Enhancing Deep Learning Robustness for Small Organ Segmentation in Medical Imaging: a Comprehensive Review and Future Roadmap

M Karthik Prabu1,a), Mathan Kumar2,b)

1Department of Computer Science and Engineering, R.V.S. College of Engineering, Dindigul, Tamil Nadu, India

2Department of Information Technology, PSNA College of Engineering and Technology, Dindigul. Tamil Nadu, India

Corresponding author: a)[karthikprabu777@gmail.com](mailto:karthikprabu777@gmail.com), b)[mathan81@gmail.com](mailto:mathan81@gmail.com)

**Abstract:** Precise segmentation of small organs in medical imaging remains a difficult task due to their small statures, unreliable boundaries, anatomical heterogeneity, and class imbalance in medical datasets. Deep learning has totally altered medical image analysis by surpassing traditional manual and semi-automated segmentation techniques, providing enhanced precision, robustness, and scalability. This study presents a comprehensive summary of recent advances in deep learning networks for small organ segmentation across various imaging modalities that include CT, MRI, PET, ultrasound, and hyper spectral imaging. Foundational architectures, including Convolutional Neural Networks (CNNs) and Fully Convolutional Networks (FCNs), along with updated variants like U-Net, UNet++, nnU-Net, HRNet, and Vision Transformers (ViTs), have exhibited exceptional success in characterizing complex anatomical structures. Hybrid architectures that integrate CNNs with transformers and attention mechanisms have led to further advancements in contextual learning and boundary localization, specifically in challenging environments. The review highlights the methodological innovations such as multi-scale feature extraction, patch-wise sampling, sparse attention, lightweight 3D U-Nets, and adversarial training, which address the research challenges such as data scarcity, noisy annotations, and domain shift across imaging devices. Evaluation metrics, including Dice coefficient, IoU, and Hausdorff distance, are discussed in relation to their clinical relevance, especially for small structures where edge accuracy is crucial. Furthermore, the paper explores the incorporation of domain knowledge, semi-supervised learning, federated learning, and generative models to overcome annotation bottlenecks and improve generalizability. By systematically mapping 100 studies published between 2016 and 2023, this review pinpoints key research trends, persistent challenges, and research gaps in clinical translation. Future directions highlight the development of explainable AI models, uncertainty quantification, compact architectures for real-time deployment, and integration with image genomics for accurate medicine. Ultimately, this roadmap underscores the potential of deep learning to achieve reliable, clinically viable small organ segmentation, advancing both diagnostic accuracy and therapeutic planning.

**Keywords:** Small organ segmentation, medical imaging, deep learning networks, CNN architecture, UNet, patch-wise sampling, lightweight 3D U-Nets, sparse attention.

# INTRODUCTION

Organs in the human body are specialized structures, each of which performs a specific task. Every single cell is an individual worker that has a specific task to perform. Cells with shared roles team up in the same place, forming tissues. Organs are made of multiple tissues which work together cooperatively to execute complex functions. Organs are grouped into systems that manage interconnected jobs. Some organs in our human body even contribute to more than one system. Heart, brain, stomach, liver, and kidney are some of the well-known organs in the human body. Organs can either be a single structure, like the heart or liver, or they can also be a collection of similar types of structures. Apart from the best-known large structures or organs, the human body consists of many other relatively smaller structures that count as organs too. The stapedius is the smallest muscle in the human body, located in the middle ear. It protects the inner ear from loud noises and also contracts to prevent one's own voice from sounding loud in our head. Malleus, incus, and stapes, collectively known as the ossicles, are the smallest bones in the human body that convey sound from the outer ear to the inner ear. Another of the smallest yet crucial organs in the human body is the pineal gland, located near the center of the brain. It is the smallest endocrine gland that produces melatonin, a hormone that regulates our sleep cycle and how we react to seasonal changes. Medical imaging refers to the various techniques and processes of imaging used to view the interior of a body. This aids clinical analysis and medical care and provides a visual representation of the function of some organs or tissues. Medical imaging seeks to reveal internal structures of the human body, as well as to diagnose, monitor, or treat medical conditions. It also builds a database of normal anatomy and physiology. This allows for identifying abnormal conditions. Pathology, which is the process of imaging removed organs and tissues, is not considered to be a part of medical imaging. Each imaging technology offers unique data about the specific area of the human body being studied or being treated. This data relates to disease, injury, or effectiveness of medical treatment.

Electroencephalography (EEG), magnetoencephalography (MEG), and electrocardiography (ECG) are some of the measurement and recording techniques that are not primarily designed to produce images. They produce data susceptible to representation as a parameter graph versus time or maps that contain data about the measurement locations. In a limited comparison, these technologies can be considered forms of medical imaging in another discipline of medical instrumentation. Most imaging methods are "non-invasive"[1], which means no instrument is introduced or enters the human body. Small organ segmentation in medical images is a challenging task. The medical images of these organs include small representations with unclear edges of the organs. They also suffer from class imbalance; there are far fewer small organ parts than the larger ones in the background. This makes model training a difficult process. Anatomical variations and image flaws add to the difficulty. Researchers employ several strategies to improve segmentation accuracy. Multi-scale feature extraction captures details at different levels. Two-stage segmentation uses a rough outline and then refines it. Including information regarding the organ's positions helps in the segmentation process. Attention methods guide models to focus on important areas. Adversarial training makes models more robust. Knowledge-guided models use existing anatomical facts. Accurate segmentation of small organs has vital applications such as radiotherapy planning and protecting healthy tissue. It also aids surgeons in planning complex operations. Segmentation also assists in computer-aided diagnosis, finding abnormalities.

Deep learning has rapidly evolved the field of abdominal organ segmentation, which offers significant benefits over traditional manual methods [2] [3]. Deep learning-based semantic organ segmentation in hyperspectral images (HSI) is an advancing field that offers detailed and robust tissue differentiation [4], leveraging the extensive spectral information beyond standard RGB or greyscale imaging. Robust deep learning-based semantic organ segmentation in hyperspectral images is more precise, performs at or above expert level, and brings notable benefits in resilience and tissue differentiation, especially in challenging clinical settings[5]. The state of autonomous deep learning-based multi-organ segmentation is thoroughly analysed in [6], which highlights the field's rapid progress and evolving approaches. Deep learning methods have far outperformed conventional approaches in the automatic segmentation of organs. It was mostly ascribed to data-driven feature extraction and end-to-end learning. The review referred to 327 papers published from January 2016 to December 2023, of which 195 met PRISMA systematic review criteria, providing a thorough mapping of research activity and leading methods in the past decade. Convolutional Neural Networks (CNNs) and Fully Convolutional Networks (FCNs) are foundational architectures, powering state-of-the-art medical segmentation strategies, especially with variants such as U-Net, UNet++, nnU-Net, HRNet, Vision Transformer (ViT), DUCK-Net, and others. Transformers (ViT) and hybrid deep architectures have emerged, improving contextual learning and attention-based localisation, especially for complex anatomical regions [7].

The U-Net framework and improved approaches (e.g., UNet++, nnU-Net) dominate biomedical segmentation, excelling at extracting object boundaries with limited annotated data. These approaches are applicable for a variety of objects, from organs to tumours, lesions, and cellular/nuclear elements. FCNs are widely employed for direct pixel-to-pixel mapping, key for semantic segmentation across various imaging modalities [8]. Hybrid models—combining CNNs, transformers, and recurrent layers—yield performance boosts for complex tasks [9]. [10] addresses a fundamental obstacle in medical image segmentation—handling imperfect datasets that are typical in clinical imaging scenarios. This study emphasises the reality that perfectly annotated medical image datasets are rare and expensive. It surveys a wide array of deep learning methodologies developed to embrace imperfections in annotation and dataset quality, enabling effective medical image segmentation despite these constraints. [11] systematically reviews how domain knowledge from medical experts and clinical understanding have been incorporated into deep learning models to enhance medical image analysis. It underscores the importance of embedding medical domain knowledge into deep learning to tackle data scarcity, improve transparency, and enhance clinical relevance in medical image analysis tasks like diagnosis, lesion detection, and segmentation.

[12] introduces AFFSegNet (Adaptive Feature Fusion Segmentation Network), a novel Transformer-based deep learning architecture designed for precise medical image segmentation, with a particular focus on microtumours and multi-organ delineation. The network effectively integrates local and global contextual features across multiple scales, addressing challenges in segmenting small and irregularly shaped tumours and multiple organs simultaneously. Its design enables it to outperform existing benchmarks across multiple challenging medical imaging datasets. [13] presents a publicly accessible deep learning model specifically designed for segmenting skeletal muscle, subcutaneous adipose tissue (SAT), and visceral adipose tissue (VAT) in CT images, covering the chest, abdomen, and pelvis regions. [4] is a comprehensive review of deep learning-based methods for medical image multi-organ segmentation; it categorizes approaches primarily into two broad groups: pixel-wise classification and end-to-end segmentation. It concluded that deep learning for multi-organ segmentation has matured significantly, delivering expert-level performance on key anatomical targets while continuing to evolve through novel architectures and learning paradigms.

A comprehensive international challenge was designed to develop and analyze generalizable algorithms for medical image segmentation across a wide range of clinical tasks. The MSD has notably contributed to updating the state-of-the-art in medical image segmentation, demonstrating that generalized deep learning models can perform robustly across different medical imaging tasks without task-specific customization [14]. [15] explores techniques and remedies in medical image analysis, highlights the challenge of label noise in medical imaging datasets and surveys methods to handle it. Overall, the paper serves as a comprehensive review of various algorithmic strategies and recommendations to enhance deep learning accuracy when training with noisy or imperfect labels in medical image analysis. [16] demonstrates powerful semi-supervised learning strategies for laparoscopic image segmentation to enhance performance when annotation resources are constrained. [17] is a seminal paper that extends DenseNets for the task of semantic segmentation. This architecture demonstrates robust performance and parameter efficiency compared to other segmentation networks of the time due to the DenseNet connectivity pattern. [18] provides a state-of-the-art overview of deep learning, especially diffusion-based approaches, in hyper spectral image analysis as of 2025. Other relevant reviews include broad enhancement and classification methods, emphasizing the integration of deep learning models for enhanced hyper spectral image understanding and utilization across various fields.

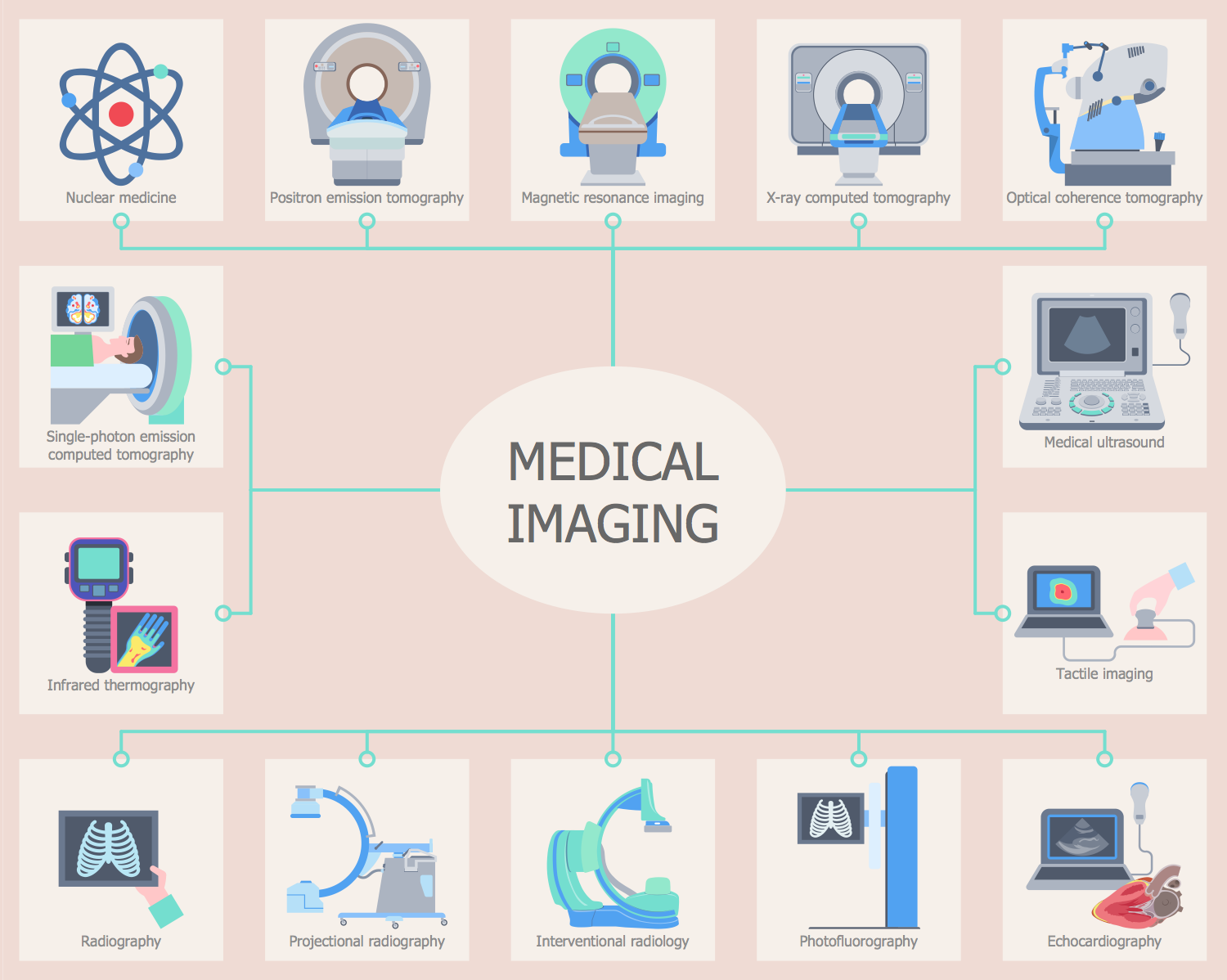
**Table1.** Previous researches on medical image segmentation.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ref.** | **Year** | **Network** | **Supervision** | **Site** | **Modality** |
| [19] | 2021 | WBNet | Weakly-supervised | Whole body (organs at risk) | CT |
| [20] | 2021 | U-Net | Supervised | Abdominal organs | MRI, CT |
| [21] | 2020 | SIFA | Unsupervised | Cardiac substructure, Abdominal organs | MRI, CT |
| [22] | 2020 | GAN | Unsupervised | Abdomen, Brain | MRI, CT |
| [23] | 2021 | CycleGAN | Unsupervised | Multiple organs | CT |
| [24] | 2021 | CycleGAN | Unsupervised | Stomach | MRI |
| [25] | 2021 | AIDE | Supervised and unsupervised | Breast | MRI |
| [26] | 2020 | 3D-Unet | Supervised | Abdominal and thoracic region | MRI |
| [27] | 2021 | U-Net | Supervised | Liver, Head, Neck | HSI |
| [28] | 2021 | LDA | Unsupervised | Sepsis diagnosis | HSI |
| [29] | 2018 | Hybrid structure (SVM, KNN, K-Means algorithm) | Supervised and unsupervised | Brain | HSI |
| [30] | 2021 | Mask R-CNN | Supervised | RLN in thyroid gland | RGB photographs |
| [31] | 2020 | U-Net | Supervised | Liver, Gall bladder, Abdominal wall | Laparoscopic images |
| [32] | 2021 | GJAM | Unsupervised | Whole body | HSI |
| [33] | 2016 | 2D-FCN | Supervised | Torso | CT |
| [34] | 2017 | DDCNN | Supervised | Bladder, intestine, colon | CT |

[35] presents a novel set of metrics designed to evaluate medical image segmentations more efficiently by focusing on boundary overlaps instead of just region overlaps. These boundary overlap metrics offer an improved toolset for evaluating medical image segmentation algorithms by addressing boundary-specific accuracy, which is crucial for clinical relevance. The deep learning networks and the imaging modalities used in the previous research papers are listed in table 1.

# SMALL ORGAN SEGMENTATION IN MEDICAL IMAGING

Medical image segmentation isolates regions of interest (ROIs) from 3D scans like Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) [36]. The primary goal of medical segmentation is to find a specific anatomy as required for a specific study. This is useful for simulating physical properties or planning CAD-designed implant placement within a patient [37]. Segmenting images using conventional techniques can be a time-consuming task [38]. New advances in AI software now simplify these routine tasks, making it easier to complete within a short duration. Medical imaging encompasses a range of widely used techniques designed to create visual representations of the interior of the human body [7]. These techniques are employed in clinical analysis, medical intervention, or to illustrate the functions of specific tissues or organs. Medical imaging is primarily used to unveil the concealed internal structures situated within or obscured by the bones and skin. It also aids in diagnosing and treating various diseases. Medical imaging is also used for establishing a database of normal anatomy, enabling the detection of abnormalities. As a discipline, medical imaging is known to be a part of biological imaging; it incorporates radiology, which includes various imaging technologies of magnetic resonance imaging, X-ray radiography, medical ultrasonography (ultrasound), elastography, endoscopy, tactile imaging, thermography, medical photography, and nuclear medicine functional imaging techniques, such as single-photon emission computed tomography and positron emission tomography. Recording and measurement techniques, such as electroencephalography, magnetoencephalography, and electrocardiography, represent additional technologies that generate data amenable to representation as parameter graphs over time or as maps detailing the locations of the measurements. These technologies are forms of medical imaging in another discipline. By 2010, as many as 5 billion medical imaging studies had been conducted globally. Being often perceived as a tool for designating the techniques that produce images of the internal aspect of the body, medical imaging can be a solution to numerous mathematical inverse problems [39]. Various techniques of medical imaging modalities used in the clinical environment are shown in figure 1.

****

**Figure 1.** Various types of medical imaging modalities.

Segmentation of medical scans like CT and MRI generally starts by collecting information from the background data and then utilises it to create a mask. Depending on the task, users then segment scans in 2D or 3D views. Segmentation software offers a wide range of tools and algorithms that are either fully manual or semi-automated operations [40]. The various tools offered include manual painting, thresholding, and region growing. Specific applications exist for cardiovascular image segmentation, with specific options for working with diverse heart cases [41]. Many medical data tasks require only a few segmentation tools. Placing medical devices can involve a few steps to segment regions of interest in a bone. Such segmentation tasks can be automated using scripts or AI [42], [43]. Complex cases, especially those that deal with unusual pathologies or complex traumas, however, may need more time and diverse software features to generate the required segmentation outcome.

Recently, segmentation software that incorporates deep learning techniques is widely used to work with medical images, both for clinical applications and general research involving scan data [43]. Accurate and fast segmentation is crucial to many applications. More recently, AI-based machine learning modules have been added to segmentation software to speed up the segmentation of orthopaedic and cardiovascular data [44]. In this context, common and time-consuming steps for extracting regions of interest can be completed within seconds, rather than relying on significant manual work. With this approach, based on medical training datasets and verified by clinicians, users can work faster and overcome typical workflow bottlenecks. This helps to speed up the process of going from image import through to models ready for analysis, design work, and simulation.

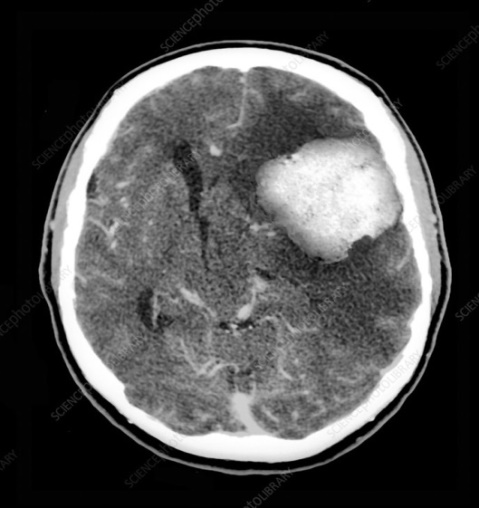
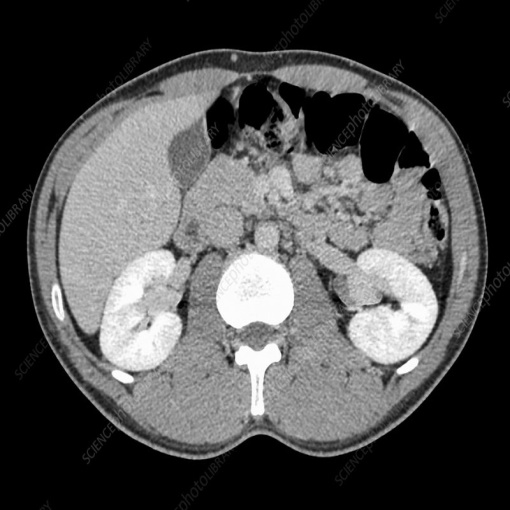
This study aims to provide a comprehensive overview of several widely used imaging modalities employed in clinical settings, including CT, MRI, PET and ultrasound [45], [46]. These are some essential imaging methods that are employed in the segmentation and imaging of different areas of the human body [47][48].

## Computed Tomography (CT)

Computed tomography is a computerised x-ray imaging procedure in which a narrow beam of x-rays is aimed at a patient. The machine spins around the body, creating signals. These signals are processed by the machine’s computer to develop cross-sectional images, also known as “slices”. These slices are tomographic images and can provide a clinician more details than conventional x-ray techniques. Once a number of successive slices are produced, the machine’s computer can digitally stack them together to form a three-dimensional (3D) image of the patient. This 3D image enables easier identification of basic structures. It also helps spot tumours and abnormalities. Unlike traditional x-rays—which use a fixed x-ray tube—the CT scanner uses a motorised x-ray source. It spins around the circular opening of a doughnut-shaped structure called a gantry[49].

During a CT scan, the x-ray tube revolves around the patient, passing narrow beams of x-rays through the body. The patient is positioned on a bed that slowly moves across the gantry. CT scanners use specialised digital x-ray detectors, which are situated directly across from the x-ray source, in place of film. The x-rays passing through the patient are picked up by the detectors, and it is transmitted to a computer. A two-dimensional image slice of the patient is constructed each time the x-ray source completes a full rotation. The CT computer uses sophisticated mathematical techniques to construct the slices. The thickness of the tissue represented in each image slice varies depending on the CT machine used but usually ranges from 1 to 10 millimetres. Once a full slice is completed, the image is saved in the computer, and the motorised bed is moved forward incrementally into the gantry. The scanning process is then repeated to produce another image slice. It continues until the required number of slices is gathered. The image slices gathered can either be displayed individually, or they can be stacked together by the computer to generate a 3D image of the patient. The 3D image thus generated shows the skeleton, organs, and tissues, as well as any abnormalities that the clinician is trying to spot. This technique possesses various benefits, including the ability to rotate the 3D image in space or to view slices in succession. This makes it easier to find the precise location where an abnormality may be located. Figure 2 displays two CT scan images providing a clear view of the human brain and abdomen.

Dense structures in the body such as bone show up clearly on X-rays, whereas soft tissues vary in their ability to block x-rays. This makes them faint or hard to view. Contrast agents were developed that when used makes such faint tissue highly visible in an x-ray or CT scan and its also safe to use in patients. Contrast agents hold materials that can block x-rays. This makes them stand out on X-ray images. For example, iodine-based intravenous (IV) contrast agents are injected into the bloodstream that highlights blood vessels. It examines the circulatory system and helps spot blockages in blood vessels including those in the heart. Barium-based compounds are used as oral contrast agents. They image the digestive system that includes the esophagus, GI tract and stomach [50]. CT scans can easily spot serious issues like haemorrhage, blood clots, or cancer [51]. Early diagnosis in such scenarios can save lives. However, CT scans use X-rays that emit ionising radiation. Radiations can harm living cells, and the risk grows with each exposure. The risk of getting cancer from X-rays is negligible. This is true unless the scan is of the abdomen or pelvis. Doctors usually prefer MRI or ultrasound for the abdomen or pelvis.

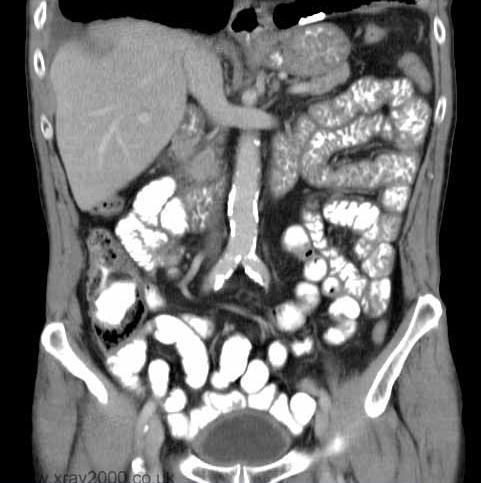
(a) (b)

**Figure 2.** CT scan images of the human brain and abdomen.

These methods do not emit any form of radiation. If MRI or ultrasound are not suitable, or in emergencies, CT scans are preferred. Contrast agents can cause allergies or even kidney problems in certain patients. IV contrast agents are not recommended to patients whose kidney function is poor, as it can worsen kidney issues, sometimes permanently. Children are more sensitive to radiation and relatively have higher risks of getting radiation cancer compared to adults.

## Magnetic Resonance Imaging (MRI)

MRI is a non-invasive medical imaging modality that generates detailed 3D anatomical images. It is a safe imaging method that is often used for detecting disease and diagnosis, and it is also used in treatment monitoring [52]. It working is based on advanced technology that excites protons. It then detects the change in the direction of the rotational axis of protons found in the body water that makes up living tissues. Magnetic Resonance Imaging, or MRI, is a powerful diagnostic tool that uses incredibly strong magnets to create detailed pictures of the inside of the body [53]. These magnets generate a powerful magnetic field which acts like a strong force, making certain particles in the body line up. These particles are called protons; they are found in the water molecules that make up most of our bodies. The MRI's magnetic field forces these protons to align themselves with it. The protons in the body point in the same direction as the magnetic field. Subsequently, a brief pulse of radiofrequency energy is sent through the patient. This energy is like a nudge, and it temporarily knocks the aligned protons out of their ordered position. The protons spin away from the magnetic field's pull. The protons are momentarily "out of equilibrium". When the radiofrequency pulse is turned off, the protons return to their original aligned state. They realign with the powerful magnetic field. As they shift back, they release the extra energy they absorbed from the radiofrequency pulse. Special sensors within the MRI machine are designed to detect this released energy. They are incredibly sensitive, and they can even measure the timing of this energy release and the amount of energy that is given off. These measurements are crucial, and they can vary depending on the protons' surroundings. The signals are influenced by the specific chemical environment and the types of molecules these protons belong to. Different tissues have different chemical compositions. This means the protons within them will release energy at different rates and amounts. For instance, water molecules in fat tissue behave differently than water molecules in muscle tissue. These subtle differences in magnetic properties allow doctors to differentiate between various types of body tissues. They can see healthy tissue, diseased tissue, or scar tissue. To get an MRI scan, a patient lies inside a large, tunnel-like machine that houses the powerful magnet. It is crucial for the patient to remain absolutely still during the entire scan. Even slight movements can blur the resulting images, making them unreadable. Sometimes, to enhance the visibility of certain structures, a special dye is used. This is called a contrast agent. These agents are often given to patients through an intravenous line, either before or during the scan. They typically contain a metal called gadolinium, which helps the protons realign with the magnetic field more quickly. A faster realignment results in a brighter signal on the MRI image. This increased brightness can make abnormalities more apparent to the radiologist.



(a)



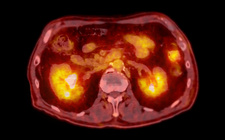
(b)

**Figure 3.** (a) (b) MRI scans of the abdominal region.

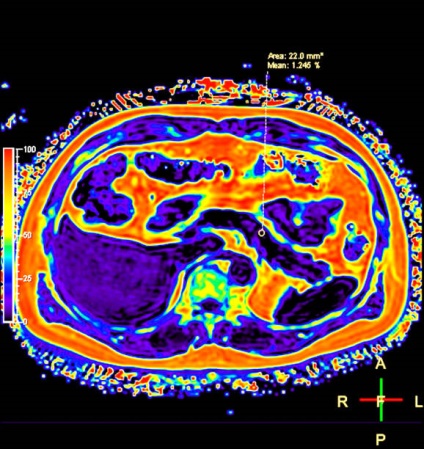
MRI scanners excel at imaging soft tissues. Unlike CT scans, MRI does not make use of harmful x-rays. MRI images can provide a clear view of the brain, spinal cord, nerves, muscles, ligaments, and tendons [54]. This makes it ideal for injuries to the knee and shoulder. MRI can also distinguish between white and grey matter in the brain. It can also spot aneurysms and tumours. MRI is the most preferred method for frequent brain imaging since it avoids radiation [55]. However, MRI is more costly than x-rays or CT. Functional MRI (fMRI) can be used to observe brain activity during tasks. It helps understand brain organisation and assess neurological status. The MRI screening of the abdominal region is shown in figure 3.

## Positron Emission Tomography (PET)

A positron emission tomography, or PET scan, is an imaging modality which shows how the organs and tissues function. The test uses a safe radioactive chemical called a radiotracer. It is injected into the body, and then the PET scanner detects the radiotracer. The scanner identifies unhealthy cells. These cells absorb more radiotracer; that often signals an abnormality. Doctors often prefer PET scans for cancer diagnosis, and they also use them to check the effectiveness of the cancer treatment. A PET scan can also evaluate certain heart and brain conditions. A PET scan is a nuclear medicine imaging test. Nuclear medicine utilises tiny, safe amounts of radioactive material. The radiotracer is injected into the patient's body through an IV. PET scans look at body functioning, and they see these activities at a molecular level. This differs from other imaging methods. PET scans can detect disease very early. Sick cells take up more radiotracer than healthy cells. These areas show up as "hot spots". The PET scanner senses this radiation. It then creates images of the problem areas. A PET/CT scan joins PET images with X-ray images from a CT scan. The combination of PET and CT scans can produce 3D images that aid in a more accurate diagnosis. PET scans are generally safe. The radiation dosages are low. It leaves the body quickly. Allergic reactions to the radiotracers are rare and mild. However, persons with diabetes may need dietary changes before a scan. PET/CT scan images of the abdominal region are shown in figure 4.



(a)



(b)

**Figure 4.** (a) (b) PET/CT scans of the abdominal region.

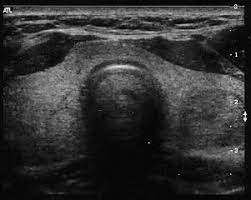
## Ultrasound

Sonography, or ultrasound, is a safe medical imaging modality. The image obtained from this test is known as a sonogram. Ultrasound employs sound waves at high speeds. These waves generate live images or videos of organs and soft tissues. It can also be used to view blood vessels. This technology enables doctors to view inside tissues without making incisions. It also avoids radiation, unlike X-rays. Ultrasound is often linked to pregnancy, while medical experts use ultrasound in a wide range of diagnoses. They examine various body parts with this test. A healthcare provider uses a transducer, also known as a probe, during an ultrasound. This device moves over the body or sometimes into a body opening. A thin layer of gel is applied to the skin that allows ultrasound waves to travel from the transducer, through the gel, and into the body. The probe converts electricity into sound waves. These high-frequency waves then go into the body's tissues. Yet these sound waves are not audible. The sound waves reflect off internal body structures. They return to the probe, changing back into electrical signals. A computer processes these signals and then creates live pictures or videos which are shown on a nearby screen.

Prenatal ultrasound, diagnostic ultrasound and ultrasound guidance for procedures are the three major classifications of the ultrasound imaging modality. Diagnostic ultrasounds enable doctors to view inside your body. It helps them find any abnormalities. They can explain symptoms like unknown pain or lumps. Ultrasounds also help clarify abnormal lab results. The type of ultrasound depends on the respective medical situation. Common diagnostic ultrasounds include abdominal ultrasound, kidney ultrasound, breast ultrasound, Doppler ultrasound, pelvic ultrasound, transvaginal ultrasound, thyroid ultrasound and transrectal ultrasound. The ultrasound images of various parts of the human body such as the (a) prenatal ultrasound, (b) abdomen, (d) breasts, and (c) thyroid are shown in the figure 5.



(a) (b)

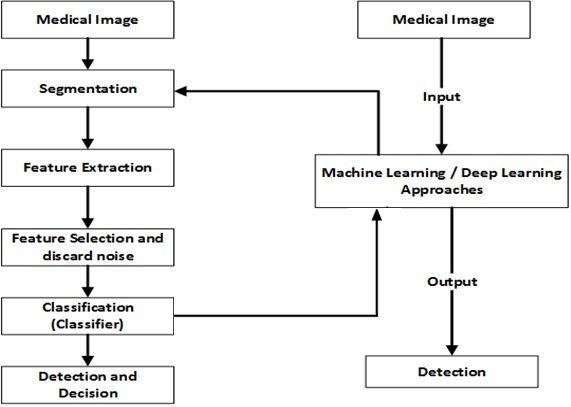


(c) (d)

**Figure 5.** (a)-(d) Ultra sound scans images of various parts of the human body.

# Deep learning in Small Organ Segmentation

Deep learning, a branch of machine learning, aids in creating models for abnormality identification [56]. These models can be automatic, semi-automatic, or hybrid. The core distinction lies in computational layers. "Deep" refers to these layers. Deep learning models utilize many layers for training. The path for assigning credit, or CAP, is vital. CAP links input to output. Deep learning models find features and patterns independently [57]. They decide which feature fits which level. Artificial neural networks extract complex data patterns. These networks mimic the brain's structure. They use many connected parts to process information. Deep learning handles data like images and text. These models can also improve their results. They adapt easily to different data and tasks. Machine learning models require supervised training. A programmer must identify the model's features. Feature extraction skill directly impacts model accuracy. This work is difficult and lengthy. Deep learning offers a key benefit: models self-define features via unsupervised learning. The system first receives prepared data, like images, text, or sound. This data is then divided into training and testing sets. The system learns features from training data. It uses these features to build a predictive model. Deep learning processes input data using nonlinear transformations. It then uses learned information to create the predictive model. Initial predictions may be imperfect. However, models grow more complex with each run. Their predictions steadily improve. This continues until desired accuracy is achieved. Back propagation is vital for boosting deep learning model accuracy. It computes loss function gradients. Then, it adjusts network weights for better performance [58]. Deep learning methods broadly fall into three main types. These are supervised, unsupervised, and reinforcement learning [59][60]. Supervised learning uses labeled data for predictions. Unsupervised learning finds patterns in unlabeled data. It then makes predictions from those patterns. Reinforcement learning makes decisions through trial and error [61]. Figure 6 illustrates a flowchart of medical image processing using deep learning techniques.



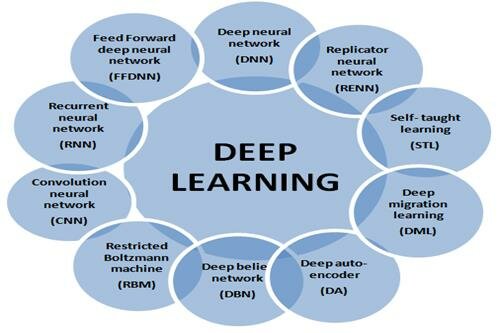
**Figure 6.** Basic flowchart of medical image segmentation using deep learning.

Supervised learning trains a model with labeled data. Each training data point has an input and its correct output. Analysts examine many such pairs to find the correct output. These methods create a function mapping inputs to outputs. After training, the model is tested with new data. Training continues until the model meets accuracy goals. More examples help improve accuracy and reduce errors. A trained and tested model can then predict outcomes for new, unseen data. Common uses include risk assessment in finance. Image sorting, online recommendations, and fraud detection also use it. Supervised learning has two main branches: classification and regression. Classification sorts data into groups. It finds a category for input data. Classification is used when the output is a distinct class. Decision trees, random forests, and support vector machines are common classification tools. Logistic regression is also widely used. Regression predicts real or continuous values. Examples include weather forecasts and market trends. This method works when input and output variables relate. Regression trees, Bayesian linear regression, and polynomial regression are frequent choices. Linear and non-linear regression is also popular.

Unsupervised learning uses data without labels. It finds patterns without guidance. The system gets raw data. Models find hidden patterns without human help. They then group data by similarities. The goal is to understand data structure. It groups similar data and compresses it. These algorithms learn without labels. They are good for complex tasks. Examples include anomaly detection and customer grouping. Clustering and association are its main types. Clustering groups data based on common traits. It is a popular method. It splits data into clusters. Types include exclusive and hierarchical clustering. Association finds links between data. It looks for if-then rules. It shows connections between data items. It helps in medical diagnosis and marketing. Common algorithms include apriori and FP-growth. Reinforcement learning learns by acting in an environment. Google Deep Mind advanced it in 2013. The agent gets signals from the environment. Its actions earn rewards or penalties. This differs from supervised learning. There is no labeled data or explicit loss function. The agent learns optimal actions. Deep reinforcement learning works for many parameters. It suits complex problems. Derivative-free methods work with fewer parameters. Businesses use it for strategy and automation. It helps agents gain maximum rewards over time. It guides actions in specific situations. It needs much computing power and time. It is less useful with ample supervised data.

## Deep Learning Networks

A deep learning network, also known as a deep neural network, is a type of artificial neural network with multiple layers between the input and output layers. These hidden layers allow the network to learn complex patterns and representations from data, enabling it to perform tasks like image recognition, natural language processing, and more [62] [63]. Deep learning builds on artificial neural networks. These mimic the human brain's structure and function. The "deep" aspect means many hidden layers. This lets networks learn data patterns step by step. Deep learning models find features on their own. This removes manual work from older machine learning. Large data sets and algorithms train these networks. Back propagation fine-tunes network settings. This boosts how well the network performs. Deep learning powers many tools. These include recognizing images and speech. It aids language tasks, robots, and more [64]. The network gets input data. This could be pictures, words, or sensor info. Data moves through several hidden layers. Each layer processes the data. The last layer gives the result. This result might be a category label. It could be a future prediction. Or it could be another needed outcome. During learning, the network updates its parts. It uses prediction errors to adjust. This loop repeats until accuracy is good. Figure 7 illustrates the various types of networks in deep learning. Deep learning models like CNN, U-Net, attention-based or transformer architectures have become dominant due to their superior performance and automation capabilities [65]. A clinically applicable deep learning framework can significantly improve the accuracy and efficiency of the segmentation tasks [66].



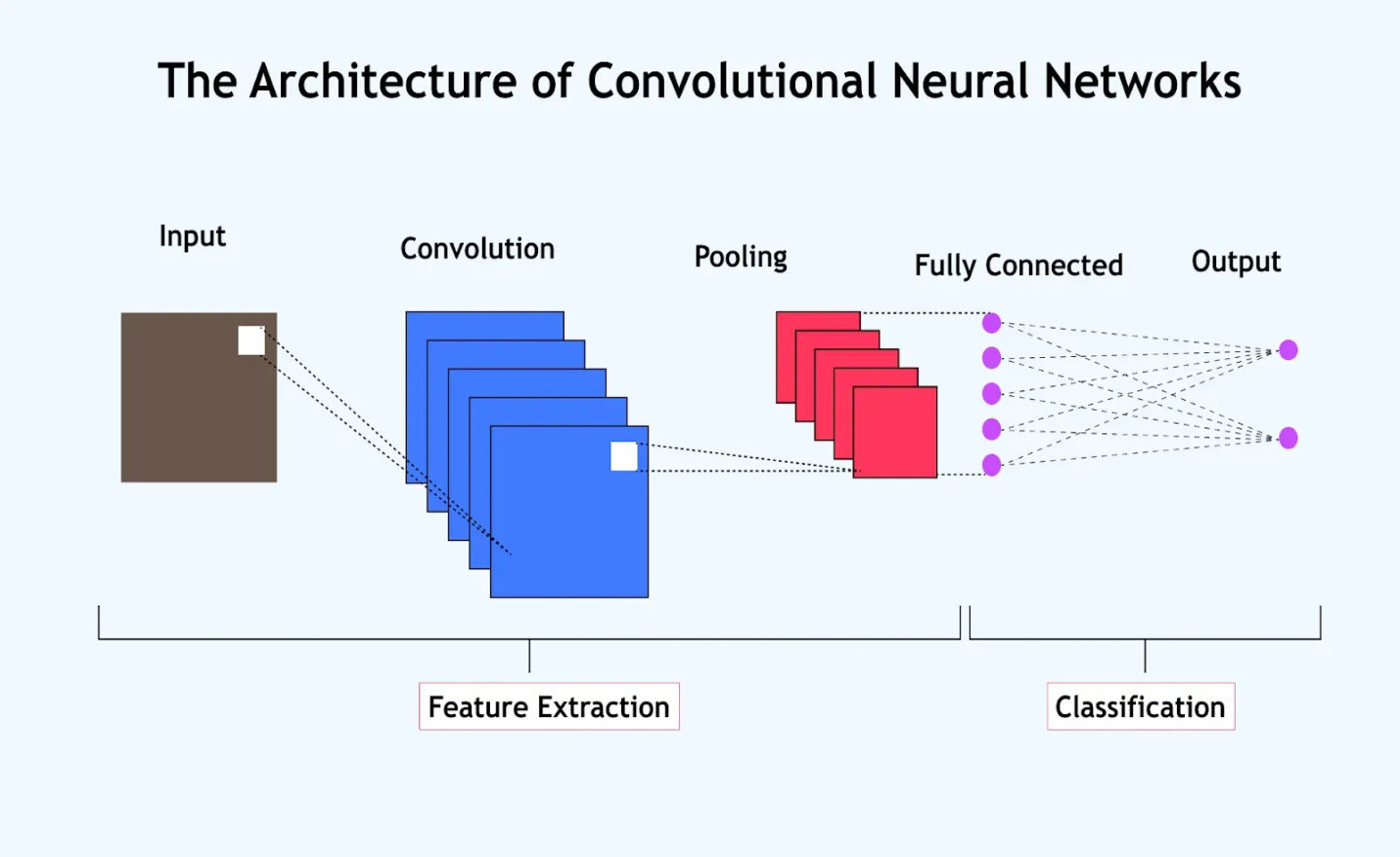
**Figure 7.** Types of deep learning networks.

### Convolutional Neural Networks (CNN)

Convolutional neural networks, or CNNs, are AI systems. They use layered networks to find and sort objects in images. CNNs, also called ConvNets, are a leading deep learning design that learns directly from data. This removes the need for manual feature selection. These networks excel at image tasks and are used for visual recognition and medical scans. They also aid in image segmentation and language processing [67]. Its design suits 2D shapes well [68]. CNNs surpass standard networks, and they find important data on their own. Human input is not required in these networks. CNN is typically composed of three layers [69], namely:

* Convolutional layer
* Pooling layer
* Fully connected layer

Convolutional layers perform the convolution operation. This process transforms pixels in a receptive field into one value. Applying convolution to an image shrinks its size, and it also consolidates field information into a single pixel. The final output of this layer is a vector. Different convolution types suit various problems. They also help learn specific features. The typical convolution is the 2D convolution layer, often called conv2D. A conv2D filter, or kernel, moves across the 2D input data. It multiplies elements directly, and then it adds up these products for one output pixel. The kernel repeats this process everywhere. It changes one 2D feature map into another. The size of input data and neural network settings are vital in machine learning. Stacking pooling layers controls these dimensions. Pooling layers process each data channel separately. They summarise values to retain key features. This summarised data feeds the next layer. Repeating this process shrinks spatial dimensions. Various methods aggregate these values. Fully connected layers, also known as dense layers, follow convolutional and pooling layers in a CNN [70]. They process features learnt by earlier layers. These features activate selectively. The dense layers then merge these to make predictions. Every neuron links to every other neuronee in dense layers. This connection scheme gives them their name. They make the final classifications. For image tasks, dense layers decide an image's category [71].

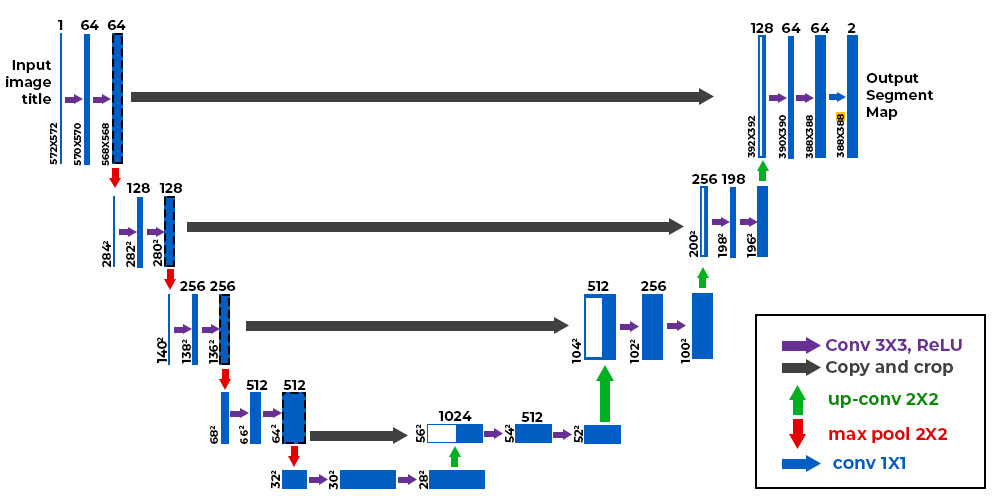


**Figure 8.** Architecture of Convolutional Neural Network

Figure 8 displays the architecture of the convolutional neural networks. Deep learning networks are fundamental to deep learning. These systems, also called deep neural networks or artificial neural networks, learn complex data patterns. They use this learning for predictions and classifications. Many types of deep learning networks exist. These include feed forward, convolutional, and recurrent networks. Others are long short-term memory, auto encoders, and transformers [72]. Generative adversarial networks are also used. Deep belief networks and graph neural networks are other examples. Each network type suits specific tasks and data. They handle image sorting, language tasks, and data sequences. This section outlines prominent deep learning networks [73]. Convolutional neural networks receive detailed attention. Their use in brain tumor classification is significant. CNNs are widely used neural networks [74]. They offer many advantages over other types.

## U-Net

The U-Net architecture, introduced in 2015, has transformed deep learning. It dominated the 2015 ISBI cell tracking challenge. U-Net excels at segmenting neuron structures in microscopy images [75]. U-Net is a convolutional neural network with an encoder-decoder architecture. It is widely used for biomedical segmentation tasks [76][77]. This architecture can process large images efficiently on modern GPUs. Its great success has led to many variations. These include LadderNet, attention-based U-Net, R2-UNet, and U-Net with residual or dense blocks [78].

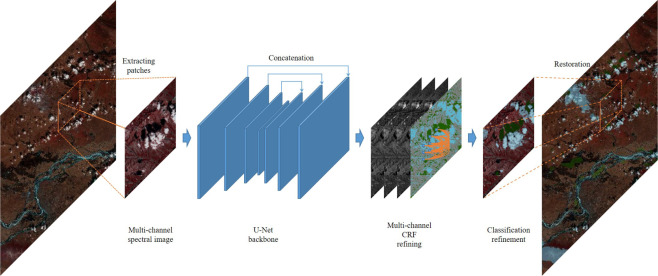
While U-Net is a major deep learning advance, understanding prior methods is important. The sliding window approach, for example, won the 2012 ISBI EM segmentation challenge. It created many sample patches from the training data. This method treated each pixel's class label as a separate task. It used a local patch around each pixel for prediction. The sliding window architecture easily adapted to different datasets. However, the sliding window approach had two key flaws U-Net fixed. First, processing pixels individually caused patch overlap and data redundancy. Second, training was very slow and resource-intensive. These issues impacted its real-world use. U-Net elegantly resolves these problems [79]. It uses fully convolutional networks[80]. U-Net captures both context and location information. Its design enables this. The architecture uses shrinking layers followed by upsampling [81]. This produces high-resolution output from input images [82].

**Figure 9.** Basic architecture of U-Net.

The U-Net architecture originates from a research paper. Its design forms a distinct 'U' shape [83]; this visual characteristic gives the network its name. The architecture employs only convolutional layers, and it avoids dense or flattened layers. The structure includes a contracting path. This is followed by an expanding path. An input image enters the U-Net and passes through convolutional layers. These use the ReLU activation function. Image dimensions decrease, for instance, from 572x572. This reduction stems from unpadded convolutions. Such convolutions limit the overall image size. The network features encoder blocks on the left. Decoder blocks are on the right. Encoder blocks reduce image size using max-pooling. Strides of 2 are used. Convolutional layers repeat here. They also increase the filter count. Decoder blocks perform upsampling. Filter counts decrease in these layers. Skip connections link earlier outputs to decoder layers. These skip connections are crucial. They preserve information loss. This improves overall output values. This method also aids faster model training [84]. The final convolution block has two layers. A final output layer follows. This layer uses a filter of 2. It displays the segmentation result. This final layer can adapt to project needs. The fundamental architecture of a U-Net segmentation model is illustrated in figure 9. U-Net architectures address the limitations of CNN in capturing long-range dependencies and global context in images [85].

### Patch Wise Sampling

Training a U-Net architecture on large-scale images frequently necessitates a patch-wise sampling strategy [86]. This technique divides expansive images into smaller, manageable sections, often referred to as patches. These patches are intentionally created with some overlap. This segmentation allows the U-Net, which has finite memory constraints, to process images far exceeding its native capacity. The workflow involves extracting these patches, feeding them individually into the U-Net for processing, and then aggregating the resulting predictions. The strategic inclusion of overlapping regions between adjacent patches is crucial for mitigating artifacts that can arise at the boundaries of individual image segments. Without overlap, the U-Net might produce inconsistent or inaccurate predictions at the edges where patches meet. Memory limitations are a common challenge for U-Nets when dealing with high-resolution imagery. Patch sampling effectively circumvents this bottleneck by breaking down the computational load into smaller, more digestible units. This granular approach enables the U-Net to better focus on local contextual information within each patch, leading to more precise segmentation outcomes. Consequently, this method broadens the applicability of the U-Net, making it suitable for a wider array of image sizes during both the training phase and deployment.



**Figure 10.** Visual representation of the process involved in patch wise sampling.

Figure 10 provides a detailed representation of the patch wise sampling process. The process begins with the systematic division of the input image. This splitting generates a collection of smaller image tiles, typically overlapping. Each of these extracted patches is then passed through the U-Net model for segmentation. After the U-Net processes all the patches, the individual segmentation maps are reassembled to reconstruct the complete prediction for the original, larger image. The overlap incorporated into the patch extraction plays a vital role in ensuring continuity and accuracy across the entire image. It smooths out the transitions between the processed patches, preventing jarring visual discontinuities. Techniques for blending the results from overlapping regions are employed to create a cohesive and seamless final prediction map. The selection of an appropriate patch size presents a trade-off between segmentation accuracy and computational cost. Smaller patches can sometimes yield higher accuracy by providing the U-Net with more focused local detail. However, this often comes at the expense of increased computational demands and can limit the U-Net's ability to learn from broader spatial context. Conversely, larger patches reduce computational overhead but may dilute local details. Careful management of patch overlap and effective blending strategies are paramount for minimizing boundary artifacts and achieving high-quality segmentation across the entire image [87].

### Light weighted 3D U-Nets

Lightweight 3D U-Nets reduce computations and parameters [88]. They enable the maintenance of accuracy for tasks such as brain tumour and seismic fault segmentation. Techniques include depth wise separable convolutions. Parallel convolutions and attention mechanisms are also employed. LATUP-N46

et uses parallel convolutions and attention for efficiency. 3D-Inception-UNet uses Inception blocks to cut parameters. TDPC-Net uses decoupled convolutions and attention for multi-scale features. Lightweight 3D U-Net splits standard convolution. It uses depthwise and pointwise stages that cut parameters and computation. Parallel convolutional blocks use different kernel sizes, and they capture multi-scale features, which avoids redundancy. Attention mechanisms focus on important data parts. They boost feature learning and also improve segmentation accuracy. Hierarchical decoupled convolutions break down convolutions. It uses smaller, efficient units which cut complexity and enable the capture of features at many scales. Balancing the network depth and width is crucial for attaining desired performance with fewer resources. Examples of lightweight 3D U-Nets include LATUP-Net, 3D-Inception-UNet, and TDPC-Net. The LATUP-Net model uses parallel convolutions. It adds an attention mechanism, cuts parameters and also matches state-of-the-art models. The 3D-Inception-UNet model adds inception blocks. It significantly reduces parameters and is also used for seismic fault segmentation. TDPC-Net uses decoupled convolutions. It uses a 3D attention mechanism. This model can handle multi-scale features, and it also uses few parameters and operations. Such lightweight 3D models are useful in medical imaging [89]. They aid in brain tumour segmentation, segmenting volumetric data and also segmenting faults in seismic data.

### Sparse-attention networks

Sparse attention integration within the U-Net architecture significantly enhances its operational efficiency and predictive accuracy by strategically reducing computational overhead. This method prioritises the analysis of critical data points rather than processing the entirety of the input. This focused approach directly translates into improved performance metrics, such as increased accuracy and faster execution times, which are vital for complex image manipulation tasks like restoration or precise object segmentation. The underlying principle is to disregard data deemed less relevant or informative. Several techniques facilitate this data prioritisation. Methods exist for intelligently grouping related data elements or effectively filtering out noise and superfluous information. In contrast, traditional full attention mechanisms demand substantial computing resources, often becoming a bottleneck for real-time applications. Sparse attention circumvents this limitation by substantially lowering computational demands. This allows the model to more effectively identify and utilise long-range dependencies within the data and subtle intricate details. This attention mechanism acts by highlighting salient regions of interest while concurrently suppressing or ignoring less important areas. The outcome is a marked improvement in the quality of results for tasks such as image segmentation and image reconstruction. The inherent design of the U-Net, which captures features at multiple resolutions, works synergistically with sparse attention. Within this framework, sparse attention pinpoints and gives preference to valuable data. This cultivates a refined, focused data representation. It effectively scrubs irrelevant or misleading data [90]. This capability preserves the model's aptitude for understanding broad contextual relationships across distant parts of an image. This is particularly beneficial for accurate image segmentation and effective image repair. The network can then zero in on critical object boundaries and fine-grained details. It also enables robust correction of degraded image sections. This allows for a deeper comprehension of pixel interrelationships throughout the entire image. Notable examples of U-Net architectures employing sparse attention include Spa-Former, DUSA-UNet, and the Dual Attention Residual U-Net. Spa-Former integrates sparse attention specifically for the purpose of image restoration tasks. DUSA-UNet applies attention mechanisms to enhance the analysis of road network data. The Dual Attention Residual U-Net leverages sparse attention to capture important contextual information, deliberately ignoring data that offers a weak or negligible signal.

# research challenges

Optimising deep learning for small organ segmentation using medical images confronts several challenges. These challenges include data scarcity, the distinctive traits of small organs, and the built-in limitations of deep learning networks. Small organs pose a unique challenge due to their size, unclear edges, and varied forms. Creating useful medical imaging data is a complex task. Patient privacy rules add on to the complexity and cost. Labelling images is a time-consuming task and needs a lot of funding; this limits the data availability [91]. The problem is worse in the case of small organs or body parts. This causes major class imbalance. The ROI is a tiny part of the medical image where most of the image shows healthy tissue. This skews the data; hence, AI models trained on this data often fail. The models might just guess the common case. This leads to many missed diagnoses, making the AI models more unreliable. Medical images are often gathered from various sources. Each source might use a different scanner. Scanners work and save images in various ways that make these images look different. These variations affect the clarity of the images and the detailings present in them. This is called domain shift. Models learn from images in the training dataset. If new images are too different, the model fails to segment them precisely [92]. A model trained on one hospital's X-rays might fail with another hospital's. One hospital might use a scanner with less power when the other might use a scanner that provides more detailed scans. The pictures are not identical, and this difference affects how well doctors can use AI. A model that spots problems in one set of images may miss them in another. It might also assume normal things are abnormalities. This makes AI tools hard to trust. AI needs to work well with different kinds of images. Without fixing domain shift, AI tools are not reliable for real-world medical use. Manual outlining of small anatomical parts is a time-consuming and unreliable task. Experts disagree on exact boundaries. This leads to inconsistent data, and noisy labels affect the performance of deep learning models. Models require clear patterns to learn, as conflicting data makes them less dependable. The optic nerve head exemplifies this issue. It has a small and intricate structure. Minor variations in manual segmentation create noise. Models trained on this data learn mixed signals, which degrade their quality and performance. Accurate and repeatable segmentation is essential. Manual methods fail to provide uniform data. AI requires precise data to function correctly.

Precisely segmenting small organs within medical images is a challenging task. Existing deep learning models often use large receptive fields, which are effective for larger structures. But it can blur or entirely miss the subtle, fine details or characteristics of diminutive organs. Small organs exhibit low contrast against their surrounding tissues. This subtle difference makes their edges unclear, creating ambiguity. This visual uncertainty significantly influences the precision and reliability of automated segmentation models. Without sharp, well-defined boundaries, deep learning algorithms struggle to confidently define the organ's precise location. The underlying anatomical variations observed across various patients exacerbate these challenges. Even common small organs present in the human body can vary greatly in their size, shape, and precise location from person to person. This inherent variability demands models that are precise. Furthermore, pathological conditions obscure these delicate structures. The physical presence of more prominent organs also poses a significant obstacle. These prominent structures can easily occlude smaller, adjacent organs. This spatial relationship makes segmenting the small organ a complex task. These combined challenges necessitate the development of a more robust and sophisticated segmentation model that must be capable of capturing minute details and adapting to significant anatomical variability and contextual clutter to achieve reliable performance in medical environments.

# evaluation metrics

Evaluation metrics are vital for assessing the performance of machine learning models. Metrics measure how well models meet their objectives. This is true for classification, prediction, and clustering tasks. For rare diseases, high accuracy is not always sufficient; recall is more crucial. It finds all actual cases that prevent wrong diagnoses. Error reduction is essential in financial models. Metrics like MAE and RMSE measure prediction accuracy. MAE averages absolute differences, and RMSE penalises larger errors more. The choice depends on risk. Common metrics covered in this article are Dice, IoU (overlap accuracy), Hausdorff or average surface distances (border precision), SSEGEP for small structures and MISm for edge cases where regions are very small or nonexistent.

The Dice coefficient measures how well two data sets overlap. It is mostly employed in machine learning for image segmentation. It compares a model's prediction to the actual answer. A higher score represents a better match. It is calculated as twice the overlap size divided by the total size of both sets. A score of 1.0 indicates a perfect match, whereas the score 0.0 represents no overlap. This score shows how accurately a model captures an object's shape and edges. It is a simple metric used to judge segmentation quality [93]. IoU (Intersection over Union) quantifies the degree to which a predicted box overlaps a ground truth box. It's a key metric employed in most of the object detection and image segmentation tasks. It is calculated by dividing the area of overlap between the two boxes by the area of their union. A higher IoU score represents a better prediction and vice versa. This metric aids in understanding the precision of object localisation.

The separation between two sets of points is quantified using the Hausdorff distance. It helps with shape comparison in machine learning. The Hausdorff distance finds the greatest distance from any point in one shape to the closest point in the other shape. It helps us see how well predicted shapes match original shapes. A smaller Hausdorff distance means a better match. It works well for tasks like image segmentation or comparing object boundaries. This metric is sensitive to outliers. SSEGEP is a metric employed for evaluating how well a machine learning model predicts trends. It calculates the accuracy of these predictions over time. Better prediction performance is indicated by a lower SSEGEP score. It's majorly used for models tracking stock prices and weather patterns, as it helps users choose the best forecasting model. MISm measures the performance of a machine learning model. It measures the Mean Inverse Score Matching. It represents how well predictions align with actual values. A lower MISm score signals better model precision. Its major applications are in regression tasks. It also penalises major errors more heavily. A well-chosen metric ensures the reliability of the model.

# FUTURE DIRECTIONS

Model training is hindered by the limited availability of labelled data. Advanced learning approaches are capable of solving this. Semi-supervised, few-shot, and partial supervision methods work well with few labels. Organ segmentation research is the best example that proves their value. These methods can be applied to enhance the model outcomes. Self-supervised learning also can be employed. Models are pre-trained on unlabelled data that teaches them patterns and features. Generative models like GANs play a vital role. GANs create fake data, which is helpful when available labelled data is scarce. It's useful for small or rare organs. Realistic fake data overcomes this scarcity [91][94]. It helps models to train over more variety, which helps them work better. Domain adaptation and federated learning can be employed to handle scanner or site differences. Federated learning helps in private multi-site collaboration. This helps in improving outcomes from varied data. Adversarial training, like IMA, guards against noisy inputs. It protects against attacks without affecting segmentation. MedRDF provides strong inference. It creates and combines noisy versions during use that build system stability. Bayesian networks or ensembles calculate model uncertainty. Uncertain results get flagged for doctors, which fosters trust. It is crucial for risky medical cases. Integrating domain knowledge into models includes prior knowledge, like anatomy rules or maps from experts [95]. It guides models to key areas and also helps in times of data scarcity. AI methods can be employed to explain decisions. Doctors require clear outputs, as they must see where the AI is unsure or why it made a prediction. It also ensures safer usage of such automated segmentation models. Creating compact models ensures quick analysis. This is a key application for clinical settings and edge tools. Add segmentation tools such as IoMT to health systems or phone apps. It helps to monitor patients continuously in real time and also helps in care settings [96]. Linking segmentation results with image-genomics insights to find non-invasive markers provides biological details beyond basic shape.

# conclusion

This study offers a thorough review of the advancing deep learning networks that are being utilised for automated medical image segmentation. It majorly focuses on small organ segmentation using images or scans obtained from distinct medical imaging sources. CT, MRI, PET, and ultrasound are among the most commonly used imaging modalities in clinical environments. These are certain non-invasive techniques employed to view the inner structures of the human body without any incision or cuts. CT and MRI are among the most prevalently used imaging techniques that play a vital role in medical environments. Such techniques, along with contrast agents, are used to get a clear view of the smaller organs present in our human body. Conventional image classification techniques include manual segmentation of medical images gathered from such imaging modalities. Manual segmentation of scans is a complex and time-consuming task. It requires an expert to be present for segmentation tasks. Deep learning models can be incorporated in segmentation tasks to optimise the outcomes. Deep learning networks such as CNN provide various architectures such as UNet. These networks can be used to develop models that can be trained with a dataset of medical images. The trained model can then be used for making predictions with unseen data. Such algorithms can be implemented in clinical environments. Such automated segmentation systems can offer various benefits over the traditional methods. This study provides an overview of the majorly used imaging modalities. It also provides a detailed overview of deep learning and its networks. The study majorly focuses on the UNet architecture and its application in small organ segmentation. Though incorporating deep learning algorithms in the medical field provides clinicians with diverse advantages, there also exist risks, such as wrong diagnoses. Data scarcity can obscure the training process of the models that, in return, might affect the model performance. It is crucial to choose proper evaluation metrics to assess the model. The final outcomes of these metrics indicate the precision and reliability of these models. The study also highlights areas that can be optimised in the future for better results. It also suggests certain areas for future enhancements which can improve the model performance and its reliability. The review also examined the previous works and found the research gaps. Thus, it provides us with a path through which we can further optimise the performance of deep learning models in medical image segmentation.

# REFERENCES

1. A. Pratondo, C. K. Chui, and S. H. Ong, “Robust edge-stop functions for edge-based active contour models in medical image segmentation,” *IEEE Signal Process Lett*, vol. 23, no. 2, pp. 222–226, Feb. 2016, doi: 10.1109/LSP.2015.2508039.
2. P. Hu, J. Li, Y. Wang, and D. Kong, “Deep learning in abdominal organ segmentation: A review,” *Intelligent Oncology*, Jul. 2025, doi: 10.1016/j.intonc.2025.06.007.
3. R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, “Medical Image Segmentation Using Deep Learning: A Survey,” Dec. 2021, doi: 10.1049/ipr2.12419.
4. Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, “A review of deep learning based methods for medical image multi-organ segmentation,” May 01, 2021, *Associazione Italiana di Fisica Medica*. doi: 10.1016/j.ejmp.2021.05.003.
5. S. Seidlitz *et al.*, “Robust deep learning-based semantic organ segmentation in hyperspectral images,” *Med Image Anal*, vol. 80, Aug. 2022, doi: 10.1016/j.media.2022.102488.
6. X. Liu, L. Qu, Z. Xie, J. Zhao, Y. Shi, and Z. Song, “Towards more precise automatic analysis: a systematic review of deep learning-based multi-organ segmentation,” Dec. 01, 2024, *BioMed Central Ltd*. doi: 10.1186/s12938-024-01238-8.
7. Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, and J. Wang, “Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods,” May 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/tomography11050052.
8. Y. Chen, L. Xing, L. Yu, H. P. Bagshaw, M. K. Buyyounouski, and B. Han, “Automatic intraprostatic lesion segmentation in multiparametric magnetic resonance images with proposed multiple branch UNet,” *Med Phys*, vol. 47, no. 12, pp. 6421–6429, Dec. 2020, doi: 10.1002/mp.14517.
9. F. Wahid, Y. Ma, D. Khan, M. Aamir, and S. U. K. Bukhari, “Biomedical Image Segmentation: A Systematic Literature Review of Deep Learning Based Object Detection Methods,” Aug. 2024, [Online]. Available: http://arxiv.org/abs/2408.03393
10. N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, “Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation,” *Med Image Anal*, vol. 63, p. 101693, Jul. 2020, doi: 10.1016/j.media.2020.101693.
11. X. Xie, J. Niu, X. Liu, Z. Chen, S. Tang, and S. Yu, “A survey on incorporating domain knowledge into deep learning for medical image analysis,” *Med Image Anal*, vol. 69, p. 101985, Apr. 2021, doi: 10.1016/j.media.2021.101985.
12. F. Zheng *et al.*, “AFFSegNet: Adaptive Feature Fusion Segmentation Network for Microtumors and Multi-Organ Segmentation,” Dec. 2024, [Online]. Available: http://arxiv.org/abs/2409.07779
13. Y. Chen *et al.*, “Automated Muscle and Fat Segmentation in Computed Tomography for Comprehensive Body Composition Analysis,” Aug. 2025, [Online]. Available: http://arxiv.org/abs/2502.09779
14. M. Antonelli *et al.*, “The Medical Segmentation Decathlon,” *Nat Commun*, vol. 13, no. 1, Dec. 2022, doi: 10.1038/s41467-022-30695-9.
15. D. Karimi, H. Dou, S. K. Warfield, and A. Gholipour, “Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis,” *Med Image Anal*, vol. 65, Oct. 2020, doi: 10.1016/j.media.2020.101759.
16. Y. Fu *et al.*, “More unlabelled data or label more data? A study on semi-supervised laparoscopic image segmentation”, doi: 10.48550/arXiv.1908.08035.
17. S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, “The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE Computer Society, Aug. 2017, pp. 1175–1183. doi: 10.1109/CVPRW.2017.156.
18. M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, “Modern Trends in Hyperspectral Image Analysis: A Review,” Mar. 12, 2018, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2018.2812999.
19. X. Chen *et al.*, “A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy,” *Radiotherapy and Oncology*, vol. 160, pp. 175–184, Jul. 2021, doi: 10.1016/j.radonc.2021.04.019.
20. A. E. Kavur *et al.*, “CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation,” *Med Image Anal*, vol. 69, p. 101950, Apr. 2021, doi: 10.1016/j.media.2020.101950.
21. C. Chen, Q. Dou, H. Chen, J. Qin, and P. A. Heng, “Unsupervised Bidirectional Cross-Modality Adaptation via Deeply Synergistic Image and Feature Alignment for Medical Image Segmentation,” *IEEE Trans Med Imaging*, vol. 39, no. 7, pp. 2494–2505, Jul. 2020, doi: 10.1109/TMI.2020.2972701.
22. W. Yuan, J. Wei, J. Wang, Q. Ma, and T. Tasdizen, “Unified generative adversarial networks for multimodal segmentation from unpaired 3D medical images,” *Med Image Anal*, vol. 64, p. 101731, Aug. 2020, doi: 10.1016/j.media.2020.101731.
23. X. Dai *et al.*, “Synthetic CT‐aided multiorgan segmentation for CBCT‐guided adaptive pancreatic radiotherapy,” *Med Phys*, vol. 48, no. 11, pp. 7063–7073, Nov. 2021, doi: 10.1002/mp.15264.
24. L. Huang, M. Li, S. Gou, X. Zhang, and K. Jiang, “Automated Segmentation Method for Low Field 3D Stomach MRI Using Transferred Learning Image Enhancement Network,” *Biomed Res Int*, vol. 2021, 2021, doi: 10.1155/2021/6679603.
25. S. Wang *et al.*, “Annotation-efficient deep learning for automatic medical image segmentation,” *Nat Commun*, vol. 12, no. 1, Dec. 2021, doi: 10.1038/s41467-021-26216-9.
26. A. Bitarafan, M. Nikdan, and M. S. Baghshah, “3D Image Segmentation With Sparse Annotation by Self-Training and Internal Registration,” *IEEE J Biomed Health Inform*, vol. 25, no. 7, pp. 2665–2672, Jul. 2021, doi: 10.1109/JBHI.2020.3038847.
27. F. Cervantes-Sanchez *et al.*, “Automatic tissue segmentation of hyperspectral images in liver and head neck surgeries using machine learning,” *Artificial Intelligence Surgery*, 2021, doi: 10.20517/ais.2021.05.
28. M. Dietrich *et al.*, “Machine learning-based analysis of hyperspectral images for automated sepsis diagnosis,” Jun. 2021, [Online]. Available: http://arxiv.org/abs/2106.08445
29. H. Fabelo *et al.*, “Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations,” *PLoS One*, vol. 13, no. 3, Mar. 2018, doi: 10.1371/journal.pone.0193721.
30. J. Gong *et al.*, “Using deep learning to identify the recurrent laryngeal nerve during thyroidectomy,” *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-93202-y.
31. P. M. Scheikl *et al.*, “Deep learning for semantic segmentation of organs and tissues in laparoscopic surgery,” in *Current Directions in Biomedical Engineering*, Walter de Gruyter GmbH, May 2020. doi: 10.1515/cdbme-2020-0016.
32. Q. Zhu *et al.*, “A Spectral-Spatial-Dependent Global Learning Framework for Insufficient and Imbalanced Hyperspectral Image Classification,” *IEEE Trans Cybern*, vol. 52, no. 11, pp. 11709–11723, Nov. 2022, doi: 10.1109/TCYB.2021.3070577.
33. X. Zhou, T. Ito, R. Takayama, S. Wang, T. Hara, and H. Fujita, “Three-Dimensional CT Image Segmentation by Combining 2D Fully Convolutional Network with 3D Majority Voting,” 2016, pp. 111–120. doi: 10.1007/978-3-319-46976-8\_12.
34. K. Men, J. Dai, and Y. Li, “Automatic segmentation of the clinical target volume and organs at risk in the planning <scp>CT</scp> for rectal cancer using deep dilated convolutional neural networks,” *Med Phys*, vol. 44, no. 12, pp. 6377–6389, Dec. 2017, doi: 10.1002/mp.12602.
35. V. Yeghiazaryan and I. Voiculescu, “Family of boundary overlap metrics for the evaluation of medical image segmentation,” *Journal of Medical Imaging*, vol. 5, no. 01, p. 1, Feb. 2018, doi: 10.1117/1.JMI.5.1.015006.
36. W. Zhu *et al.*, “AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy,” *Med Phys*, vol. 46, no. 2, pp. 576–589, Feb. 2019, doi: 10.1002/mp.13300.
37. S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, “Medical Image Analysis using Convolutional Neural Networks: A Review,” Nov. 01, 2018, *Springer New York LLC*. doi: 10.1007/s10916-018-1088-1.
38. R. Zhang, G. Li, Z. Li, S. Cui, D. Qian, and Y. Yu, “Adaptive Context Selection for Polyp Segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2020, pp. 253–262. doi: 10.1007/978-3-030-59725-2\_25.
39. F. Hutter, L. Kotthoff, and J. Vanschoren, “The Springer Series on Challenges in Machine Learning Automated Machine Learning Methods, Systems, Challenges.” [Online]. Available: http://www.springer.com/series/15602
40. J. Chen *et al.*, “TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation,” Feb. 2021, [Online]. Available: http://arxiv.org/abs/2102.04306
41. R. Poplin *et al.*, “Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning,” *Nat Biomed Eng*, vol. 2, no. 3, pp. 158–164, Feb. 2018, doi: 10.1038/s41551-018-0195-0.
42. A. Reinke *et al.*, “Common Limitations of Image Processing Metrics: A Picture Story,” Dec. 2023, [Online]. Available: http://arxiv.org/abs/2104.05642
43. F. Isensee *et al.*, “nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation”, doi: 10.48550/arXiv.1809.10486.
44. X. Zhu, Z. Cheng, S. Wang, X. Chen, and G. Lu, “Coronary angiography image segmentation based on PSPNet,” *Comput Methods Programs Biomed*, vol. 200, p. 105897, Mar. 2021, doi: 10.1016/j.cmpb.2020.105897.
45. B. Huang, F. Yang, M. Yin, X. Mo, and C. Zhong, “A Review of Multimodal Medical Image Fusion Techniques,” 2020, *Hindawi Limited*. doi: 10.1155/2020/8279342.
46. X. Chen and L. Pan, “A Survey of Graph Cuts/Graph Search Based Medical Image Segmentation,” *IEEE Rev Biomed Eng*, vol. 11, pp. 112–124, 2018, doi: 10.1109/RBME.2018.2798701.
47. A. M. Saranathan and M. Parente, “Threshold based segmentation method for hyperspectral images,” in *2013 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, IEEE, Jun. 2013, pp. 1–4. doi: 10.1109/WHISPERS.2013.8080656.
48. H. Tang *et al.*, “Clinically applicable deep learning framework for organs at risk delineation in CT images,” *Nat Mach Intell*, vol. 1, no. 10, pp. 480–491, Sep. 2019, doi: 10.1038/s42256-019-0099-z.
49. D. Shen, G. Wu, and H.-I. Suk, “Deep Learning in Medical Image Analysis,” *Annu Rev Biomed Eng*, vol. 19, no. 1, pp. 221–248, Jun. 2017, doi: 10.1146/annurev-bioeng-071516-044442.
50. J. Yang *et al.*, “Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017,” *Med Phys*, vol. 45, no. 10, pp. 4568–4581, Oct. 2018, doi: 10.1002/mp.13141.
51. C. Kaushal, M. K. Islam, S. A. Althubiti, F. Alenezi, and R. F. Mansour, “A Framework for Interactive Medical Image Segmentation Using Optimized Swarm Intelligence with Convolutional Neural Networks,” *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/7935346.
52. I. Rizwan I Haque and J. Neubert, “Deep learning approaches to biomedical image segmentation,” Jan. 01, 2020, *Elsevier Ltd*. doi: 10.1016/j.imu.2020.100297.
53. S. Bakas, A. Jakab, and S. Bauer Helbling, “Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge.” [Online]. Available: https://www.researchgate.net/publication/331576745
54. A. G. Roy, N. Navab, and C. Wachinger, “QuickNAT: Segmenting MRI Neuroanatomy in 20 seconds”, doi: 10.48550/arXiv.1801.04161.
55. M. Khened, V. A. Kollerathu, and G. Krishnamurthi, “Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers,” *Med Image Anal*, vol. 51, pp. 21–45, Jan. 2019, doi: 10.1016/j.media.2018.10.004.
56. L. Maier-Hein *et al.*, “BIAS: Transparent reporting of biomedical image analysis challenges,” *Med Image Anal*, vol. 66, Dec. 2020, doi: 10.1016/j.media.2020.101796.
57. J. C. Á. Cerón, G. O. Ruiz, L. Chang, and S. Ali, “Real-time instance segmentation of surgical instruments using attention and multi-scale feature fusion,” *Med Image Anal*, vol. 81, Oct. 2022, doi: 10.1016/j.media.2022.102569.
58. J. van der Veen *et al.*, “Benefits of deep learning for delineation of organs at risk in head and neck cancer,” *Radiotherapy and Oncology*, vol. 138, pp. 68–74, Sep. 2019, doi: 10.1016/j.radonc.2019.05.010.
59. Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, “Deep learning in medical image registration: a review,” *Phys Med Biol*, vol. 65, no. 20, p. 20TR01, Oct. 2020, doi: 10.1088/1361-6560/ab843e.
60. S. Calderon-Ramirez, S. Yang, and D. Elizondo, “Semisupervised Deep Learning for Image Classification with Distribution Mismatch: A Survey,” *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 6, pp. 1015–1029, Dec. 2022, doi: 10.1109/TAI.2022.3196326.
61. A. Esteva *et al.*, “A guide to deep learning in healthcare,” Jan. 01, 2019, *Nature Publishing Group*. doi: 10.1038/s41591-018-0316-z.
62. J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, “Segment anything in medical images,” *Nat Commun*, vol. 15, no. 1, Dec. 2024, doi: 10.1038/s41467-024-44824-z.
63. W. Lei, X. Wei, X. Zhang, K. Li, and S. Zhang, “MedLSAM: Localize and Segment Anything Model for 3D CT Images,” Oct. 2024, [Online]. Available: http://arxiv.org/abs/2306.14752
64. A. Hatamizadeh *et al.*, “UNETR: Transformers for 3D Medical Image Segmentation,” Oct. 2021, [Online]. Available: http://arxiv.org/abs/2103.10504
65. Y. Alzahrani and B. Boufama, “Biomedical Image Segmentation: A Survey,” *SN Comput Sci*, vol. 2, no. 4, p. 310, Jul. 2021, doi: 10.1007/s42979-021-00704-7.
66. S. Cai, Y. Tian, H. Lui, H. Zeng, Y. Wu, and G. Chen, “Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network,” *Quant Imaging Med Surg*, vol. 10, no. 6, pp. 1275–1285, Jun. 2020, doi: 10.21037/qims-19-1090.
67. G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” Dec. 01, 2017, *Elsevier B.V.* doi: 10.1016/j.media.2017.07.005.
68. G. Wang *et al.*, “Interactive Medical Image Segmentation Using Deep Learning with Image-Specific Fine Tuning,” *IEEE Trans Med Imaging*, vol. 37, no. 7, pp. 1562–1573, Jul. 2018, doi: 10.1109/TMI.2018.2791721.
69. T. Zhou, S. Ruan, and S. Canu, “A review: Deep learning for medical image segmentation using multi-modality fusion,” *Array*, vol. 3–4, Sep. 2019, doi: 10.1016/j.array.2019.100004.
70. D. Nie *et al.*, “Medical Image Synthesis with Deep Convolutional Adversarial Networks,” *IEEE Trans Biomed Eng*, vol. 65, no. 12, pp. 2720–2730, Dec. 2018, doi: 10.1109/TBME.2018.2814538.
71. J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks,” in *Proceedings of the IEEE International Conference on Computer Vision*, Institute of Electrical and Electronics Engineers Inc., Dec. 2017, pp. 2242–2251. doi: 10.1109/ICCV.2017.244.
72. H. Wu *et al.*, “CvT: Introducing Convolutions to Vision Transformers,” Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.15808
73. Q. Dou, C. Chen, H. Chen, H. Kong, C. Ouyang, and P.-A. Heng, “Unsupervised Cross-Modality Domain Adaptation of ConvNets for Biomedical Image Segmentations with Adversarial Loss,” 2017, doi: 10.48550/arXiv.1804.10916.
74. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
75. J. Ma, “Cutting-edge 3D Medical Image Segmentation Methods in 2020: Are Happy Families All Alike?,” Jan. 2021, [Online]. Available: http://arxiv.org/abs/2101.00232
76. M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, “Recurrent residual U-Net for medical image segmentation,” *Journal of Medical Imaging*, vol. 6, no. 01, p. 1, Mar. 2019, doi: 10.1117/1.JMI.6.1.014006.
77. Y. Zhang, S. Miao, T. Mansi, and R. Liao, “Task Driven Generative Modeling for Unsupervised Domain Adaptation: Application to X-ray Image Segmentation,” 2018, pp. 599–607. doi: 10.1007/978-3-030-00934-2\_67.
78. Y. He, D. Yang, H. Roth, C. Zhao, and D. Xu, “DiNTS: Differentiable Neural Network Topology Search for 3D Medical Image Segmentation”, doi: 10.48550/arXiv.2103.15954.
79. N. S. Punn and S. Agarwal, “Modality specific U-Net variants for biomedical image segmentation: a survey,” *Artif Intell Rev*, vol. 55, no. 7, pp. 5845–5889, Oct. 2022, doi: 10.1007/s10462-022-10152-1.
80. X. Liu, Y. Zhang, H. Jing, L. Wang, and S. Zhao, “Ore image segmentation method using U-Net and Res\_Unet convolutional networks,” *RSC Adv*, vol. 10, no. 16, pp. 9396–9406, Mar. 2020, doi: 10.1039/c9ra05877j.
81. S. Asgari Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, “Deep semantic segmentation of natural and medical images: a review,” *Artif Intell Rev*, vol. 54, no. 1, pp. 137–178, Jan. 2021, doi: 10.1007/s10462-020-09854-1.
82. D. Dai *et al.*, “I <math altimg="si18.svg" display="inline" id="d1e3125"> <msup> <mrow/> <mrow> <mn>2</mn> </mrow> </msup> </math> U-Net: A dual-path U-Net with rich information interaction for medical image segmentation,” *Med Image Anal*, vol. 97, p. 103241, Oct. 2024, doi: 10.1016/j.media.2024.103241.
83. H. Cao *et al.*, “Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation,” 2023, pp. 205–218. doi: 10.1007/978-3-031-25066-8\_9.
84. Y. Cao, S. Liu, Y. Peng, and J. Li, “DenseUNet: densely connected UNet for electron microscopy image segmentation,” *IET Image Process*, vol. 14, no. 12, pp. 2682–2689, Oct. 2020, doi: 10.1049/iet-ipr.2019.1527.
85. H. Cao *et al.*, “Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation,” 2023, pp. 205–218. doi: 10.1007/978-3-031-25066-8\_9.
86. H. Y. Zhou *et al.*, “nnFormer: Volumetric Medical Image Segmentation via a 3D Transformer,” *IEEE Transactions on Image Processing*, vol. 32, pp. 4036–4045, 2023, doi: 10.1109/TIP.2023.3293771.
87. K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
88. R. Karthik, R. Menaka, A. Johnson, and S. Anand, “Neuroimaging and deep learning for brain stroke detection - A review of recent advancements and future prospects,” *Comput Methods Programs Biomed*, vol. 197, p. 105728, Dec. 2020, doi: 10.1016/j.cmpb.2020.105728.
89. H. Xiao, L. Li, Q. Liu, X. Zhu, and Q. Zhang, “Transformers in medical image segmentation: A review,” *Biomed Signal Process Control*, vol. 84, p. 104791, Jul. 2023, doi: 10.1016/j.bspc.2023.104791.
90. X. Zhou, R. Takayama, S. Wang, T. Hara, and H. Fujita, “Deep learning of the sectional appearances of 3D <scp>CT</scp> images for anatomical structure segmentation based on an <scp>FCN</scp> voting method,” *Med Phys*, vol. 44, no. 10, pp. 5221–5233, Oct. 2017, doi: 10.1002/mp.12480.
91. Y. Ganin *et al.*, “Domain-Adversarial Training of Neural Networks,” May 2016, [Online]. Available: http://arxiv.org/abs/1505.07818
92. E. Gibson *et al.*, “Inter-site Variability in Prostate Segmentation Accuracy Using Deep Learning,” 2018, pp. 506–514. doi: 10.1007/978-3-030-00937-3\_58.
93. S. Nikolov *et al.*, “Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study,” *J Med Internet Res*, vol. 23, no. 7, p. e26151, Jul. 2021, doi: 10.2196/26151.
94. S. Liang *et al.*, “Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning,” *Eur Radiol*, vol. 29, no. 4, pp. 1961–1967, Apr. 2019, doi: 10.1007/s00330-018-5748-9.
95. Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, and J. Wang, “Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods,” May 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/tomography11050052.
96. S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, “Image Segmentation Using Deep Learning: A Survey,” *IEEE Trans Pattern Anal Mach Intell*, pp. 1–1, 2021, doi: 10.1109/TPAMI.2021.3059968.