Review

Deep learning in dermatopathology: applications for skin disease diagnosis and classification

Sana Fatima[1,2] · Muhammad Usman Akram[2] · Sabah Mohammad[2] · Saad Bin Ahmed[2]

Received: 12 December 2024 / Accepted: 12 May 2025
Published online: 28 August 2025

Abstract
Medical image segmentation is pivotal in disease diagnosis and treatment planning across various imaging modalities, including MRI, CT, ultrasound, X-ray, dermoscopy, and histopathology. This systematic literature review, conducted using the PRISMA framework, provides a comprehensive analysis of Deep Learning approaches applied to medical image segmentation, with a focus on dermato-pathology for skin disease diagnosis and classification. Transformer-based models have shown notable improvements over traditional CNN architectures, achieving up to 79.95% accuracy in multitask cancer detection tasks, surpassing CNN-based models that achieved 74.05%. In liver lesion segmentation using CT scans, attention-enhanced U-Net models achieved a 93.4% Dice Similarity Coefficient (DSC) for liver tissue and 77.8% for tumor segmentation. In dermoscopy, self-supervised transformer-based models like G2LL exceeded 80% accuracy, while U-Net-based models for skin lesion segmentation achieved up to 93.32% accuracy. Histopathology image analysis further demonstrated that models incorporating attention mechanisms, such as the PistoSeg framework, improved segmentation precision by up to 7.15% compared to conventional methods. Across various modalities, Deep Learning models consistently outperform traditional methods, with improvements ranging from 5 to 15% in accuracy and segmentation metrics. Despite challenges such as computational demands and the need for large annotated datasets, Deep Learning continues to revolutionize medical image segmentation, offering higher diagnostic precision and outlining future research directions to bridge existing gaps.

Keywords Interpretable models · Medical imaging · Semantic segmentation · Deep learning · Skin histology · Skin lesions

1 Introduction

In computer vision and image processing, segmentation refers to the process of dividing an image into distinct segments or regions. The goal of segmentation is to isolate objects, making it easier to detect and recognize them individually. This process aims to create a more meaningful and easier-to-analyze representation of the image [1]. There are various kinds of image segmentation, but recent research demonstrates that semantic segmentation has been particularly successful in medical imaging [2]. Image segmentation involves dividing the image into distinct regions or parts, each representing different tissues and organs. The information represented by the distinct parts must be interpreted by the prediction models. Such kind of models are useful in time-critical decisions. Deep Learning (DL) is the subset of Machine learning

✉ Saad Bin Ahmed, sbinahm@lakeheadu.ca | [1]Computer Science Department, Faculty of Science and Environmental Studies, Lakehead University, Thunder Bay, ON, Canada. [2]Department of Computer Engineering, College of Electrical & Mechanical Engineering, National University of Sciences and Technology (NUST), Islamabad, Pakistan.

(ML) and uses an artificial neural network with a deep neural network that plays an important role in modeling the complex pattern and the representation of data [3]. The DL algorithms automatically learn and extract the features that differentiate between the various classes of objects in an image. The semantic segmentation has significantly addressed in computer vision along with Deep Learning models that are employed in computational pathology which is influenced by factors such as quality of data, type of output, and the learning methodologies [4].

Semantic segmentation plays a crucial role in medical imaging by classifying each pixel based on the region of interest in the image, allowing for the identification of disease diagnosis at a semantic level [5]. It also provides solutions for medical imaging, autonomous driving, and satellite imagery [6]. Medical images play an important role in clinical monitoring and disease diagnosis. The segmentation of structured biopsy images is particularly important for automated diagnostic systems. Among many medical conditions, skin diseases encompass a wide range of diseases that have become increasingly common. They can significantly impact a patient's overall health, mental well-being, and quality of life [7]. Some skin diseases, such as cancer, have the potential to become life-threatening, making early detection crucial for timely and effective treatment. The non-melanoma skin cancers, including squamous cell carcinoma (SCC), basal cell carcinoma (BCC), and intraepidermal carcinoma (IEC), account for about 90% of the cases [8]. The dermatologists emphasize that accurate semantic segmentation of skin histology images is essential for diagnosing various skin conditions, including cancer, as manual methods are time-consuming [9]. The main objective of the proposed work is to investigate the importance of medical image analysis in semantic segmentation. This paper highlights the recent work by considering the following questions.

- What's the application of medical images in semantic segmentation?
- How to use Deep Learning in medical image semantic segmentation?

This paper is organized into five sections. The research methodology is explained in Sect. 2, whereas the related work has been summarized in Sect. 3, the explanation of medical image modalities is discussed in Sect. 4, the various evaluation metrics have been discussed in Sect. 5, while the research gap is presented in Sect. 6, and the conclusion in Sect. 7.

2 Related work

In this section, we have discussed the applicability of deep neural network architecture with reference to medical image segmentation.

2.1 Image segmentation

The image segmentation is a technique used to analyze images by dividing a digital image into distinct segments and organizing the data meaningfully within each segment. The hierarchy of image segmentation and its sub-types are shown in Fig. 1.

The types of image segmentation are designed to identify specific edges and shapes of various objects and regions within the image and label each pixel individually. The four main types of image segmentation are:

- Semantic segmentation
- Object detection
- Instance segmentation
- Panoptic segmentation

In *Semantic Segmentation*, the pixel labels are assigned to correspond to their respective regions. For instance, all pixels associated with a car would be labeled as the class named "automobile". The goal is to identify and classify each pixel based on its specific object class, making semantic segmentation a crucial component in Deep Learning, artificial intelligence, and machine learning [4, 10]. Despite its difficulties, semantic segmentation has demonstrated strong results across various fields, including agriculture, medicine, transportation [11]. In *object detection*, the semantic segmentation is essential as it operates at a pixel level, understanding the fine details of an image. Most techniques require labeling every pixel in an image with an object class, and predictions encompass both the class and the boundaries of objects [12]. The final output reveals the spatial relationships among objects within the image, such as sky, land, and forest [13].

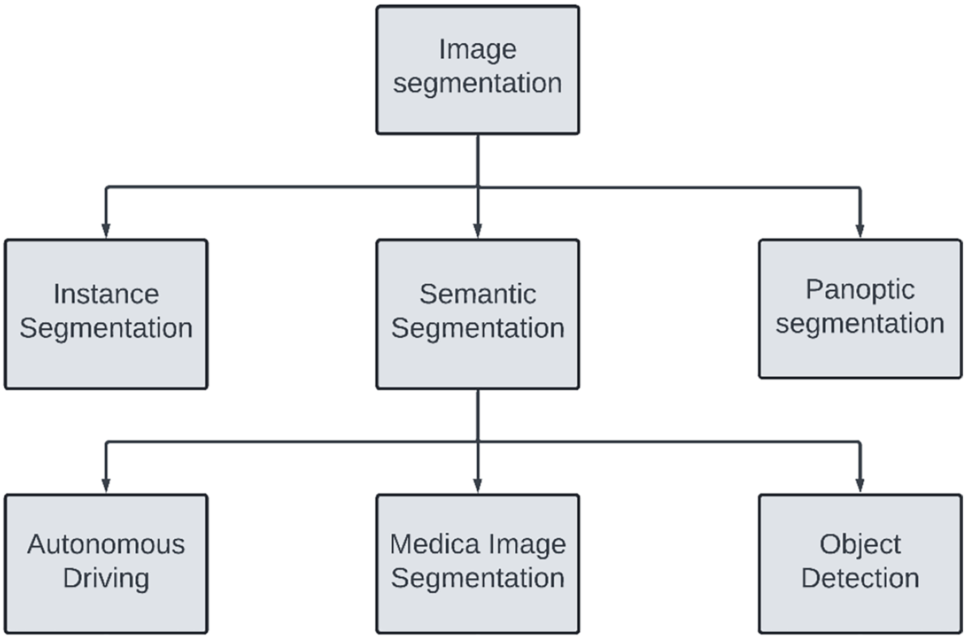


Fig. 1 Image segmentation and its sub-types

Instance Segmentation is akin to semantic segmentation but specifically distinguishes between different objects of the same class [14]. Its primary focus is on identifying and separating entities and objects within semantic categories, such as roads, animals, people, and cars. In computer vision, object instance segmentation is a recent development [15]. It can be further categorized into two types i.e.; detection-level segmentation and image-level segmentation. *Panoptic Segmentation* is a segmentation approach that combines semantic and instance segmentation, labeling each pixel and defining across different instances of the same class. It is particularly beneficial in safety-critical systems since it allows complete object identification and detection in a single frame, removing the risks related to false recognition [16]. Applications of panoptic segmentation include autonomous driving, where it's essential to distinguish between road and sidewalk, as well as between multiple vehicles and pedestrians, and medical imaging, where precise segmentation of organs and tissues helps improve diagnosis.

2.2 Semantic segmentation using deep learning

Traditional machine learning and image processing methods offer solutions to semantic segmentation problems. However, with the extensive applications and advancements in Deep Learning, its benefits for image semantic segmentation have attracted significant attention [17]. Recent advancements in Deep Learning have significantly improved semantic segmentation through the use of neural networks [4]. Deep neural networks have proven particularly effective at semantic segmentation, which involves labeling each region or pixel in an image as either an object or a non-object [18]. Semantic segmentation is a widely used approach known for its capability to analyze complex images at the pixel level. It has various applications, including autonomous driving [19], object detection [12], and medical imaging [20]. However, this study focuses on medical images, as semantic segmentation plays a vital part in tasks including tumor identification [21], organ segmentation [22], and other abnormality detection to support accurate diagnosis and medical care. In medical image analysis, semantic segmentation is often referred to as pixel-level classification [23]. The use of artificial intelligence in medical imaging is rapidly expanding and playing an increasingly important role in this area. It has the potential to revolutionize healthcare services [24]. Also, it's helping doctors to address complex problems and significantly improve diagnostic accuracy and efficiency in medical imaging [25]. Medical image segmentation is used to differentiate between various structures or regions of interest in medical images [26]. It is vital for clinical diagnosis and analysis, as it assists the doctors in finding the correct diagnosis [27]. It facilitates earlier identification of medical issues and more accurate diagnostics [28].

Our study is focused on the four important medical image modalities represented in Fig. 2. These modalities are radiology, ultrasound (UV), dermoscopy, and histopathology. The explanation of each modality is presented in Sect. 4.

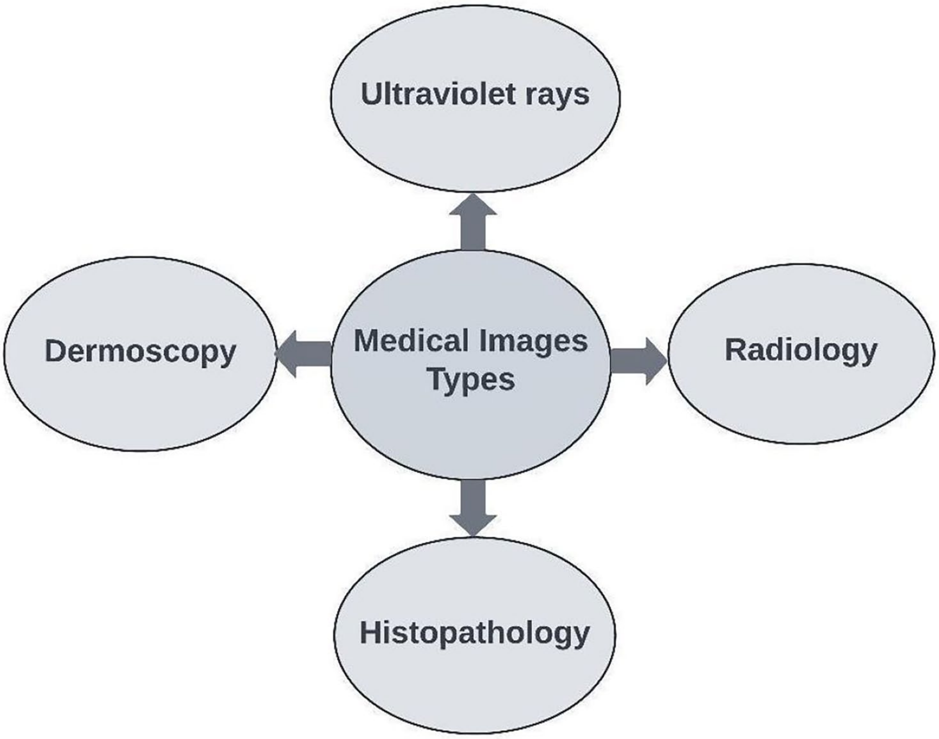
2.3 Deep learning in medical image segmentation

Deep Learning has shown remarkable performance in medical image segmentation [29]. It leads to improved healthcare efficiency and lower costs. A significant area of Deep Learning study, encouraged by the increasing amount of big data, higher processing speeds, and rising Deep Learning algorithms, assists doctors in the detection of skin cancers [30] by analyzing the medical x-ray images. As mentioned earlier, medical image segmentation aims to extract regions of interest (ROIs) from the image data [31] while there are several approaches available for image segmentation. Among them, *Convolutional Neural Networks* (CNNs) are commonly used for medical image segmentation, particularly with the *U-Net* architecture [32]. The CNNs strategies have significantly impacted various fields of medical research such as radiology [33], pathology [34], dermatology [35] that improve the accuracy of diagnosis and proved to be the significant resource in defining treatment plan. Several kinds of deep neural network structures have been designed to assist in medical image segmentation, each adapted to specific imaging modalities [36]. The Deep Learning techniques offer advantages over traditional machine learning models due to the structure of the neural networks [37]. Figure 3 illustrates the process of Deep Learning application in medical images.

The transformer network is another Deep Learning architecture that employs network architecture employs a self-attention mechanism that excels with large datasets [38], delivering improved accuracy compared to traditional methods. The automated systems, especially those utilizing AI and Deep Learning, can diagnose skin diseases much more accurately than traditional, manual methods. [39]. The emphasis of this paper is on medical image segmentation specifically relevant to skin diseases [40]. The skin diseases, such as lesions, scales, and various other symptoms, can have a significant impact on patient's overall health. The most common skin diseases include skin cancers, such as malignant melanoma (melanoma skin tumors) [41] and non-melanoma skin cancers [42], which affect approximately 90% of those diagnosed with this disease. Other skin conditions include acne, genetic disorders like sickle-cell anemia, bacterial infections, psoriasis, fungal infections, and leprosy. These skin diseases are typically diagnosed through biopsy procedures and examinations by pathologists [43]. There are various techniques available for diagnosing these skin conditions [44].

Medical image segmentation also has a key role in histology images [45]. The *Diffusion Convolutional Neural Networks* (DCNN) model is specifically designed for the histological images segmentations like U-Net [46], U-Net++ [47] and reported higher accuracy on the complex dataset [48]. It is evident by assessing the research work presented in the

Fig. 2 Medical image types



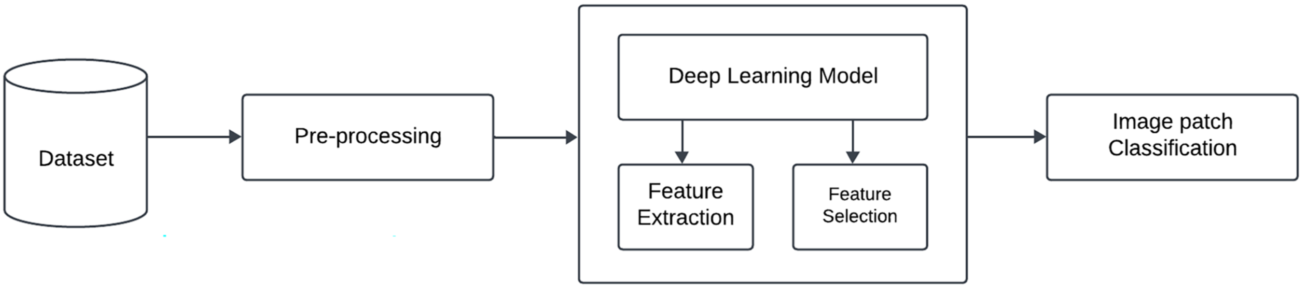


Fig. 3 Depiction of deep learning applications in medical images

last 5–7 years that various Deep Learning methods have been applied to medical image segmentation using different datasets as shown in Table 1.

3 Research methodology

We have reviewed the presented work by the PRISMA guidance. This framework optimizes the efficacy and predictability of the proposed work. The PRISMA technique is a common approach for assuring completion and transparency in research studies. It applies a uniform flowchart for monitoring the study of identification, screening, eligibility, and inclusion, ensuring that the selection process is transparent and neutral. PRISMA is further divided into four categories. As shown in Fig. 4, the strategy started with the *identification* phase, whereby records were collected from numerous databases, including IEEE, Google Scholar, Elsevier, Springer, and Frontier. We have selected articles specifically published from the year 2015 to 2024 that focused on the use of medical image segmentation. But we have emphasized on the work presented in last 5 years. After identifying related data, the next step is to select the relevant papers and this is called a *screening* process. It consist of removing duplication by conducting an initial evaluation according to the title and abstract. By doing so, the papers that were not aligned with the study objectives were not considered. During the *eligibility* phase, the full-text papers were evaluated in detail, with a focus on their applicability towards medical image-based segmentation, specifically for skin disease diagnosis. In the final *inclusion* process, the data chosen were added to the analysis, with a focus on two primary groups such as different skin diseases reported in medical images and multiple kinds of whole-slide skin images. The use of the PRISMA technique enabled a robust and consistent selection of papers for inclusion in the presented review, which improved the validity and reliability of the findings.

When evaluating the existing work, the hypothesis is developed by addressing the following research questions.

1. Which methods and techniques are available for medical image segmentation?
2. Why medical image segmentation is important nowadays?
3. In terms of accuracy and efficiency, how do transformer-based architectures for medical image segmentation compare with traditional CNN-based methods?
4. When compared to the traditional manual method, how much automated systems improve the accuracy of skin disease diagnosis?

These research questions cover key characteristics of medical image segmentation that have already presented in the literature [2, 4, 6–9, 23, 49, 50]. This study systematically reviews 79 number of papers published between 2015 and 2024, focusing on deep learning advancements in medical image segmentation.

3.1 Segmentation methods for medical images

To provide a clear understanding of the key segmentation methods used, this section includes pseudo-code outlining the workflows of both the CNN-based and Transformer-based architectures.

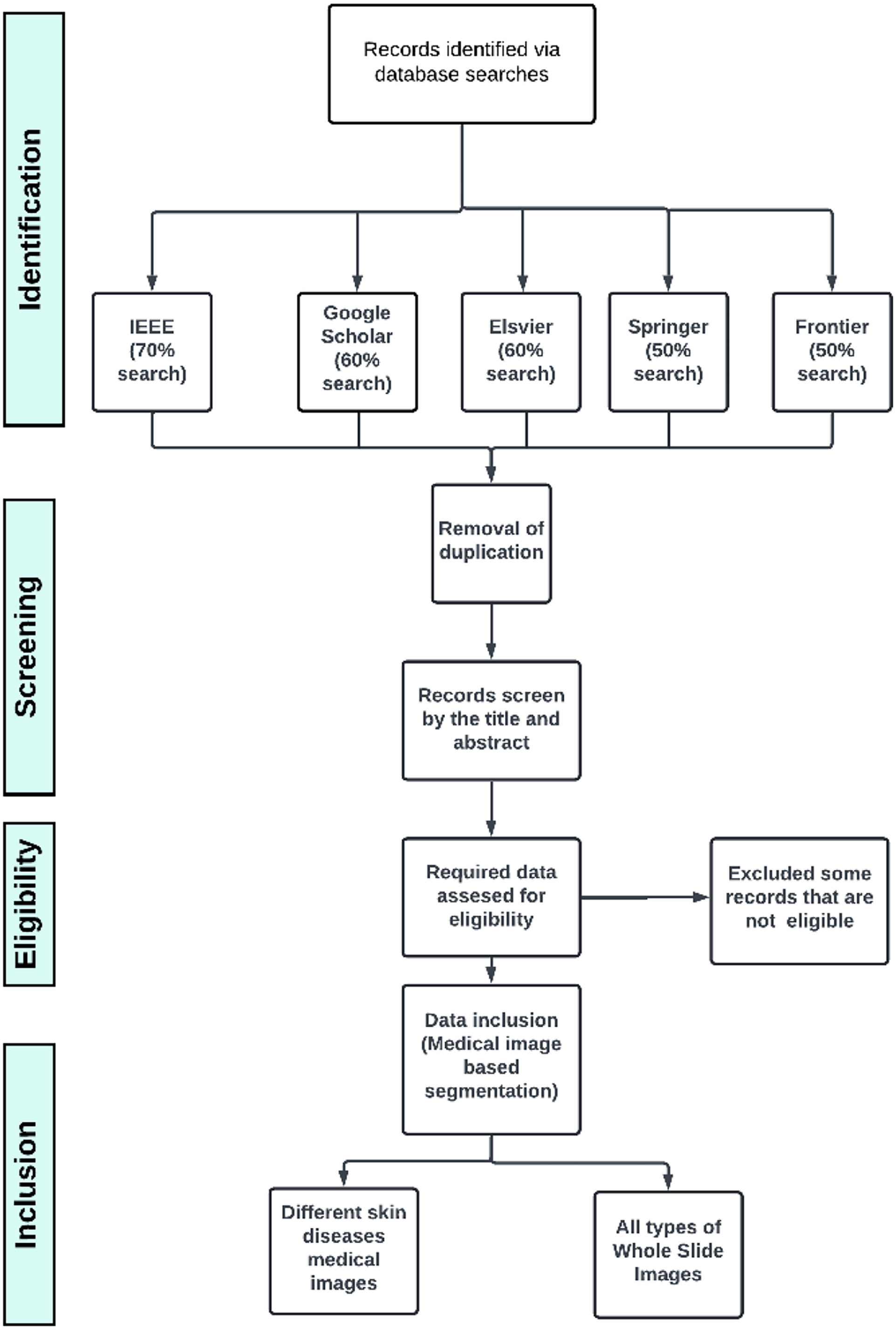

Identification
Screening
Eligibility
Inclusion
Records identified via database searches
IEEE (70% search)
Google Scholar (60% search)
Elsvier (60% search)
Springer (50% search)
Frontier (50% search)
Removal of duplication
Records screen by the title and abstract
Required data assesed for eligibility
Excluded some records that are not eligible
Data inclusion (Medical image based segmentation)
Different skin diseases medical images
All types of Whole Slide Images

Fig. 4 PRISMA schematic representation

3.1.1 U-Net architecture

U-Net is a widely used deep learning architecture for medical image analysis, particularly for tasks like segmentation in skin histology images. Its encoder-decoder structure, combined with skip connections, makes it highly effective in distinguishing between different tissue structures. The encoder captures high-level semantic features through successive convolutional and pooling layers, while the decoder progressively restores spatial details using upsampling layers [28]. The skip connections help retain fine-grained information by directly linking early encoder layers to corresponding decoder layers, ensuring precise boundary detection. This is crucial in skin histology, where accurate

Table 1 Key studies of existing work

Studies	Year	Dataset	Model
Geng et al. [51]	2022	Clinical fetal ultrasound images	SMNet
Subramanian et al. [33]	2023	CT & MRI images	Mobile Net, VGG Net, Dense Net
Xu et al. [52]	2019	BUID(3D breast ultrasound images)	CNN
Vakanski et al. [53]	2020	510 BUS images	U-Net
Seif et al. [54]	2022	LiTS 2017 Challenge dataset (CT images)	U-Net
Turhan et al. [34]	2021	PanNuke	SegNet, UNet, DeepLabV3
Li et al. [55]	2021	2018 Kaggle challenge data Science Bow	CRU-Net
Han et al. [56]	2022	LUAD HistoSeg, BCSS-WCSS	Cam-based, Multi-layer pseudo-supervisor
Gu et al. [57]	2022	Glas	HSRT
Fang et al. [58]	2023	WSSS4LUAD, BCSS-WSSS	PistoSeg
Chen et al. [59]	2023	ISIC-2019, ISIC-2017	G2LL, SSL (MAE, DINO)
Kaur et al. [60]	2021	ISIC-2016, ISIC-2017, PH2	FCNN
Fedorenko et al. [35]	2022	ISIC-2018	U-Net
Iranpoor et al. [61]	2020	PH2	U-Net
Yildiz et al. [62]	2022	CoNIC Challenge 2022	U-Net
Thomas et al. [49]	2021	Queensland Uni dataset (290 H & E slides)	U-Net
Xu et al. [63]	2021	Brain Histological, Skin Histology	Tiscut
Li et al. [64]	2021	Camelyon16	ResNet101, MobileNetV2, U-Net, Ensemble method
Wazir et al. [48]	2022	MoNuSeg, GlaS	ENC & DEC Network, Quick Attention Module, Multi Loss Function
Rasool et al. [65]	2021	Oral cell carcinoma tissues	ENC (ResNeXt blocks), DEC (combining feature learning & fusion mapping back to pixel map)
Asaf et al. [50]	2023	10x magnificent images (Queensland University)	U-Net, EfficientNetB3
Gite et al. [66]	2023	Montgomery County & Shenzhen Hospital X-ray set	U-Net++ (with other benchmarks: FCN, SegNet, U-Net)
Sun et al. [67]	2021	BrainWeb20, IBSR18	SemiGCN, AdvSemiSeg
Pang et al. [68]	2020	T2-weighted volumetric MR images	2D ResUNet, 3D GCSN
Sun et al. [69]	2022	IBSR18, BraTS2015	FSMPN

segmentation of cellular structures and tissue regions is necessary for identifying abnormalities such as cancerous lesions.

Another key advantage of U-Net is its ability to perform well with limited annotated medical data. Unlike traditional deep learning models that require extensive labeled datasets, U-Net is highly data-efficient and benefits from techniques like data augmentation and transfer learning. Additionally, its fully convolutional nature allows it to handle variable-sized input images, making it adaptable to different histology datasets. The combination of robust feature extraction and fine-detail preservation ensures that U-Net provides reliable and interpretable segmentation results, aiding dermatopathologists in diagnosing and analyzing skin diseases with high accuracy [28, 47]. Algorithm 1 summarized the steps considered by U-Net architecture for medical image segmentation.

Algorithm 1 U-Net for Medical Image Segmentation

```
Input: Preprocessed medical image $I$
Output: Segmentation mask $S$
Initialize U-Net with encoder-decoder structure
for each epoch $e$ in $E$ do
    for each batch $B$ in training data do
        Apply convolution + ReLU in encoder layers
        Downsample using max-pooling
        Apply convolution + ReLU in decoder layers
        Upsample using transposed convolution
        Add skip connections
        Compute Dice Loss
        Backpropagate and update weights
    end for
end for
return $S$
```

3.1.2 Transformer-based segmentation (Swin-UNet)

Transformer-based segmentation models like Swin-UNet have shown significant promise in medical image analysis, particularly for skin histology images. Unlike traditional CNN-based models like U-Net, Swin-UNet utilizes the self-attention mechanism of Transformers, allowing it to capture long-range dependencies within an image [34, 38]. This is especially beneficial for skin histology, where tissue structures and cellular formations have complex spatial relationships. The hierarchical design of the Swin Transformer preserves both global and local contexts, enabling more precise segmentation of intricate histological patterns. Additionally, the patch-based processing in Swin-UNet improves computational efficiency while maintaining high-resolution feature representation, which is critical for identifying fine-grained details in histopathological images.

3.1.3 Mathematical justification of transformer superiority

While CNN-based architectures have been widely used in medical image segmentation, their reliance on fixed-size convolutional kernels limits their ability to capture long-range dependencies efficiently. CNNs learn spatial features through convolutional operations that focus on local neighborhoods. Each layer captures features within a specific receptive field, and deeper layers aggregate information from a larger area. However, this hierarchical feature extraction process results in a loss of fine-grained details in deeper layers and restricts the model's ability to recognize distant dependencies. The computational complexity of a CNN layer for an image of size $H \times W$ with a kernel of size $k \times k$ is given by:

$$\mathcal{O}(HWk^2C^2) \tag{1}$$

where C represents the number of channels. The dependency on small receptive fields necessitates deeper networks to capture long-range dependencies, increasing the number of parameters and computational cost.

In contrast, Transformers process images using the self-attention mechanism, which allows every pixel to interact with all others in a given region. The self-attention operation is mathematically defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

where Q, K, V are the query, key, and value matrices, respectively, and d_k is the dimensionality of the key. Unlike CNNs, which aggregate information through stacked convolutional layers, self-attention allows each pixel to directly relate to any other pixel in the image. However, the computational complexity of full self-attention in a naïve Transformer scales as:

$$\mathcal{O}(HWC^2) + \mathcal{O}(H^2W^2C) \tag{3}$$

which becomes prohibitively expensive for high-resolution medical images.

To mitigate this, Swin-UNet introduces shifted window attention, an efficient mechanism that partitions an image into non-overlapping windows, applying self-attention locally within each window. This method reduces computational overhead while still allowing long-range interactions by shifting window positions across layers. The complexity of Swin-UNet's windowed self-attention is given by:

$$\mathcal{O}(HWC^2) + \mathcal{O}(M^2C) \tag{4}$$

where M is the window size, significantly lowering computation compared to full self-attention. By preserving hierarchical structures and long-range dependencies while reducing computational burden, Swin-UNet provides an optimal balance between efficiency and contextual awareness.

Overall, Swin-UNet's ability to model long-range dependencies, integrate multi-scale contextual information, and dynamically adapt to image variations makes it highly suitable for histopathological image segmentation. It captures both local tissue structures and broader histological patterns, making it more robust to staining variations and imaging inconsistencies compared to CNN-based models.

3.1.4 Algorithm implementation

Algorithm 2 presents the steps for implementing a transformer-based segmentation model for medical image analysis.

Algorithm 2 Swin-UNet for Medical Image Segmentation

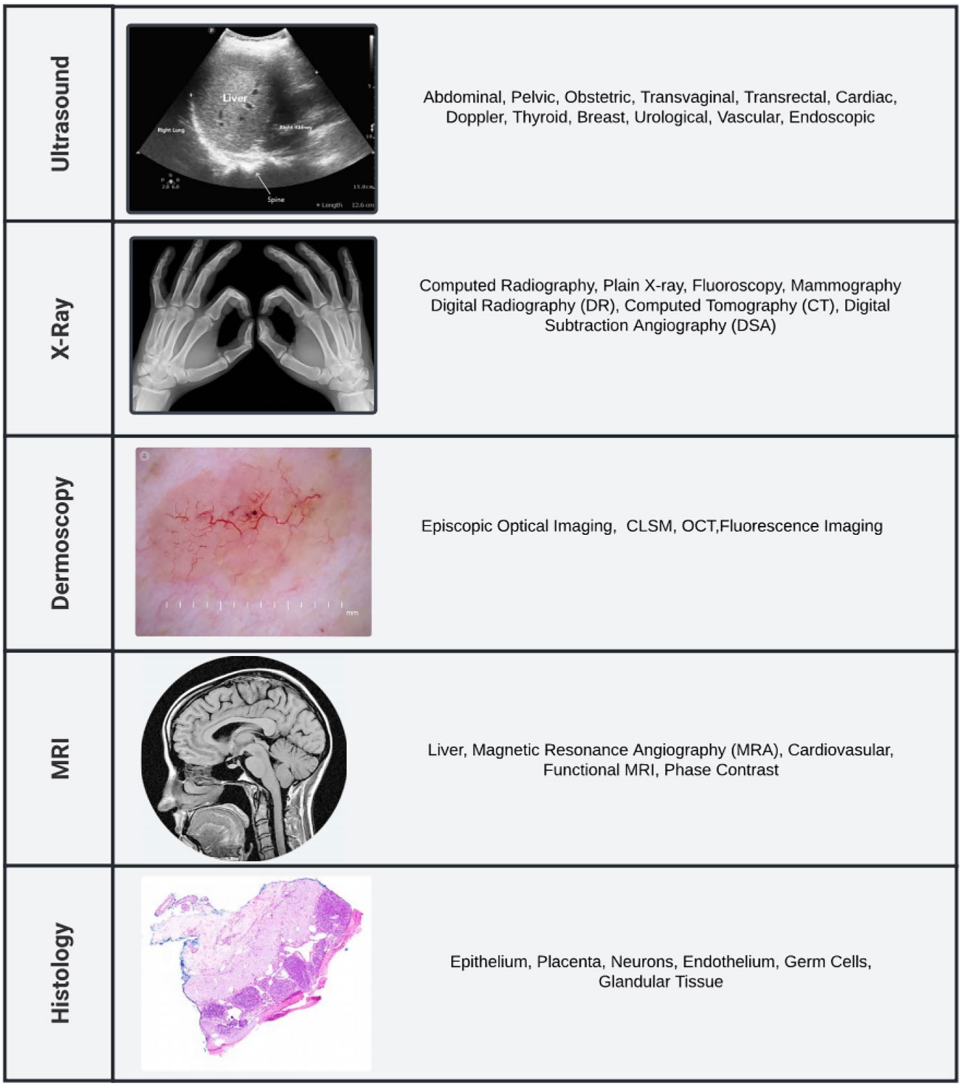
```
Require: Preprocessed medical image $I$
Ensure: Segmentation mask $S$
Partition the image into non-overlapping patches
Apply shifted window attention mechanism
for each epoch $e$ in $E$ do
    for each batch $B$ in training data do
        Pass through transformer encoder layers
        Apply hierarchical self-attention
        Upsample in decoder with attention-based skip connections
        Compute Dice Loss
        Backpropagate and update weights
    end for
end for
return $S$
```

4 Existing datasets in medical imaging modalities

There are various methods discussed and used in different structures based on a given medical condition, causing the design of customized algorithms. The output of these algorithms utilizes different kinds of dataset. A summary of the existing work discussing the algorithms for different image modalities is presented in Table 1. The medical image modalities we considered include, radiology, ultrasound, dermatology, and histology. Figure 5 visualizes the internal structures of the focused body parts that are facilitating disease monitoring.

4.1 Radiology images

The radiology images are used to diagnose the disease and assist in patient treatment by capturing the affected part of the body, helping to identify illnesses [70], structural damage, or abnormalities. There are several types of images such as X-ray [66], magnetic resonance imaging (MRI) [69], computer tomography (CT) [54], and mammography [71]. These images can assist in diagnosing cancer effectively. The cancer cells can develop in any part of the body, but affect the lungs, brain, liver, stomach, breasts, colon, prostate, rectum, and skin. Detecting cancer at an early stage increases the chances of patient survival [72]. The clinicians have several modalities available for diagnosis, including physical examinations, biopsies, imaging techniques, and lab tests.



Ultrasound	Abdominal, Pelvic, Obstetric, Transvaginal, Transrectal, Cardiac, Doppler, Thyroid, Breast, Urological, Vascular, Endoscopic
X-Ray	Computed Radiography, Plain X-ray, Fluoroscopy, Mammography Digital Radiography (DR), Computed Tomography (CT), Digital Subtraction Angiography (DSA)
Dermoscopy	Episcopic Optical Imaging, CLSM, OCT,Fluorescence Imaging
MRI	Liver, Magnetic Resonance Angiography (MRA), Cardiovasular, Functional MRI, Phase Contrast
Histology	Epithelium, Placenta, Neurons, Endothelium, Germ Cells, Glandular Tissue

Fig. 5 Modality of medical types

Table 2 summarizes the studies presented in radiology. Subramanian et al. [33] utilized imaging techniques such as CT scans and MRI for detecting cancer cells. They performed experiments using pre-trained CNN variants like MobileNetV3, VGGNet (Visual Geometry Group), and DenseNet (Densely Connected Networks) on ImageNet dataset which is divided into two tasks. In the first task, fine-tuning is applied, while the second task manipulates the model's multitasking capability. As a result, the model maintains the knowledge gained from Task 1 and applies its learning to Task 2 to generate better predictions. The approach incorporates the incremental learning concept, which allows the model to retain previous knowledge while acquiring new information. They best accuracy was yielded on MobileNetV3. Seif et al. [54] also proposed a Deep Learning model that employs an additional attention-based UNet mechanism to enhance performance on CT images. Gite et al. [66] experimented with various Deep Learning architectures to evaluate chest X-ray images and reported higher accuracy to determine good segmentation results, the details are summarized in Table 2. Sun et al. [67] introduced semi-supervised Deep Learning methods. The performance was measured by multiscale graph cut loss function and achieving highly efficient results on neuro-imaging. Pang et al. [68] proposed a two-stage multi-class architecture for MRI images to improve the results in spine parsing as explained in their paper. Sun et al. [69] proposed the Feature Space Message Passing Network (FSMPN) framework for capturing long-range dependencies. To compare it with the Deep Learning traditional models, FSMPN achieved the outperforming results on Brain MRI images.

4.2 Ultrasound images

Semantic segmentation of medical images is crucial for diagnosing and treating diseases. The primary goal of ultrasound images is to identify abnormalities that may have a significant impact on fetal growth [17]. Despite advances in ultrasound technology, specifically recognizing anomalies in prenatal imaging remains challenging, requiring significant expertise from physicians [73]. Relevant studies utilizing Deep Learning models for disease diagnosis using ultrasound are summarized in Table 3. Geng et al. [51] proposed a novel method, a structured multi-scale residual fusion network for better semantic segmentation on UV images. In this paper, it is presented that Symmetric Multi-Task Network (SMNet) yielded significant performance on a small dataset as compared to the state-of-the-art techniques. Xu et al. [52] proposed a CNN Deep Learning-based semantic segmentation approach for 3D breast UV images into the four main tissues, indicating that automated systems can significantly reduce the time required for analysis and provide efficient results. Similarly, Vakanski et al. [53] also focused on breast UV images. To improve the precision and reliability of breast tumor segmentation in UV images, they presented to incorporate visual saliency into a DL-based model (U-Net), yielded improved performance on segmentation method.

4.3 Dermoscopy images

Dermoscopy is another application of medical images, that is used for the examination of skin lesions [74]. Dermoscopy is an important tool in dermatology that diagnoses and enhances the accuracy of skin diseases [75] and also leads to enhanced patient care and early detection of skin cancer. Melanoma skin cancer [41] is one of the most dangerous skin diseases, requiring precise segmentation of skin lesions in dermoscopy images for accurate diagnosis and treatment [76]. Recently, automated transformer-based methods for skin lesion segmentation have been employed, achieving higher accuracy in segmentation tasks [77]. The relevant studies are illustrated in Table 4. Chen et al. [59] proposed the global-to-local (G2LL) self-supervised learning approach for the transformer-based segmentation on skin lesions and got outstanding results on a publicly available dataset of dermoscopy. Also, the proposed approach gives outstanding results as compared to the state-of-the-art self-supervised approach. Kaur et al. [60] proposed the fully connected neural network (FCNN) Deep Learning model for the lesion segmentation dermoscopic skin cancer images. The proposed methods provide higher accuracy in segmentation and are used in clinical settings to better understand the nature of cancerous lesions. Fedorenko et al. [35] proposed the U-Net for the semantic segmentation of the dermoscopic images of pigment skin lesions and achieved high accuracy. Also, it improves the automated classification by efficiently handling the variation in skin color. Iranpoor et al. [61] also proposed the U-Net model on dermoscopy skin images, improving the performance of the pre-trained encoder and modifications of the pooling layer. This proposed method achieves high accuracy in training and testing data and also enhances the efficiency of the existing CNN model.

Table 2 Radiology in medical image

Data modality	Target detection	Dataset size	Description	Obtained result	Limitation
CT & MRI [33]	Comprehensive cancer detection	5000 images of each type of cancer	Pre-trained models used to detect Cancer disease	MobileNet (without multitasking): 74.05%, MobileNet (with multitasking): 79.95%	Pre-trained models might be less precise if their weights are not properly fine-tuned. Also, it relies on specific imaging data
CT [54]	Liver lesions on abdomen CT images	201 CT scans	Deep Learning for semantic segmentation of lesions in the liver, also evaluates the impact of attention modules on segmentation performance	DSC (liver): 0.934%, DSC (tumor): 0.778%	Utilization of attention mechanisms didn't guarantee improved performance on their own
X-Ray [66]	Tuberculosis (TB) disease	Montgomery 138 images, Shenzhen Hospital 662 images	Comparison of different Deep Learning models (U-Net, FCN, U-Net++, SegNet)	U-Net++: 98%	Limited size of data for TB disease
MRI [67]	Cerebrospinal Fluid (CSF)	IBSR18 has 18 volumes, BrainWeb20 has 20 volumes	Semi-supervised system for learning using a multi-scale graph cut loss function	Proposed models: approximately 96% on BrainWeb20 dataset	Limited labeled data
MRI [68]	Spine parsing (vertebrae and IVDs)	215 patients (T2-weighted MR images)	Using two-stage architecture (3D GCSN + 2D ResUNet)	DSC vertebrae: 87.32%, DSC IVDs: 87.78%, Overall Accuracy: 87.49%	Manual distinction may add variability, subject to expert interpretations
MRI [69]	Brain tissue	IBSR18 has 18 volumes, BraTS2015 has 274 subjects	Used FSMPN to compare with existing segmentation results	DSC: over 86%	Limited generalization; performance may vary between datasets

Table 3 Ultrasound in medical image

Data modality	Target detection	Dataset size	Description	Obtained result	Limitation
Fatel UV [51]	Heart segmentation with lung as a reference	105 UV images	SMNet enhances segmentation accuracy via structured information	SMNet outperforms MSRF about of dice, IOU, precision, and recall	Limited dataset size
BUI [52]	Breast tissue segmentation	20,000 images derived from 250 slices	Automated segmentation using CNNs boosts tissue classification	CNN: over 80%	Manual segmentation is dependent on proficient radiologists; might still need validation
BUS [53]	Breast Tumor segmentation	510 BIU images	Adding visual saliency to a U-Net model for better segmentation	DSC: 90.5%	Rely on the quality of prominence maps; low-quality mappings can reduce performance.

Table 4 Dermoscopy in medical image

Data modality	Target detection	Dataset size	Description	Obtained result	Limitation
Dermoscopy Imaging [59]	Segmentation of Skin lesion	ISIC(2019) has 25331 images and ISIC(2017) has 2750 images	Used self-supervised Learning G2LL approach for transformer-based segmentation models	Achieved more than 80%	Limited data size that affects on model training, also for future use of the semi-supervised learning approach.
Dermoscopic [60]	Melanoma & neoplasm detection	40 melanoma images, 160 nevus images	Used the PH2 dataset for evaluation and comparison	Achieved Accuracy: 0.963%	Limited dataset might influence interpretation.
Dermoscopic Imaging [35]	Pigmented skin lesions	ISIC(2018) has 2694 images	Used U-Net a convolutions neural network for categorizing pigmented lesions	Attained accuracy: 93.32%	The system operates as another tool, not as a standalone diagnostic tool.
Dermoscopic Imaging [61]	Skin lesion image segmentation	PH2 database has 200 images (80 for common nevi, 80 atypical nevi & 40 melanomas)	Used enhanced U-Net structures with the pre-trained encoding	Attained precision: 89%	Not Specified

4.4 Histopathology images

Another primary application of medical images is histopathology, particularly within the field of pathology. This involves examining tissue samples under a microscope [78] to understand and diagnose diseases. It provides essential information for diagnosing various conditions, including cancer [79] and infectious diseases. There are few notable works presented in recent years [9, 50, 56, 58, 62, 63]. In histology images, researchers study the microscopic structure of tissue samples, examining various organs and tissue types, such as the liver, lungs, and kidneys. Some researchers have applied automated deep-learning models to histology images. There are a few notable works presented in recent years as summarized in Table 5. The visualization of skin images is possible in high-resolution images when applying Deep Learning approaches. The network can learn, interpret, and represent meaningful insights that help to better understand the model's performance.

Li et al. [55] proposed an automated deep-learning method for detecting cell nuclei, achieving efficient results on tissue slices. The pathological tissue slices are crucial for determining the extent of disease progression and informing treatment decisions when healthcare providers diagnose complex conditions. Since manual processing of pathological areas is time-consuming and subjective, researchers need to utilize computer-aided diagnosis as an intelligent support tool. Han et al. [56] proposed a model that employs whole-slide histology images of breast cancer and lung adenocarcinoma, utilizing patch-level classification and a weakly semi-supervised semantic segmentation approach for these tissue slices. Similarly, Gu et al. [57] is focused on histopathology images to identify morphological features of glands, such as their size and contour by using a weakly supervised learning technique. They proposed histopathology segmentation with transformer (HistoSegRest) approach and achieved the correct segmentation on the gland region that uses the weak supervision with image label histopathology data. For long-range dependencies, it uses transfer-based self-attention and the GlaS dataset (Gland Segmentation in Colon Histology Images). They reported improved segmentation accuracy. Fang et al. [58] proposed to use the weakly supervised semantic segmentation on the histopathology images. They used the PistoSeg model to segment the histology images. Yildiz et al. [62] proposed to use the Deep Learning-based segmentation model in colon histology images using the U-Net model for medical image segmentation. They perform multi-class semantic segmentation on the provided dataset and gets enhanced accuracy using the semantic segmentation approach. Li et al. [64] aim is to develop a fast and accurate model that detects and segments the regions of breast cancer in WSI (whole-slide images) using the Camelyon16 dataset. They performed preprocessing by using MobileNetV2 and ResNet101 and for learning, they predicted the correct segmentation by the U-Net model. Wazir et al. [48] proposed a neural network architecture for multiscale objects that improves segmentation accuracy on the histology images using an encoder and decoder architecture with quick attention and multi-loss function. Additionally, Rasool et al. [65] proposed the semantic segmentation on the oral cell carcinoma tissues using the 18 WSI to correctly identify and segment the micro-vessels and nerves in histology images. They reported encouraging results on the efficiency of cancer prognosis and diagnosis. Turhan et al. [34] also employed the Deep Learning-based semantic segmentation approach on histology images. Some of these histology images specifically focus on skin histology, where skin tissue samples are analyzed to study disorders and diagnose conditions and other dermatological diseases. Xu et al. [63] proposed and analyzed the unsupervised method, of tissue clustering based on morphological traits. These traits are used for the segmentation of histological images such as tumor or non-tumor regions. Their proposed technique has been assessed using datasets related to brain and skin histology, demonstrating performance comparable to U-Net models. Thomas et al. [49] applied an interpretable Deep Learning U-Net model for skin histology images of 12 dermatological classes using the Queensland University dataset and reported high results. Similarly, Asaf et al. [50] proposed deep semantic segmentation methods to enhance the accuracy of segmentation, particularly in detection of the skin cancer. They used the transformer-based semantic segmentation and compared the results with the traditional model.

5 Evaluation measures

Evaluation metrics are statistical tools used to assess the performance of Deep Learning models and data analysis algorithms. They play a crucial role in measuring the accuracy, precision, and overall effectiveness of algorithms, particularly in medical image segmentation tasks. A confusion matrix is commonly employed to evaluate a classification

Table 5 Histopathology in medical image

Data modality	Target detection	Dataset size	Description	Obtained result	Limitation
The pathological Tissue Slices [55]	Semantic Segmentation on Cells Nuclei	670 images	Modified U-Net with channel attention system	U-Net: 93.06%	Highly precise needs for cell nuclei segmentation
Histology [56]	Tissue level segmentation	54 patients data, LUAD-HistoSeg has 16678 patches and BCSS-WSSS has 31826 patches	Two-step models with complementary and alternative medicine for pseudo masks	Overall Accuracy: 0.81%	Needs additional specimens for accuracy.
Histology [58]	Tissue Segmentation	WSSS4LUAD: 10091 images and BCSS-WSSS: 23422 images	The PistoSeg framework fixes boundary mistakes in CAM-based methods	Exceeding modern methods in efficiency	Future work to develop accurate synthetic datasets will be needed.
Histology [62]	Nuclei segmentation	4981 images	Using Multi-class semantically segmented with the U-Net architecture.	Overall Accuracy: 69.61%	Further study needs instance-based segmentation.
Histology WSI images[64]	Segmentation of breast cancer and Tumor detection	378 WSIs of sentinel lymph nodes	Combine segmentation and classification methods	ResNet101: 98.3%, MobileNetV2: 97.2%	Not Specified.
Histology [48]	Glands, nuclei, and biomarkers segmentation	44 MoNuSeg samples and 165 GlaS images	An encoder-decoder structure with quick attention module.	Proposed network in MoNuSeg: 1.99% increase & in GlaS: 7.15% increase	Not specified.
Histology [65]	Micro-vessels and nerves segmentation	18 WSIs of oral cell carcinoma	Utilizing ResNeXt blocks in a hybrid segmentation semantic architecture	Outperformed the state-of-the-art network.	Not specified
Histology [34]	Nuclie Segmentation and classification	481 Visual fields	Comparison of Deep Learning models.	F1/DSC: 80.2%	Imbalance dataset
Histology and Skin Histology [63]	Melanoma detection	80 lung images, 35 brain images, 30 skin images	Using the tiscut for the unsupervised segmentation	U-Net (brain): 70%, U-net(skin): 85%	Efficiency of annotating manually for ground truth has not been assessed for other tissue types.
Histology and Skin Histology [49]	Non-melanoma skin cancer detection	290 H & E stained images	Using the Interpretable Deep Learning semantic segmentation strategies	Overall accuracy: over 97%	Needs the extensively labeled data.

model's performance. It outlines the relationships between true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). True positives represent correctly identified outcomes, while false positives refer to instances incorrectly labeled as positive. Various evaluation metrics have been utilized in recent medical image segmentation studies. Each provides unique insights into model performance. The details of these metrics are summarized in Table 6.

5.1 Dice similarity coefficient

The Dice Similarity Coefficient (DSC) is a widely used metric in image segmentation, particularly in medical imaging, to quantify the overlap between a predicted segmentation and the ground truth. It is defined as the ratio of twice the intersection of the two sets to the sum of their sizes. Mathematically, it is expressed as in Eq. (1).

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \tag{5}$$

where *X* represents the predicted region and *Y* represents the actual region. The DSC ranges from 0 to 1, where 1 indicates a perfect match and 0 means no overlap. This metric is particularly useful for imbalanced datasets, as it penalizes both false positives and false negatives while emphasizing true positives. Compared to the Jaccard Index (IoU), the Dice score gives more weight to overlapping regions, making it a preferred choice for evaluating segmentation models in medical imaging applications such as tumor detection and organ segmentation.

5.2 Mean intersection over union

The Mean Intersection over Union (mIoU) is another metric in image segmentation to evaluate the performance of a model by measuring the overlap between predicted and ground truth segmentations. It is computed as the average Intersection over Union (IoU) across all classes in a multi-class segmentation task. Mathematically, IoU is defined as the ratio of the intersection of the predicted and actual regions to their union as expressed in Eq. (2):

$$mIoU = \frac{|X \cap Y|}{|X \cup Y|} \tag{6}$$

where *X* represents the predicted segmentation and *Y* is the ground truth. The mean IoU (mIoU) is then calculated by averaging the IoU values across all classes. The metric ranges from 0 to 1, where 1 indicates perfect segmentation and 0 means no overlap. Unlike simpler accuracy metrics, mIoU accounts for false positives and false negatives, making it a robust evaluation measure for semantic segmentation tasks in applications like medical imaging, autonomous driving, and satellite image analysis.

5.3 Precision and recall

Precision and recall are two key metrics used in image segmentation to evaluate a model's performance in correctly identifying segmented regions. Precision measures the accuracy of the positive predictions by calculating the proportion of correctly segmented pixels (true positives) out of all pixels predicted as positive (true positives + false positives). It is given by:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

where *TP* (True Positives) are correctly segmented pixels, and *FP* (False Positives) are incorrectly segmented pixels. High precision means fewer false positives.

On the other hand, recall (also called sensitivity) measures how well the model captures all the actual segmented regions by computing the proportion of correctly segmented pixels out of all ground truth positives (true positives + false negatives). It is given by:

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

Table 6 Evaluation metrics

Metric	Studies	Calculation method	Interpretation
Dice Similarity Coefficient	[35, 51, 53–55, 59, 60, 62, 63, 66–68]	$\frac{2\lvert X\cap Y\rvert}{\lvert X\rvert+\lvert Y\rvert}$	A Dice score of 0 indicates that the predicted segmentation has no overlap with the ground truth, implying no resemblance. A Dice score of 1 indicates that the predicted segmentation completely matches the ground truth, indicating overall similarity.
Mean IOU	[50, 51, 56, 58, 59, 61, 62, 66]	$\frac{\lvert X\cap Y\rvert}{\lvert X\cup Y\rvert}$	The mIoU score shows how well the predicted segmentation matches the ground truth. The High mIoU indicates the more accurate result and the low value shows the poor prediction, between the prediction and the ground truth.
Precision	[33, 51, 52, 61, 62, 65, 66]	$\frac{TP}{TP+FP}$	Measure the precision on the 0 or 1 scale, the higher value of precision means the model returns more relevant results than the irrelevant ones.
Recall	[33, 51, 52, 61, 62, 65, 66]	$\frac{TP}{TP+FN}$	It represents the amount of actual positive instances in which the model can accurately comprehend.
Sensitivity	[54, 65, 66]	$\frac{TP}{TP+FN}$	It estimates the proportion of an actual positive case that the algorithms correctly identifies.
Jaccard Index	[35, 53–55, 63, 65]	$\frac{\lvert X\cap Y\rvert}{\lvert X\cup Y\rvert}$	Also known as mIoU. It identifies which elements of the two sets are distinctive and which are shared by comparing them.
F1-Score	[33, 34, 48, 50, 52, 56–58, 61, 65]	$2\cdot\frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}$	F1 Score = 0: the smallest possible result suggesting that there are no true positives identified by the model. F1 Score = 1: Excellent recall and precision.
Cohen's Kappa	[34]	$\kappa=\frac{p_o-p_e}{1-p_e}$	Kappa value=1 is the perfect agreement. Kappa value=0 means the agreement by chance or the below 0 means the worse than chance.

where *FN* (False Negatives) are pixels that should have been segmented but were missed. High recall means fewer false negatives.

In image segmentation, there is often a trade-off between precision and recall. A high precision but low recall means the model is conservative, predicting only highly certain segmentations, while a high recall but low precision means the model is over-segmenting, capturing more than necessary. To balance both, F1-score is often used as a harmonic mean of precision and recall.

As reported in other papers [50, 58, 59, 62, 66] Mean Intersection over Union (MIoU) is a widely used metric that measures the average between the actual and predicted segments. Furthermore, some work used Jaccard Index measure to measure the similarity between the predicted and actual segmentation.

5.4 Computational trade-offs between CNNs and transformers

While transformers such as Swin-UNet offer superior segmentation accuracy by leveraging self-attention mechanisms, they come with increased computational demands. Compared to CNN-based architectures like U-Net, transformer-based models require significantly higher memory due to the quadratic complexity of self-attention operations. This can lead to increased inference time and greater GPU memory consumption, particularly when handling high-resolution medical images.

Conversely, CNN-based architectures exhibit lower computational costs due to their localized convolutional operations, making them more suitable for real-time applications and resource-constrained environments. However, their limited receptive field can hinder their ability to capture long-range dependencies, which are crucial for segmenting complex histopathological structures. Transformers mitigate this limitation by dynamically modeling global and local context, albeit at the expense of higher computational overhead.

Despite these trade-offs, recent advancements such as hierarchical transformers and efficient self-attention mechanisms have reduced computational complexity while retaining the benefits of long-range feature extraction. As a result, the choice between CNNs and transformers depends on the specific use case, balancing accuracy, computational efficiency, and model interpretability in medical image analysis.

6 Explainable AI (XAI) in medical image segmentation

While deep learning models have demonstrated remarkable performance in medical image segmentation, their black-box nature poses challenges for clinical adoption. Explainable AI (XAI) techniques aim to address this issue by providing insights into model decisions, enabling medical professionals to trust and interpret predictions effectively.

6.1 Saliency maps and Grad-CAM for model interpretability

Saliency maps and Gradient-weighted Class Activation Mapping (Grad-CAM) are widely used techniques to visualize the regions that influence model predictions. Saliency maps highlight important areas in an image by computing the gradient of the output with respect to the input pixels. This helps identify which regions contribute most to the segmentation outcome. Grad-CAM, on the other hand, generates heatmaps over feature maps, offering a class-specific localization of relevant structures. These methods allow pathologists to validate whether the model is focusing on clinically significant regions, such as tumor boundaries or inflamed tissue.

6.2 Enhancing clinical decision-making with XAI

The integration of XAI in medical image segmentation enhances the interpretability of automated systems, reducing ambiguity in decision-making. By providing visual explanations, models can assist clinicians in verifying segmentation results and understanding misclassifications. This is particularly crucial in high-stakes scenarios such as skin cancer diagnosis, where incorrect segmentation could lead to misdiagnosis.

6.3 Future directions for XAI in medical segmentation

Future advancements in XAI could focus on incorporating model-agnostic interpretability frameworks to improve trust and transparency in deep learning-based segmentation. Techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) could be explored to quantify the contribution of individual features in segmentation outcomes. Additionally, integrating attention-based visualization methods within transformer architectures could further enhance explainability by highlighting long-range dependencies in histopathological images.

By leveraging XAI, deep learning models for medical image segmentation can transition from black-box predictors to clinically relevant tools that provide both accurate and interpretable results, fostering greater acceptance among medical professionals.

7 Research gap

In this review paper, numerous works have been analyzed explaining the amount of work performed in medical image segmentation. The Deep Learning approaches were discussed with their potential role in medical disease diagnosis. There are numerous research gaps identified and discussed.

1. Most of the recent research is focused on specific imaging modalities like CT scans, MRI, and X-rays, with limited exploration of multi-modal techniques. These multi-model techniques could potentially enhance diagnostic precision by incorporating information from multiple sources.
2. Another potential constraint is the availability of the dataset which hinders the researcher's focus on certain cancer types while neglecting other critical diseases such as Chronic obstructive pulmonary disease (COPD) and pneumonia.
3. There is also an abundance of research focused on explainable AI (XAI) approaches, which are vital to establishing confidence as well as transparency in healthcare settings. But there is a need to develop XAI model that can assist in medical image diagnosis.
4. Many Deep Learning models are validated utilizing controlled datasets instead of real-world clinical validation, creating issues regarding their generalizability and utility across different patient categories. It is important to address these flaws for proposing reliable Deep Learning-based medical diagnostics that are trustworthy interpretable, and accessible at a same time in a variety of clinical settings.
5. We identified that, there is a lack of studies on the potential benefits of multi-task learning techniques, which could improve a segmentation model's sensitivity to various types of ultrasound images.
6. Radiologists aren't highly trained in using AI-based model training, which inhibits their ability to use crucial medical knowledge in the application domain. In addition, there aren't many tailored loss functions that might be used to encode clinical data and enhance the model's reliability and performance.

Addressing these gaps can lead to devise strategy for ultrasound algorithms for image segmentation which could be more effective and useful in clinical settings. Several significant gaps also exist in the segmentation of skin lesions for dermoscopy images. The prominent key challenge is to deal with numerous approaches that are evaluated on small datasets, restricting their capacity to perform effectively in real-world scenarios. Furthermore, many approaches fail to properly address problems such as hair interference and skin image analysis. The available techniques are often used as supplementary tools rather than fully integrated into diagnostic systems, emphasizing the need for more practical and reliable options for clinical applications.

In this paper, skin histology images are considered as a potential case study in detecting skin diseases. While recent studies have highlighted the potential benefits of Deep Learning models, many continue to struggle with class imbalance, leading to lower accuracy for those with fewer skin lesions as explained in Sect. 4. The skin cancer is explained through multiple classes as represented in Fig. 6.

By assessing the recently proposed work, it is evident that there is an absence of research into model interpretation that may enhance diagnosis transparency, particularly for complex cases. Future studies should address these limitations by improving model interpretability which will eventually improve the class balance using elegant

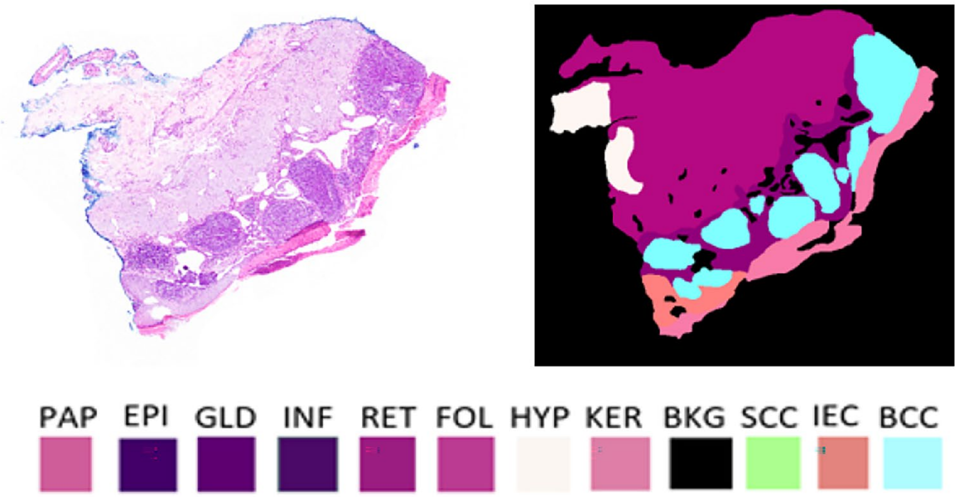


Fig. 6 Multi-class skin cancer

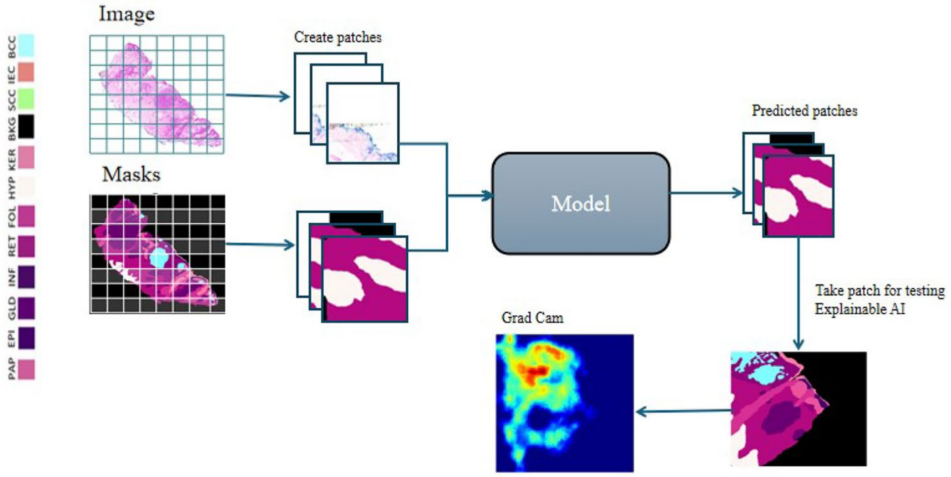


Fig. 7 Proposed transformer architecture and explainable AI

augmentation methods. Recent studies have used transformer-based models to analyze histology images, but this option might lack in generating explanation to demonstrate how well the model performs. We have established a hypothesis that, if histology images were given to specialized transformer-based architecture then it could improve the diagnosis accuracy. The hypothesis was evaluated and produced encouraging results if we could present the results with correct interpretation through an explainable AI (XAI) model as depicted in Fig. 7.

Furthermore, one of the highlighted problems is transfer learning, which uses a large database to train models for more specific tasks. However, transfer learning often does not evaluate the model's performance on previous tasks. As the models might lose information from prior learning while adjusting to new ones. This constraint can lead to reduced performance on previous as the algorithm focuses on new categories. To address this problem, we need to look for solutions that consider the prior expertise while learning from new data.

8 Implementation and empirical analysis

To address the identified research gaps, we implemented an experimental comparison between CNN-based and Transformer-based architectures for skin lesion and histopathology image segmentation. The objective is to evaluate the effectiveness of attention mechanisms and transformer models in improving segmentation accuracy over conventional CNN models.

8.1 Dataset

The experiments were conducted on three publicly available datasets. The ISIC 2018/2019 dataset was used for skin lesion segmentation in dermoscopy images, providing a comprehensive collection of annotated skin images for melanoma detection. The GlaS dataset focuses on histopathology images for gland segmentation, widely used for evaluating segmentation algorithms in colorectal cancer diagnosis. Additionally, the Camelyon16 dataset, consisting of whole-slide histopathology images, was utilized for tumor detection tasks, offering large-scale data for evaluating segmentation performance on complex tissue structures.

All datasets used in this study are publicly accessible, ensuring reproducibility and transparency in experimental evaluations. Ethical approvals of each of the abovementioned data repositories have been stated in the dataset link. However, dataset biases must be acknowledged, as the distribution of images may not fully represent all demographic groups, skin tones, or histopathological variations. Such biases could potentially affect model generalization and fairness in clinical applications. Future work should consider curating diverse datasets to mitigate these limitations and enhance model robustness across different populations.

8.2 Experimental setup

For preprocessing, all images were resized to 256 × 256 pixels, normalized, and subjected to data augmentation techniques such as rotation, flipping, and contrast adjustments to enhance model generalization. The experimental setup involved two distinct architectures: a CNN-based model using U-Net with standard convolutional blocks, and a Transformer-based model employing Swin-UNet, which integrates self-attention mechanisms into the U-Net framework for enhanced feature extraction.

The models were trained using the Adam optimizer with a learning rate of 0.001. A batch size of 16 was used, and training was conducted over 50 epochs. The Dice Loss function was employed as the primary loss metric, focusing on optimizing overlap between the predicted and ground truth segmentations.

8.3 Results

Table 7 presents a comparative analysis of segmentation performance between the CNN-based and Transformer-based models, evaluated using Dice Similarity Coefficient (DSC), Mean Intersection over Union (mIoU), and F1-Score.

8.4 Discussion

Despite significant advancements in Deep Learning for skin histology image analysis, several challenges remain, including limited annotated data, high computational costs, and generalizability issues of models. To address the issue of data scarcity, future research should explore self-supervised learning (SSL) and synthetic data generation. SSL techniques, such as contrastive learning, can enable models to learn robust feature representations from unlabeled data before fine-tuning on limited annotated samples. Additionally, generative adversarial networks (GANs) and diffusion models can be employed to create realistic synthetic histopathology images, augmenting training datasets and improving model robustness. Using federated learning can also allow decentralized institutions to train models collaboratively while preserving patient privacy.

The high computational cost associated with transformer-based models compared to CNNs is another major bottleneck. Future research should focus on efficient transformer architectures, such as lightweight Vision Transformers (ViTs) and hybrid CNN-Transformer models, which integrate convolutional layers for low-level feature extraction with transformer layers for high-level spatial understanding. Model pruning, knowledge distillation, and quantization techniques can further optimize Deep Learning architectures, enabling their deployment in resource-constrained clinical environments. In addition, exploring edge computing and cloud-based AI inference systems can help distribute computational workloads efficiently, reducing the dependency on high-end hardware.

Table 7 Comparison of CNN and Transformer-based models

Model	DSC	mIoU	F1-score
U-Net (CNN)	0.85	0.78	0.82
Swin-UNet (Transformer)	0.91	0.85	0.88

Generalizability remains a critical challenge, as Deep Learning models often struggle with variations in staining techniques, imaging modalities, and patient demographics. To improve the robustness of the model, future research should emphasize domain adaptation and short-shot learning. Domain adaptation techniques, such as adversarial training and style transfer, can help models learn invariant features across different datasets, enhancing their ability to generalize. Few-shot learning approaches, including meta-learning and prototypical networks, can allow models to adapt quickly to new cases with minimal labeled examples. Furthermore, explainability methods such as attention visualization and saliency mapping should be integrated to enhance model interpretability, ensuring their adoption in clinical decision making. By addressing these challenges, Deep Learning models can become more reliable, interpretable, and deployable in real-world dermatopathology applications.

Our empirical analysis indicate that Swin-UNet achieves a 7% increase in DSC and a 9% improvement in mIoU over the traditional U-Net. The enhanced performance can be attributed to the attention mechanisms in transformers, which better capture long-range dependencies in medical images. However, transformer-based models have higher computational requirements, limiting their feasibility for real-time applications. This experimental study demonstrates the advantages of transformer-based models in medical image segmentation.

9 Conclusion

This study presents a comprehensive review and experimental analysis of Deep Learning techniques applied to medical image segmentation, with a particular focus on dermoscopy and histopathology images. The research highlights the comparative performance of CNN-based and Transformer-based architectures, addressing key gaps identified in the literature. The experimental results demonstrate that Transformer-based models, such as Swin-UNet, outperform traditional CNN-based architectures like U-Net in complex segmentation tasks. Specifically, the Swin-UNet achieved a 7% improvement in the Dice Similarity Coefficient (DSC) and a 9% increase in the Mean Intersection over Union (mIoU) compared to the CNN-based U-Net. These enhancements underscore the effectiveness of attention mechanisms in capturing long-range dependencies within medical images, leading to more accurate and precise segmentation.

While Transformer-based models offer notable improvements in segmentation performance, they also introduce higher computational complexity, which may limit their application in real-time clinical settings. Future research could explore hybrid models that combine the efficiency of CNNs with the representational power of transformers, aiming to balance accuracy with computational efficiency.

In conclusion, this study not only reinforces the potential of advanced Deep Learning models in medical image segmentation but also provides empirical evidence supporting the adoption of Transformer-based architectures for more accurate disease diagnosis. These findings contribute to the ongoing development of automated, reliable, and efficient diagnostic tools in medical imaging.

Acknowledgements This research was conducted in collaboration between Lakehead University, Thunder Bay, Canada and the National University of Sciences and Technology (NUST), Islamabad, Pakistan. The authors gratefully acknowledge the support of the MITACS Globalink Research Award, secured by Ms. Sana Fatima.

Author contributions The presented work is a part of the project which is collaboratively supervised by S.B.A from Lakehead University and M.U.A from NUST. We also extend our gratitude to S.M, Chair of the Computer Science Department at Lakehead University, for reviewing and providing valuable insights that enhanced the quality of this work. S.F implemented the concept.

Funding The authors appreciate the financial support provided by Lakehead University through the grant code #: 11-50-25082439-61101.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Fatma K, Benaissa I, Zitouni A, Zinne-eddine B. Assessing the performance of u-net in 3D medical image segmentation. In 2024 8th international conference on image and signal processing and their applications (ISPA). IEEE; 2024. pp. 1–6.
2. Qureshi I, et al. Medical image segmentation using deep semantic-based methods: a review of techniques, applications and emerging trends. Inf Fus. 2023;90:316–52.
3. Sarker I. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci. 2021;2(6):420.
4. Hao S, Zhou Y, Guo Y. A brief survey on semantic segmentation with deep learning. Neurocomputing. 2020;406:302–21.
5. Borba P, de Carvalho Diniz F, da Silva NC, de Souza Bias E. Building footprint extraction using deep learning semantic segmentation techniques: experiments and results. In 2021 IEEE international geoscience and remote sensing symposium IGARSS. IEEE; 2021. pp. 4708–4711.
6. Yu H, Xiao Z, Fang Z. A real-time semantic segmentation network with multi-path structure. In 2022 10th international conference on information systems and computing technology (ISCTech). IEEE; 2022. pp. 523–527.
7. Li Q, et al. Large-scale epidemiological analysis of common skin diseases to identify shared and unique comorbidities and demographic factors. Front Immunol. 2024;14:1309549.
8. Thomas SM, Lefevre JG, Baxter G, Hamilton NA. Non-melanoma skin cancer segmentation for histopathology dataset. Data Brief. 2021;39: 107587.
9. De A, Mishra N, Chang H-T. An approach to the dermatological classification of histopathological skin images using a hybridized cnn-densenet model. PeerJ Comput Sci. 2024;10: e1884.
10. Li B, Shi Y, Qi Z, Chen Z. A survey on semantic segmentation. In 2018 IEEE international conference on data mining workshops (ICDMW). IEEE; 2018. pp. 1233–1240.
11. Guo Y, Yang B. A survey of semantic segmentation methods in traffic scenarios. In 2022 international conference on machine learning, cloud computing and intelligent mining (MLCCIM). IEEE; 2022. pp. 452–457.
12. Zhao Z-Q, Zheng P, Xu S-T, Wu X. Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst. 2019;30:3212–32.
13. Thoma M. A survey of semantic segmentation. arXiv preprint arXiv:1602.06541 2016.
14. Hafiz AM, Bhat GM. A survey on instance segmentation: state of the art. Int J Multimed Inf Retriev. 2020;9:171–89.
15. Sharma R, Saqib M, Lin C-T, Blumenstein M. A survey on object instance segmentation. SN Comput Sci. 2022;3:499.
16. Kusumawardhana DB, Trilaksono BR, Abdurrohman H. Panoptic segmentation datasets for rail-based autonomous vehicle under mixed-traffic scenario. In 2022 12th international conference on system engineering and technology (ICSET). IEEE; 2022. pp. 19–24.
17. Guanlin D. Research on semantic segmentation algorithm based on deep learning control tools. In 2020 international conference on computer communication and network security (CCNS). 2020. pp. 35–38 https://doi.org/10.1109/CCNS50731.2020.00016.
18. Lateef F, Ruichek Y. Survey on semantic segmentation using deep learning techniques. Neurocomputing. 2019;338:321–48.
19. Siam M, et al. A comparative study of real-time semantic segmentation for autonomous driving. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018. pp. 587–597.
20. Asgari Taghanaki S, Abhishek K, Cohen JP, Cohen-Adad J, Hamarneh G. Deep semantic segmentation of natural and medical images: a review. Artif Intell Rev. 2021;54:137–78.
21. Trajanovski S, Shan C, Weijtmans PJ, de Koning S, Ruers TJ. Tongue tumor detection in hyperspectral images using deep learning semantic segmentation. IEEE Trans Biomed Eng. 2020;68:1330–40.
22. Küstner T, et al. Semantic organ segmentation in 3d whole-body MR images. In 2018 25th IEEE international conference on image processing (ICIP). IEEE; 2018. pp. 3498–3502.
23. Valizadeh A, Shariatee M. The progress of medical image semantic segmentation methods for application in covid-19 detection. Comput Intell Neurosci. 2021;2021:7265644.
24. Alowais SA, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ. 2023;23:689.
25. Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. Bioengineering (Basel). 2023;10(12):1435.
26. Armand TPT, Bhattacharjee S, Choi H-K, Kim H-C. Transformers effectiveness in medical image segmentation: a comparative analysis of unet-based architectures. In 2024 international conference on artificial intelligence in information and communication (ICAIIC). IEEE; 2024. pp. 238–242.
27. Wang R, et al. Medical image segmentation using deep learning: a survey. IET Image Proc. 2022;16:1243–67.
28. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: a nested u-net architecture for medical image segmentation. In Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer; 2018. pp. 3–11.
29. Rayed ME, et al. Deep learning for medical image segmentation: state-of-the-art advancements and challenges. Inform Med Unlock 2024;101504.
30. Nie Y, et al. Recent advances in diagnosis of skin lesions using dermoscopic images based on deep learning. IEEE Access. 2022;10:95716–47.
31. Yee D, et al. Medical image compression based on region of interest using better portable graphics (bpg). In 2017 IEEE international conference on systems, man, and cybernetics (SMC). IEEE; 2017. pp. 216–221.
32. Chen L, et al. Drinet for medical image segmentation. IEEE Trans Med Imaging. 2018;37:2453–62.

33. Subramanian M, Cho J, Sathishkumar VE, Naren OS. Multiple types of cancer classification using CT/MRI images based on learning without forgetting powered deep learning models. IEEE Access. 2023;11:10336–54.
34. Turhan H, Bilgin G. Semantic segmentation of histopathological images with fully and dilated convolutional networks. In 2021 medical technologies congress (TIPTEKNO). IEEE; 2021. pp. 1–4.
35. Fedorenko VV, Lyakhova UA, Nagornov NN, Efimenko GA, Kaplun DI. Semantic segmentation system of pigmented skin lesions based on convolutional neural networks. In 2022 11th Mediterranean conference on embedded computing (MECO). IEEE; 2022. pp. 1–5.
36. Khan MZ, Gajendran MK, Lee Y, Khan MA. Deep neural architectures for medical image semantic segmentation. IEEE Access. 2021;9:83002–24.
37. Alzubaidi L, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8:1–74.
38. Boulila W, et al. A transformer-based approach empowered by a self-attention technique for semantic segmentation in remote sensing. Heliyon 2024;10.
39. Rezk E, Haggag M, Eltorki M, El-Dakhakhni W. A comprehensive review of artificial intelligence methods and applications in skin cancer diagnosis and treatment: emerging trends and challenges. Healthc Anal 2023;100259.
40. Choy S. et al. Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease. npj Digit Med 2023;6(1).
41. Patel RH, Foltz EA, Witkowski A, Ludzik J. Analysis of artificial intelligence-based approaches applied to non-invasive imaging for early detection of melanoma: a systematic review. Cancers. 2023;15:4694.
42. Trager MH, Gordon ER, Breneman A, Weng C, Samie FH. Artificial intelligence for non-melanoma skin cancer. Clin Dermatol 2024.
43. Nofallah S, et al. Automated analysis of whole slide digital skin biopsy images. Front Artif Intell. 2022;5:1005086.
44. Manoorkar P, Kamat D, Patil P. Analysis and classification of human skin diseases. In 2016 international conference on automatic control and dynamic optimization techniques (ICACDOT). IEEE; 2016. pp. 1067–1071.
45. Xu Y, et al. Weakly supervised histopathology cancer image segmentation and classification. Med Image Anal. 2014;18:591–604.
46. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18. Springer; 2015. pp. 234–241.
47. Zhou Z, Siddiquee M, Tajbakhsh N, Liang J. Unet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging. 2019;39:1856–67.
48. Wazir S, Fraz MM. Histoseg: quick attention with multi-loss function for multi-structure segmentation in digital histology images. In 2022 12th international conference on pattern recognition systems (ICPRS). IEEE; 2022. pp. 1–7.
49. Thomas SM, Lefevre JG, Baxter G, Hamilton NA. Interpretable deep learning systems for multi-class segmentation and classification of non-melanoma skin cancer. Med Image Anal. 2021;68: 101915.
50. Asaf MZ, et al. A modified deep semantic segmentation model for analysis of whole slide skin images. 2023.
51. Geng J, et al. Exploring structural information for semantic segmentation of ultrasound images. In 2022 IEEE/CIC international conference on communications in China (ICCC). IEEE; 2022. pp. 570–575.
52. Xu Y, et al. Medical breast ultrasound image segmentation by machine learning. Ultrasonics. 2019;91:1–9.
53. Vakanski A, Xian M, Freer PE. Attention-enriched deep learning model for breast tumor segmentation in ultrasound images. Ultrasound Med Biol. 2020;46:2819–33.
54. Seif SR, Karimian A, Arabi H, Zaidi H. Impact of attention modules in deep learning-based semantic segmentation: evaluation for liver lesion segmentation from ct images. In 2022 IEEE nuclear science symposium and medical imaging conference (NSS/MIC). IEEE; 2022. pp. 1–3.
55. Li Y. Cru-net: a deep learning network for semantic segmentation of pathological tissue slices. In 2021 IEEE international conference on artificial intelligence and industrial design (AIID). IEEE; 2021. pp. 46–50.
56. Han C, et al. Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels. Med Image Anal. 2022;80: 102487.
57. Gu W, Wang S, Zhao S, Wan L, Zhu Z. Histosegrest: a weakly supervised learning method for histopathology image segmentation. In Proceedings of the 2022 5th international conference on image and graphics processing, 2022. pp. 189–195.
58. Fang Z, et al. Weakly-supervised semantic segmentation for histopathology images based on dataset synthesis and feature consistency constraint. In Proceedings of the AAAI conference on artificial intelligence. 2023; vol. 37, pp. 606–13.
59. Chen F, et al. G2ll: global-to-local self-supervised learning for label-efficient transformer-based skin lesion segmentation in dermoscopy images. In 2023 IEEE 20th international symposium on biomedical imaging (ISBI). IEEE; 2023. pp. 1–5.
60. Kaur R, Hosseini HG, Sinha R. Lesion border detection of skin cancer images using deep fully convolutional neural network with customized weights. In 2021 43rd annual international conference of the IEEE engineering in medicine & biology society (EMBC). IEEE; 2021. pp. 3035–3038.
61. Iranpoor R, Mahboob AS, Shahbandegan S, Baniasadi N. Skin lesion segmentation using convolutional neural networks with improved u-net architecture. In 2020 6th Iranian conference on signal processing and intelligent systems (ICSPIS). IEEE; 2020. pp. 1–5.
62. Yildiz S, Memiş A, Varl S. Nuclei segmentation in colon histology images by using the deep cnns: a u-net based multi-class segmentation analysis. In 2022 medical technologies congress (TIPTEKNO). IEEE; 2022. pp. 1–4.
63. Xu H, Liu L, Lei X, Mandal M, Lu C. An unsupervised method for histological image segmentation based on tissue cluster level graph cut. Comput Med Imaging Graph. 2021;93: 101974.
64. Li C, Lu X. Computer-aided detection breast cancer in whole slide image. In 2021 international conference on computer, control and robotics (ICCCR). IEEE; 2021. pp. 193–198.
65. Rasool A, Fraz MM, Javed S. Multiscale unified network for simultaneous segmentation of nerves and micro-vessels in histology images. In 2021 International conference on digital futures and transformative technologies (ICoDT2). IEEE; 2021. pp. 1–6.
66. Gite S, Mishra A, Kotecha K. Enhanced lung image segmentation using deep learning. Neural Comput Appl. 2023;35:22839–53.
67. Sun J, Zhang Y, Zhu J, Wu J, Kong Y. Semi-supervised medical image semantic segmentation with multi-scale graph cut loss. In 2021 IEEE international conference on image processing (ICIP). IEEE; 2021. pp. 624–628.

68. Pang S, et al. Spineparsenet: spine parsing for volumetric mr image by a two-stage segmentation framework with semantic image representation. IEEE Trans Med Imaging. 2020;40:262–73.
69. Sun J, Zhang K, Niu S, Zhang Y, Kong Y. Feature space message passing network for medical image semantic segmentation. In ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2022. pp. 1081–1085.
70. Pot M, Kieusseyan N, Prainsack B. Not all biases are bad: equitable and inequitable biases in machine learning and radiology. Insights Imaging. 2021;12:13.
71. Xi P, Shu C, Goubran R. Abnormality detection in mammography using deep convolutional neural networks. In 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA). IEEE; 2018. pp. 1–6.
72. Deepa R, Arunkumar S, Jayaraj V, Sivasamy A. Healthcare's new frontier: Ai-driven early cancer detection for improved well-being. AIP Adv 2023;13.
73. Yousefpour Shahrivar R, Karami F, Karami E. Enhancing fetal anomaly detection in ultrasonography images: a review of machine learning-based approaches. Biomimetics. 2023;8:519.
74. Celebi ME, Codella N, Halpern A. Dermoscopy image analysis: overview and future directions. IEEE J Biomed Health Inform. 2019;23:474–8.
75. Li H, Pan Y, Zhao J, Zhang L. Skin disease diagnosis with deep learning: a review. Neurocomputing. 2021;464:364–93.
76. Gulzar Y, Khan SA. Skin lesion segmentation based on vision transformers and convolutional neural networks-a comparative study. Appl Sci. 2022;12:5990.
77. Eskandari S, Lumpp J. Inter-scale dependency modeling for skin lesion segmentation with transformer-based networks. arXiv preprint arXiv:2310.13727 2023.
78. Zhou X, et al. A comprehensive review for breast histopathology image analysis using classical and deep neural networks. IEEE Access. 2020;8:90931–56.
79. Dabeer S, Khan MM, Islam S. Cancer diagnosis in histopathological image: Cnn based approach. Inform Med Unlock. 2019;16: 100231.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.