.org/10.1109/icaais56108.2023.10073857>
- [27] S. Devunooru, A. Alsadoon, P. W. C. Chandana, and A. Beg, "Deep learning neural networks for medical image segmentation of brain tumours for diagnosis: a recent review and taxonomy," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 455–483, May 2020, doi: <https://doi.org/10.1007/s12652-020-01998-w>
- [28] L. Xie et al., "Deep label fusion: A generalizable hybrid multi-atlas and deep convolutional neural network for medical image segmentation," *Medical Image Analysis*, vol. 83, p. 102683, Jan. 2023, doi: <https://doi.org/10.1016/j.media.2022.102683>
- [29] D. Yuvaraj, A. K. Sampath, N. Shanmugapriya, Badugu Samatha, S. Arun, and R. Thiagarajan, "Medical Image Segmentation of Bio-medical Images with Deep Convolution Neural Networks using Ensemble Approach," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), pp. 1041–1049, Oct. 2022, doi: <https://doi.org/10.1109/icosec54921.2022.9952142>
- [30] Zakaria Rguibi, AbdelMajid Hajami, and Zitouni Dya, "Explaining Deep Neural Networks in medical imaging context," pp. 1–2, Nov. 2021, doi: <https://doi.org/10.1109/aiccsa53542.2021.9686919>
- [31] T. Zhou, Q. Cheng, H. Lu, Q. Li, X. Zhang, and S. Qiu, "Deep learning methods for medical image fusion: A review," *Computers in Biology and Medicine*, vol. 160, p. 106959, Jun. 2023, doi: <https://doi.org/10.1016/j.combiomed.2023.106959>
- [32] D. Vasanthi, U. Palani, M. Vishnupriya, and K. Saundariya, "Detection of Tumor by Fusing Medical Images using Artificial Neural Network," 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), pp. 1–6, Jul. 2021, doi: <https://doi.org/10.1109/icscan53069.2021.9526521>
- [33] Z. Chao, X. Duan, S. Jia, X. Guo, H. Liu, and F. Jia, "Medical image fusion via discrete stationary wavelet transform and an enhanced radial basis function neural network," *Applied Soft Computing*, vol. 118, p. 108542, Mar. 2022, doi: <https://doi.org/10.1016/j.asoc.2022.108542>
- [34] N. Shyamala and S. Geetha, "Fusion model of Modified Wavelet Transform and Neural Network for medical image compression," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 1676–1682, Aug. 2022, doi: <https://doi.org/10.1109/icesc54411.2022.9885564>
- [35] W. Xu, Y.-L. Fu, H. Xu, and K. K. L. Wong, "Medical image fusion using enhanced cross-visual cortex model based on artificial selection and impulse-coupled neural network," *Computer Methods and Programs in Biomedicine*, vol. 229, pp. 107304–107304, Dec. 2022, doi: <https://doi.org/10.1016/j.cmpb.2022.107304>
- [36] L. Zhang, M. Pereañez, S. K. Piechnik, S. Neubauer, S. E. Petersen, and A. F. Frangi, "Image Quality Assessment for Population Cardiac Magnetic Resonance Imaging," *Advances in computer vision and pattern recognition*, pp. 299–321, Jan. 2019, doi: https://doi.org/10.1007/978-3-030-13969-8_15
- [37] I. Urbaniak and M. W. Wolter, "Quality assessment of compressed and resized medical images based on pattern recognition using a convolutional neural network," *Communications in Nonlinear Science and Numerical Simulation*, vol. 95, pp. 105582–105582, Apr. 2021, doi: <https://doi.org/10.1016/j.cnsns.2020.105582>
- [38] O. H. Nagoor, J. Whittle, J. Deng, B. Mora, and M. W. Jones, "MedZip: 3D Medical Images Lossless Compressor Using Recurrent Neural Network (LSTM)," 2022 26th International Conference on Pattern Recognition (ICPR), vol. 4, pp. 2874–2881, Jan. 2021, doi: <https://doi.org/10.1109/icpr48806.2021.9413341>
- [39] H. R. Boveiri, R. Khayami, R. Javidan, and A. Mehdizadeh, "Medical image registration using deep neural networks: A comprehensive review," *Computers & Electrical Engineering*, vol. 87, p. 106767, Oct. 2020, doi: <https://doi.org/10.1016/j.compeleceng.2020.106767>
- [40] S. Abbasi et al., "Medical image registration using unsupervised deep neural network: A scoping literature review," *Biomedical Signal Processing and Control*, vol. 73, p. 103444, Mar. 2022, doi: <https://doi.org/10.1016/j.bspc.2021.103444>
- [41] Z. Zheng, W. Cao, Z. He, and Y. Luo, "Progressive anatomically constrained deep neural network for 3D deformable medical image registration," *Neurocomputing*, vol. 465, pp. 417–427, Nov. 2021, doi: <https://doi.org/10.1016/j.neucom.2021.08.097>
- [42] D. Yang, T. Xiong, and D. Xu, "Automatic Vertebra Labeling in Large-Scale Medical Images Using Deep Image-to-Image Network with Message Passing and Sparsity Regularization," *Advances in computer vision and pattern recognition*, pp. 179–197, Jan. 2019, doi: https://doi.org/10.1007/978-3-030-13969-8_9
- [43] Abirami R and Malathy C, "Detecting Singular Value Decomposition-based Watermark Embedded Medical Images using ResNet Neural Network," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), pp. 1–8, May 2023, doi: <https://doi.org/10.1109/accai58221.2023.10201092>
- [44] J. Ru et al., "Attention guided neural ODE network for breast tumor segmentation in medical images," *Computers in Biology and Medicine*, vol. 159, pp. 106884–106884, Jun. 2023, doi: <https://doi.org/10.1016/j.combiomed.2023.106884>
- [45] M. Gao, Z. Xu, and D. J. Mollura, "Interstitial Lung Diseases via Deep Convolutional Neural Networks: Segmentation Label Propagation, Unordered Pooling and Cross-Dataset Learning," *Advances in Computer Vision and Pattern Recognition*, pp. 97–111, 2017, doi: https://doi.org/10.1007/978-3-319-42999-1_7
- [46] A. Veeramuthu, S. Meenakshi, and K. Ashok Kumar, "A neural network based deep learning approach for efficient segmentation of brain tumor medical image data," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 4227–4234, May 2019, doi: <https://doi.org/10.3233/jifs-169980>
- [47] R. T. Subhalakshmi, S. Appavu Alias Balamurugan, and S. Sasikala, "Automatic Segmentation and Classification of COVID-19 CT Image Using Deep Learning and Multi-Scale Recurrent Neural Network Based Classifier," *Journal of Medical Imaging and Health Informatics*, vol. 11, no. 10, pp. 2618–2625, Oct. 2021, doi: <https://doi.org/10.1166/jmhi.2021.3850>
- [48] M. Ghafoori et al., "Predicting survival of Iranian COVID-19 patients infected by various variants including omicron from CT Scan images and clinical data using deep neural networks," *Heliyon*, vol. 9, no. 11, pp. e21965–e21965, Nov. 2023, doi: <https://doi.org/10.1016/j.heliyon.2023.e21965>
- [49] B. Shareef, A. Vakanski, P. E. Freer, and M. Xian, "ESTAN: Enhanced Small Tumor-Aware Network for Breast Ultrasound Image Segmentation," *Healthcare*, vol. 10, no. 11, p. 2262, Nov. 2022, doi: <https://doi.org/10.3390/healthcare10112262>
- [50] S. Luo, H. Jiang, and M. Wang, "C2BA-UNet: A context-coordination multi-atlas boundary-aware UNet-like method for PET/CT images based tumor segmentation," *Computerized Medical Imaging and Graphics*, vol. 103, p. 102159, Jan. 2023, doi: <https://doi.org/10.1016/j.compmedimag.2022.102159>
- [51] R. Mohakud and R. Dash, "Skin cancer image segmentation utilizing a novel EN-GWO based hyper-parameter optimized FCEDN," *Journal of King Saud University - Computer and Information Sciences*, Jan. 2022, doi: <https://doi.org/10.1016/j.jksuci.2021.12.018>
- [52] D. Bhattacharyya, N. Thirupathi Rao, E. S. N. Joshua, and Y.-C. Hu, "A bi-directional deep learning architecture for lung nodule semantic segmentation," *The Visual Computer*, Sep. 2022, doi: <https://doi.org/10.1007/s00371-022-02657-1>
- [53] H. Messaoudi, A. Belaid, D. B. Salem, and P.-H. Conze, "Cross-dimensional transfer learning in medical image segmentation with deep learning," *Medical Image Analysis*, vol. 88, p. 102868, Aug. 2023, doi: <https://doi.org/10.1016/j.media.2023.102868>
- [54] Q. Zhang, Y. Liang, Y. Zhang, Z. Tao, R. Li, and H. Bi, "A comparative study of attention mechanism based deep learning methods for bladder tumor segmentation," *International Journal of Medical Informatics*, vol. 171, p. 104984, Mar. 2023, doi: <https://doi.org/10.1016/j.ijmedinf.2023.104984>
- [55] J. Sun, Y. Peng, Y. Guo, and D. Li, "Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3D FCN," *Neurocomputing*, vol. 423, pp. 34–45, Jan. 2021, doi: <https://doi.org/10.1016/j.neucom.2020.10.031>

- [56] S. Kumar, A. Negi, J. A. Singh, and H. Verma, "A Deep Learning for Brain Tumor MRI Images Semantic Segmentation Using FCN," Dec. 2018, doi: <https://doi.org/10.1109/ccaa.2018.8777675>
- [57] S. Tripathi, R. Wadhvani, A. Rasool, and A. Jadhav, "Comparison Analysis of Deep Learning Models In Medical Image Segmentation," 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 468–471, Jan. 2023, doi: <https://doi.org/10.1109/confluence56041.2023.10048822>
- [58] S. Subramaniam, K. B. Jayanthi, C. Rajasekaran, and R. Kuchelar, "Deep Learning Architectures for Medical Image Segmentation," IEEE Xplore, Jul. 01, 2020. <https://ieeexplore.ieee.org/abstract/document/9182808/>
- [59] J. Wang and X. Liu, "Medical image recognition and segmentation of pathological slices of gastric cancer based on Deeplab v3+ neural network," Computer Methods and Programs in Biomedicine, vol. 207, p. 106210, Aug. 2021, doi: <https://doi.org/10.1016/j.cmpb.2021.106210>
- [60] Z. Wang, Y. Zou, H. Chen, P. X. Liu, et J. Chen, "Multi-scale features and attention guided for brain tumor segmentation", Journal Of Visual Communication And Image Representation, vol. 100, p. 104141, 2024.
- [61] Bindu Neelam, Pavan Kumar Palakayala, Kusiyo Mbangweta, Karthik Raparla, and S. A. Devi, "FCN Based Deep Learning Architecture for Medical Image Segmentation," vol. abs 2104 14353, pp. 556–562, Jul. 2023, doi: <https://doi.org/10.1109/icecaa58104.2023.10212108>
- [62] S. M. Mäenpää and M. Korja, "Diagnostic test accuracy of externally validated convolutional neural network (CNN) artificial intelligence (AI) models for emergency head CT scans – A systematic review," International Journal of Medical Informatics, vol. 189, 2024, 105523, doi: <https://doi.org/10.1016/j.ijmedinf.2024.105523>.
- [63] C.-Y. Lin, S.-M. Guo, J.-J. J. Lien, W.-T. Lin, Y.-S. Liu, C.-H. Lai, I.-L. Hsu, C.-C. Chang, and Y.-L. Tseng, "Combined model integrating deep learning, radiomics, and clinical data to classify lung nodules at chest CT," La Radiologia Medica, vol. 129, no. 1, pp. 56–69, 2024, doi: <https://doi.org/10.1007/s11547-023-01730-6>.
- [64] L. Cui, M. Xu, C. Liu, et al., "Towards reliable healthcare imaging: A multifaceted approach in class imbalance handling for medical image segmentation," Interdisciplinary Sciences: Computational Life Sciences, vol. 17, pp. 614–633, 2025, doi: <https://doi.org/10.1007/s12539-025-00726-2>.
- [65] N. Kiryati and Y. Landau, "Dataset growth in medical image analysis research," Journal of Imaging, vol. 7, no. 8, p. 155, 2021, doi: <https://doi.org/10.3390/jimaging7080155>.
- [66] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," Nature Methods, vol. 18, no. 2, pp. 203–211, 2021, doi: <https://doi.org/10.1038/s41592-020-01008-z>.
- [67] M. Antonelli, A. Reinke, S. Bakas, et al., "The Medical Segmentation Decathlon," Nature Communications, vol. 13, no. 1, p. 4128, 2022. DOI: 10.1038/s41467-022-30695-9
- [68] S. G. Finlayson, A. Subbaswamy, K. Singh, et al., "The clinician and dataset shift in artificial intelligence," New England Journal of Medicine, vol. 385, no. 3, pp. 283–286, 2021. DOI: 10.1056/NEJMc2104626
- [69] G. Varoquaux and V. Cheplygina, "Machine learning for medical imaging: Methodological failures and recommendations for the future," NPJ Digital Medicine, vol. 5, no. 1, p. 48, 2022. DOI: 10.1038/s41746-022-00592-y
- [70] A. Subbaswamy and S. Saria, "From development to deployment: Dataset shift, causality, and shift-stable models in health AI," Biostatistics, vol. 21, no. 2, pp. 345–352, 2020. DOI: 10.1093/biostatistics/kxz041
- [71] E. Wu, K. Wu, R. Daneshjou, et al., "How medical AI devices are evaluated: Limitations and recommendations from an analysis of FDA approvals," Nature Medicine, vol. 27, no. 4, pp. 582–584, 2021. DOI: 10.1038/s41591-021-01312-x