Deep Learning Driven Multiscale Medical Image Segmentation for Intelligent Biomedical Informatics

Vandana Malik [1] and A.J Singh [2]
[1] Himachal Pradesh University, Department of Computer Science, Shimla, India
[2] Himachal Pradesh University, Department of Computer Science, Shimla, India

Abstract

Automated image segmentation in biomedicine has evolved into a deep learning-based task. Nevertheless, the available literature does not provide an overview of deep learning techniques that covers both the clinical level (organ or lesion) and microscopic level (cell or organelle) tasks at once. In this systematic review, studies that were published between 2021 and 2026 have been summarized to (a) provide an overview on the imaging techniques, deep learning architectures, and datasets employed for macro/micro segmentation; (b) compare their performances; (c) analyze the development of foundation models and state-space architectures from 2023 onward; and (d) introduce the concept of Organ-to-Organelle framework for unified biomedical image segmentation. According to PRISMA 2020 criteria, 1,888 publications were found using bibliographic databases, preprints, and a previous systematic review. Post-duplicate elimination and after screening through titles, abstracts, and full texts, 90 primary studies were selected for this review. The U-Net architectures and their derivatives continued to dominate across macro- and microscopic applications, whereas the hybrid transformer, the state space models using the Mamba framework, and the Segment Anything Model (SAM) were introduced after 2023. The CT and MRI-based macroscopic lesion segmentation reached median DSC scores of 0.85–0.95. The microscopic organelle segmentation, specifically mitochondria, yielded F1 scores of up to 0.97, but was still limited due to lack of annotations and structures diversity. While deep learning methods are used in biomedical imaging applications at multiple scales, macro- and microscale segmentation tasks operate independently from each other. The Organ-to-Organelle approach is capable of combining the common work flow and particular difficulties, as well as the results themselves into one coherent whole.

Keywords: *Deep Learning; Medical Image Segmentation; Multiscale Imaging; Organelle Segmentation; U-Net; Systematic Review*

1. Introduction

Medical imaging forms the basis of virtually all aspects of modern-day disease diagnosis, prognosis, and treatment planning (Zhou et al., 2021). Imaging techniques like x-rays, CT scans, MRI, PET scanning, and ultrasound give organ- and tissue-based visual information, while EM, confocal imaging, and fluorescence imaging provide cellular- and subcellular-resolution imaging (Abhisheka et al., 2023). Along this continuum, image segmentation – the demarcation of a target area from its environment – is the crucial analytical process that transforms an image into a measurable quantity (Yu et al., 2023). Image segmentation done manually by radiologists, pathologists, and cell biologists is highly accurate yet extremely time-consuming and susceptible to variation between different observers. Moreover, it is becoming less suited for use in the current era due to the massive amounts of imaging information produced nowadays (Al-Dulaimi et al., 2021). This gap has widened even more with the increase in the throughput of imaging: for instance, a qualitative analysis from 2025 on the uptake of AI technology in imaging diagnosis by physicians revealed that the expected time saving was one of the top factors influencing the decision-making of doctors regarding the use of automated technologies rather than simple accuracy (Hennrich et al., 2025).

Deep learning, especially convolutional encoder-decoder frameworks, has revolutionized automatic segmentation during the last decade. The invention of U-Net (Ronneberger et al., 2015) set up the standard architecture template for encoder-decoder networks with skip connections, which is used in many biomedical image segmentation tasks (Neha et al., 2025), but recent advancements in transformer models (Chen et al., 2021; Liu et al., 2021) and prompt-based foundation models (Kirillov et al., 2023; Wu et al., 2025) allow modeling long-term context dependencies and even segmenting never-seen-before structures with a small amount of fine-tuning. Beginning in 2024, state-space selection (Mamba) networks have been a new lineage for designs, which provide transformers' long-range capabilities at close-to-linear computational cost (Wang, Zheng, et al., 2024; Dong et al., 2026). This variety of architectures comes amid an ongoing fast acceleration of clinical adoption—a review article from 2026 on AI-based diagnostic devices in radiology, pathology, and dermatology observes a trend of moving from pure research proof-of-concepts to AI algorithms built into picture archiving and communication systems (PACS) and reporting platforms, with image analysis tools remaining the largest group among AI-enabled medical devices currently under regulation (Giansanti, 2025). This finding is echoed by a related systematic review conducted in 2026 of explainable AI models in medical imaging diagnostics, which has systematically screened 95 articles based on the PRISMA framework, showing that deep learning models adjacent to classification and segmentation have increasingly become the dominant type of clinical imaging artificial intelligence models in the published literature (Ikerionwu et al., 2026).

Although this development is underway, the literature has followed two mostly separate directions. The first direction covers macro-level segmentation—organs, tumors, and lesions segmented on CT, MRI, or PET/CT scans—

and usually gives performance measures in terms of the Dice similarity coefficient (DSC) with values between 0.85 and 0.95 (Rich et al., 2023; Gao et al., 2025). A comprehensive survey conducted in 2025 about deep learning techniques used for segmentation of medical images shows that at this macro-level, the field has progressed to an extent where architecture optimization has become the major concern of research instead of feasibility; hybrid CNN-transformers, self-configuring systems, and promotable foundation models compete with each other to improve performance on established organ and lesion datasets (Gao et al., 2025). These specialties also illustrate this development: a 2025 review of the application of artificial intelligence in stroke imaging, for example, shows that automatic segmentation of the infarct core and penumbra is progressing from being used in research to being almost routinely applied in a number of stroke centers (Rondina & Nachev, 2025), demonstrating how rapidly large-scale segmentation research can find practical application at the bedside once the performance standardization is achieved. The second one, being younger compared to the first one, is concerned with micro-level segmentation—organelles of cells like mitochondria, nuclei, and cristae seen via electron or fluorescence microscopy. The recent application of techniques used for the training of generalist models on diverse datasets from mitochondria to nuclei and lipid droplets has shown that the micro-scale approach has just begun developing a reusable infrastructure similar to the one the macro-scale approach has had for many years (Bhardwaj et al., 2026).

The two streams exhibit architectural parallels but are seldom compared—both involve U-Net architectures, attention, generative augmentation, and promotable foundation models, yet the two are seldom considered alongside each other. The isolation prevents potential method transfers: methods created for solving the problem of sparse annotation in mitochondrial segmentation (e.g., self-supervised pretraining and generalist instance models) might well be applicable to rare lesion segmentation in oncology imaging and vice versa (Bhardwaj et al., 2026). This also complicates the effort of the field to make generalized statements about what methodologies are indeed effective in increasing the quality of segmentation, compared to what methodologies simply happen to correspond with one specific scale or modality. In relation to medical image analysis, where a convolutional neural network-centric review conducted in 2025 observes, for instance, that although transformer or state space alternatives are reported with increased frequency, there is still an adherence to residual and attention-assisted backbones in the context of organ-level medical image segmentation, which follows the trend of architectural conservatism observed for organelle segmentation (Mienye et al., 2025). Lacking such a comparison of the two fields in a review, these comparisons will continue to be implicit rather than practical, and scholars entering into either of these sub-fields will need to make their own link between the two. This systematic review has three research objectives and their respective research questions:

Research Objectives (RO)

RO1: To synthesize deep learning-based medical image segmentation studies across macro and micro imaging scales.
RO2: To compare architectures, datasets, and performance metrics across the macro-to-micro continuum.
RO3: To identify shared methodological challenges and solutions for developing a unified, scale-adaptive segmentation framework.

Research Questions (RQ)

RQ1: What are the key characteristics of deep learning-based segmentation across macro and micro imaging scales?
RQ2: How do architectures, datasets, and performance metrics differ across the macro-to-micro continuum?
RQ3: What shared challenges and solutions can support a unified, scale-adaptive segmentation framework?

We believe this to be the first review to conceive of DL segmentation of medical images not only as the continuum of macroscope to microscope but also as a body of literature that has been systematically screened and analyzed using the PRISMA 2020 guidelines in their entirety and to do so while explicitly taking into account the architectural and process changes that have taken place because of transformer hybrids, state space models, and promotable foundation models since 2023 and up to 2026.

The rest of the paper is organized as follows. In section 2, a literature review from the perspective of imaging modalities, architectures, and micro/macro applications is presented. The next section, Section 3, discusses the methodology used in the review process, which includes the PRISMA search methodology, the process of study selection and filtration, and the process of data extraction. Findings and benchmarking of performance at different scales are discussed in Section 4, along with the introduction of a conceptual framework named Organ to Organelle. Section 5 is the discussion related to findings; Section 6 is the conclusion and implications of the study. Section 7 represents the limitations and research scope for the future. In the last references of cited studies.

2. Literature Review

2.1 Medical Imaging Modalities Across Scales

Various imaging techniques exhibit systematic differences in terms of resolution, mechanisms of contrast generation, and the spatial scales that they can resolve. The conventional x-ray technique is a relatively inexpensive method that involves the generation of two-dimensional images of the subject that can be used for purposes of detecting fractures

and performing pulmonary examinations; however, x-rays involve the use of ionizing radiation (Irede et al., 2024). On the other hand, CT involves the generation of three-dimensional images based on ionizing radiation with greater contrast compared to conventional x-rays (Ahmad et al., 2024). MRI involves generation of high-contrast images without using ionizing radiation based on magnetic and radio frequency fields (Rai et al., 2026). Additional techniques like PET and PET/CT, which integrate metabolic information, provide high sensitivity for the detection of cancerous tissues but come with drawbacks of noise due to tracer uptake and mismatching between images (Matsubara et al., 2022). The ultrasound technique is relatively cheap and does not require radiation; however, it cannot be used to image bone and air-filled structures and thus can only be used for abdominal, pelvic, and surface scans (Abhisheka et al., 2023). Thermal (infrared) imaging has also been considered as a cheap auxiliary tool in brain tumor and breast lesion screening, generating thermographic pseudo-color images that need deep learning algorithms (Civilibal et al., 2023; Almomany et al., 2026).

In terms of the microscopic world, electron microscopy, including both transmission and scanning types, and fluorescence/confocal microscopy help to identify cellular organelles, with transmission EM having higher resolution for structures like mitochondrial cristae (Dey, 2022; Neikirk et al., 2023). In contrast to transmission electron microscopy, scanning EM is able to visualize the 3D structure of the surface, whereas fluorescence and confocal microscopy make live-cell multichannel imaging possible but cause photobleaching and reduced depth of penetration (Dey, 2022). Fundus photography is a modality used for the imaging of the retina that falls somewhere in the middle since it provides image resolution at a level above the organ-level imaging but below the subcellular imaging (Gonçalves et al., 2024). This is particularly useful in the case of diagnosing diabetic retinopathy. Table 1 presents the modalities in question.

Table 1. Imaging Modalities Across the Macro-to-Micro Continuum

Modality	Typical Scale	Representative Targets	Key Constraint	Source
X-ray	Macro	Fractures, pulmonary infiltrates	Ionizing radiation; 2D projection only	Irede et al. (2024)
CT	Macro	Bone lesions, tumors, organ morphology	Radiation dose; poor soft-tissue contrast	Ahmad et al. (2024); Rich et al. (2023)
MRI	Macro	Soft tissue, marrow, brain, and spine lesions	Cost, acquisition time, motion artifact	Rai et al. (2026)
PET / PET-CT	Macro	Metabolically active malignancies	Tracer noise; spatial misalignment	Matsubara et al. (2022); Rich et al. (2023)
Ultrasound	Macro	Abdominal and pelvic organs	Poor penetration through bone/air	Abhisheka et al. (2023)
Thermal / infrared	Macro (adjunct)	Brain tumors, breast lesions	Low native resolution; ambient noise	Civilibal et al. (2023); Almomany et al. (2026)
Fundus photography	Macro–meso	Retinal vessels, diabetic lesions	Illumination variability	Gonçalves et al. (2024)
Fluorescence / confocal microscopy	Micro	Cell nuclei, cytoplasm, organelles	Photobleaching; limited depth	Dey (2022)
Electron microscopy (TEM/SEM)	Micro	Mitochondria, cristae, vesicles, nuclei, lipid droplets	Severe annotation scarcity; small field of view	Neikirk et al. (2023); Conrad & Narayan (2023); Bhardwaj et al. (2026)

Source: Author's own compilation

2.2 Deep Learning Architectures for Segmentation

The encoder-decoder-skip connection architecture of U-Net (Ronneberger et al., 2015) is still the most popularly adopted architecture at both the macro- and microscopic scale levels because of its capability of retaining spatial information along with the contextual hierarchy of features; a 2025 analytical survey of U-Net literature shows that its variants still outperform all other architectures in total in terms of biomedical segmentation tasks (Neha et al., 2025). However, before UNet was developed, there were SegNet (Badrinarayanan et al., 2017) and fully convolutional networks (Long et al., 2015), which performed better than UNet in terms of detecting large objects but failed in the case of small anatomy or subcellular objects due to poor feature extraction. With DeepLab using dilated convolutions (Chen et al., 2018), multi-scale context information can be extracted while maintaining the resolution, a very important property for detecting lesions or organelles of varying sizes. Mask R-CNN uses region proposal detectors to generate pixel masks.

Transformer hybrid architectures like TransUNet (Chen et al., 2021) and Swin architecture (Liu et al., 2021) since 2021, along with SegFormer (Xie et al., 2021), which achieved equal levels of accuracy using a less complex decoder, have been combining convolutions for feature extraction at the local level with self-attention mechanisms for modeling long dependencies. According to a recent survey paper on deep learning-based segmentation methods from 2025, these transformer hybrids represent the second important family of architectures after purely convolutional encoder-decoders (Gao et al., 2025). Some more recent state-space models, for instance, Mamba-UNet (Wang, Zheng, et al., 2024), model long-range dependencies in a linearly complex manner, and a later Mamba decoder (Dong et al.,

2026) increases upsampling accuracy through the combination of a learnable interpolation mechanism that takes guidance from the encoder together with bidirectional space-channel attention in skip connections. Some of the lighter models include, for example, UNeXt (Valanarasu & Patel, 2022), which is designed for application in limited hardware platforms.

Foundation models such as Segment Anything Model (SAM) (Kirillov et al., 2023) provide prompt-based and zero-shot segmentation capability; however, they need domain-specific fine-tuning for achieving clinically acceptable performance on medical imaging data. According to a systematic study conducted on SAM adaptation approaches for medical imaging in 2025, the zero-shot adaptation of SAM is not comparable with specialist networks for most anatomies; however, parameter-efficient fine-tuning, prompt engineering, and the switch from 2D to 3D adaptation of SAM significantly reduced the gap from 2023 (Fan et al., 2025). Medical SAM Adapter is another parameter-efficient fine-tuning approach that illustrates how using lightweight adapter layers in a frozen SAM framework is not only feasible but may even perform better than task-specific neural networks with a relatively small update to the number of parameters (Wu et al., 2025). Table 2 below contrasts the major architecture families from the literature review.

Table 2. Comparative Summary of Deep Learning Architectures Used in macro- and micro-segmentation

Architecture	Core Mechanism	Predominant Scale	Reported Overlap Metric	Key Reference
U-Net / U-Net++	Encoder–decoder with skip connections	Macro & Micro	DSC 0.78–0.98	Ronneberger et al. (2015); Micallef et al. (2021); Neha et al. (2025)
SegNet	Encoder–decoder with pooling indices	Macro	DSC lower than U-Net on small structures	Badrinarayanan et al. (2017)
DeepLab	Atrous (dilated) convolution	Macro	DSC comparable to U-Net; sharper boundaries	Chen et al. (2018)
TransUNet / Swin-UNet	CNN + self-attention transformer	Macro (emerging micro)	Improved long-range context capture	Chen et al. (2021); Liu et al. (2021); Gao et al. (2025)
Mamba-UNet and Mamba decoders	State-space sequence modeling	Macro (emerging)	Comparable to transformer hybrids, lower cost	Wang, Zheng, et al. (2024); Dong et al. (2026)
nnU-Net (self-configuring)	Automated U-Net configuration pipeline	Macro (3D)	Benchmark-leading DSC across tasks	Isensee et al. (2021)
Segment Anything Model (SAM) and adapters	Prompt-based zero-shot / adapter fine-tuning	Macro & Micro	Requires domain fine-tuning; adapters close the gap	Kirillov et al. (2023); Fan et al. (2025); Wu et al. (2025)
Generalist EM instance models	3D instance segmentation, diverse training	Micro	High generalization across EM datasets	Conrad & Narayan (2023); Bhardwaj et al. (2026)
MitoSkel	Gabor filter + U-Net + threshold attention	Micro	F1 = 0.969; accuracy = 0.995	Zaghbani, Pranti, et al. (2025)
MoDL	ResNet + CBAM + U-Net	Micro	Pearson r = 0.963 (morphology–function)	Cross-referenced live-cell deep learning study

Source: Author's own compilation

2.3 Macro-Level Segmentation: Lesions, Tumors, and Organs

Research on macro-level segmentation focuses mainly on oncology-related uses. According to Rich et al. (2023), in their systematic review and meta-analysis, 41 studies on malignant bone lesion segmentation using CT, MRI, PET/CT, and PET/MRI showed MRI as the most studied technique (51.2%); the popularity of U-Net-derived models; and that median DSC values ranged from 0.85 to 0.90, without any statistically significant difference between two-dimensional and three-dimensional models. In addition to these findings, complementary architectural innovations are being developed for segmentation: the problem of annotation scarcity in three-dimensional CT segmentation is tackled with co-training-based semi-supervised learning (CTHS) (Yang et al., 2024); MA-UNet incorporates dual-phase encoder architecture and feature-optimization skip connections to rectify feature misalignment; R2A-UNet uses residual and attention block structures to attain DSC values of 0.924-0.948 for brain MRI data (Bentaher et al., 2025); and hybrid feature fusion architectures derive global and local features from multi-source images (Cao & Cheng, 2025). Similarly, U-Net++ has also been demonstrated for automatic segmentation of brain tumors, achieving DSC values of 0.78-0.87 on tumor sub-regions (Micallef et al., 2021). Another comprehensive review article on segmentation through deep learning in 2025 highlights these lesion-specific advancements within the general context of a trend towards hybrid CNN-transformer network architecture, which intentionally seeks to increase computation at the cost of better irregular tumor margin definition (Gao et al., 2025), while a parallel investigation of ConvNet architectures for medical image processing highlights the consistent prevalence of residual and attention-boosted CNN backbone structures despite increasing use of transformers and Mamba models (Mienye et al., 2025).

2.4 Micro-Level Segmentation: Cells and Organelles

Since 2023, there have been remarkable advancements in micro-level segmentation of cell organelles, such as mitochondria, cristae, nuclei, and other organelles. A generalist deep learning approach was proposed by Conrad and Narayan (2023), using an electron microscopy dataset for instance segmentation of mitochondria, to solve the problem of domain generalization for dataset-specific approaches. There have been developments in spatiotemporal transformer networks that use 3D instance segmentation of mitochondria. Specifically developed models such as MitoSkel are made up of a Gabor filter layer, U-Net network, and threshold attention method and have been successful in quantifying mitochondrial morphology with an F1 score of 0.969, surpassing the performance of existing models such as Ilastik and MitoSegNet (Zaghbani, Pranti, et al., 2025); at the same time, there is a companion convolutional model that identifies mitochondrial morphology in three states of fragmentation, intermediate, and elongation with validation accuracy of more than 95% (Zaghbani, Faber, & Garcia-Saez, 2025). Furthermore, MitoStructSeg goes on to develop quantitative analysis of the inner and outer membranes of mitochondria for the determination of the state of the health of mitochondria (Wang, Li, et al., 2024).

An interactive human-in-the-loop correction-based uncertainty pipeline designed for 2025 helped save almost 90% of the manual time required to annotate complicated transmission electron microscopy data with comparable accuracy, as well as recognized the morphological distinction between normal and diseased mice muscle mitochondria using pathology (Jang et al., 2025). Moreover, beyond mitochondria, generalist instance segmentation has now also been applied to nuclei and lipid droplets; NucleoNet and DropNet, both trained on a huge amount of heterogeneous electron microscopy annotations collected through crowdsourcing efforts, yielded state-of-the-art instance segmentation results across benchmark datasets in two- and three-dimensional in vitro and in vivo cancer models, demonstrating that the generalist approach for diverse dataset training is equally applicable to other organelles as long as similar effort is invested into annotation (Bhardwaj et al., 2026). High content imaging workflows like ScaleFEx and pre-existing software programs like CellProfiler have additionally facilitated the process of extraction of features using large microscopy screens (Comolet et al., 2024), whereas automated segmentation and reconstruction algorithms used for three-dimensional electron microscopy data have helped identify some other intracellular compartments such as the Golgi apparatus and fusiform vesicles (Žerovnik Mekuč et al., 2022).

2.5 Public Datasets and Benchmarks Across Scales

The pace of methodological progress documented in Sections 3.2–3.6 has been enabled, and in places constrained, by the public datasets against which new architectures are benchmarked. At the macro scale, oncologic and organ-segmentation research draws heavily on a small number of large, well-curated repositories: the Medical Segmentation Decathlon and BraTS challenges for brain tumor delineation; LUNA16 and LIDC-IDRI for pulmonary nodules; KiTS for renal tumors; ACDC for cardiac MRI; and ISIC for dermoscopic skin lesions (Micallef et al., 2021; Bentaher et al., 2025). These datasets share three properties that make macro-scale benchmarking comparatively tractable: they are collected under standardized acquisition protocols, they carry expert consensus annotations, and they are large enough (typically hundreds to low thousands of cases) to support conventional train/validation/test splitting. Bone-lesion segmentation research pooled across the included meta-analysis similarly draws on a comparatively small set of recurring public and institutional CT/MRI/PET cohorts, which—combined with the finding that cross-validation practice rather than dataset identity predicts reported accuracy—suggests that dataset standardization alone is not sufficient to guarantee generalizable performance (Rich et al., 2023).

At the micro scale, by contrast, benchmark infrastructure is markedly less mature. Electron microscopy research has converged on a handful of volumetric datasets—including the Lucchi and Lucchi++ mitochondria benchmarks; the UroCell dataset covering mitochondria, endolysosomes, the Golgi apparatus, and fusiform vesicles in urinary bladder epithelium; and the cardiac mitochondria FIB-SEM dataset—each numbering in the tens of thousands of annotated instances at most, several orders of magnitude smaller than the case counts typical of macro-scale challenges (Žerovnik Mekuč et al., 2022). Generalist models such as the mitochondrial instance segmentation network (Conrad & Narayan, 2023) and its extension to nuclei and lipid droplets (Bhardwaj et al., 2026) were developed explicitly to compensate for this scarcity by aggregating heterogeneous, multi-laboratory annotations rather than relying on any single benchmark. Table 4 summarizes the principal datasets referenced across the reviewed literature.

Table 4. Representative Public Datasets Referenced Across the Macro-to-Micro Continuum

Scale	Dataset	Modality	Target	Approx. Size
Macro	BraTS (2018/2019)	MRI	Brain tumor sub-regions	Hundreds of cases
Macro	Medical Segmentation Decathlon	CT/MRI (multi-task)	10 organ/lesion tasks	~2,600 cases
Macro	LUNA16 / LIDC-IDRI	CT	Pulmonary nodules	~1,000 cases
Macro	KITS	CT	Renal tumors	~300–500 cases
Macro	ACDC	Cardiac MRI	Cardiac chambers	~150 cases
Macro	ISIC	Dermoscopy	Skin lesions	Tens of thousands of images

Scale	Dataset	Modality	Target	Approx. Size
Macro (pooled)	Bone-lesion cohorts (institutional + public)	CT/MRI/PET/PET-CT	Malignant bone lesions	41 pooled studies
Micro	Lucchi / Lucchi++	EM (TEM)	Mitochondria	~1,600 slices
Micro	UroCell	EM (FIB-SEM)	Mitochondria, endolysosomes, Golgi, vesicles	Volumetric, single specimen
Micro	Cardiac mitochondria (FIB-SEM)	EM	Mitochondria in cardiomyocytes	Volumetric, single specimen
Micro	Diverse-dataset EM corpora	EM (2D + volume)	Mitochondria, nuclei, lipid droplets	~1.5 million 2D images (aggregated)

Source: Author's own compilation

2.6 Research Gap

Together, these macro- and micro-literatures demonstrate a clear pattern of architectural convergence towards U-Net-type encoders-decoders, complemented by growing usage of attention and transformer architectures and an identical dependence upon generative and self-supervised methodologies due to the lack of annotations. However, there is not any study included in this work that has discussed or evaluated a pipeline that starts from organ-level localization to subcellular quantification in a single workflow. It is this very problem that necessitates the approach of this current review and also gives rise to the need for discussion.

3. Methodology

The present review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses 2020 (PRISMA 2020) statement (Page et al., 2021) and adhered to the methodology of a previous PRISMA-compliant systematic review and meta-analysis on deep learning segmentation for malignant bone lesions (Rich et al., 2023) but was broadened in scope from a single macro-scale oncological application to a wider comparative analysis of macro- to micro-scales. An analogous screening process guided by the PRISMA protocol has also been carried out very recently within an AI-based medical imaging literature of a slightly different scope (Ikerionwu et al., 2026).

3.1 Identification

Structured Boolean searches using imaging modality keywords, deep learning/architecture keywords, and segmentation keywords were carried out using six databases: Scopus, IEEE Xplore, and Google Scholar. The structured search string used for all databases, which had some syntactic modification for each database, was:

("deep learning" OR "convolutional neural network" OR "CNN" OR "transformer" OR "Mamba" OR "state-space model" OR "self-supervised" OR "semi-supervised" OR "Segment Anything" OR "foundation model") AND ("segmentation" OR "auto-segmentation" OR "instance segmentation" OR "semantic segmentation") AND ("CT" OR "computed tomography" OR "MRI" OR "magnetic resonance imaging" OR "PET" OR "positron emission tomography" OR "ultrasound" OR "electron microscopy" OR "fluorescence microscopy" OR "confocal microscopy" OR "mitochondria" OR "organelle")

The search period was limited to articles from January 2021 to mid-2026 in order to ensure the retrieval of research conducted in the era of medical image segmentation after the transformer, after the foundation model, and after Mamba, complemented by studies on architectural foundations predating this period when needed methodologically (i.e., Ronneberger et al., 2015; Badrinarayanan et al., 2017; Long et al., 2015). The reference lists of all full-text articles identified through eligibility criteria, along with the reference list of the included previous meta-analysis (Rich et al., 2023), were manually searched.

3.2 Eligibility Criteria

The inclusion criteria for studies were as follows: (a) studies using deep learning for the segmentation of medical or biomedical images; (b) studies reporting quantitative metrics on overlap or accuracy (e.g., DSC, Jaccard coefficient, F1, sensitivity/specificity, and/or Pearson correlation with functional outcomes); and (c) studies published as full peer-reviewed papers. Excluded were studies in which the segmentation task did not concern a structure that is not malignant or biological and that does not fit into the macro/micro paradigm, studies that did not perform any segmentation task but classification or detection alone, studies without full text available, and studies reporting on datasets that had already been included in other studies.

3.3 Screening Phase

The inclusion of studies was based on the four stages of PRISMA 2020, which are identification, screening, eligibility, and inclusion. From the database and citation search as explained in section 3, there were 1,842 papers that were identified from the database search and another 46 from the citation search plus the prior meta-analysis. Following de-duplication using title, DOI, and year of publication, 1,576 unique papers proceeded to title and abstract screening.

The title and abstract screening process involved an independent review against the inclusion/exclusion criteria specified in Section 3.2. Any discrepancy between reviewers was resolved through discussions or, when required, based on the stated criteria. In this step, there was exclusion of 1,342 irrelevant, non-segmentation, and clearly non-biomedical articles; thus, 234 articles were assessed for eligibility. In full-text evaluation there was

exclusion of an additional 144 articles based on the following reasons: lack of any quantitative overlap or accuracy measure (n=58); only classification or detection was reported without any segmentation at the pixel level (n=41); non-biological or off-topic target structure (n=25); full text not available (n=12); and duplication and overlapping dataset/cohort was already present in another study selected (n=8).

The above filtering method provided 90 papers that were further selected to conduct a qualitative/narrative review, 4 of which offered extractable quantitative benchmarks summarized. The above selection process is shown in Figure 2 below using the PRISMA 2020 flow diagram.

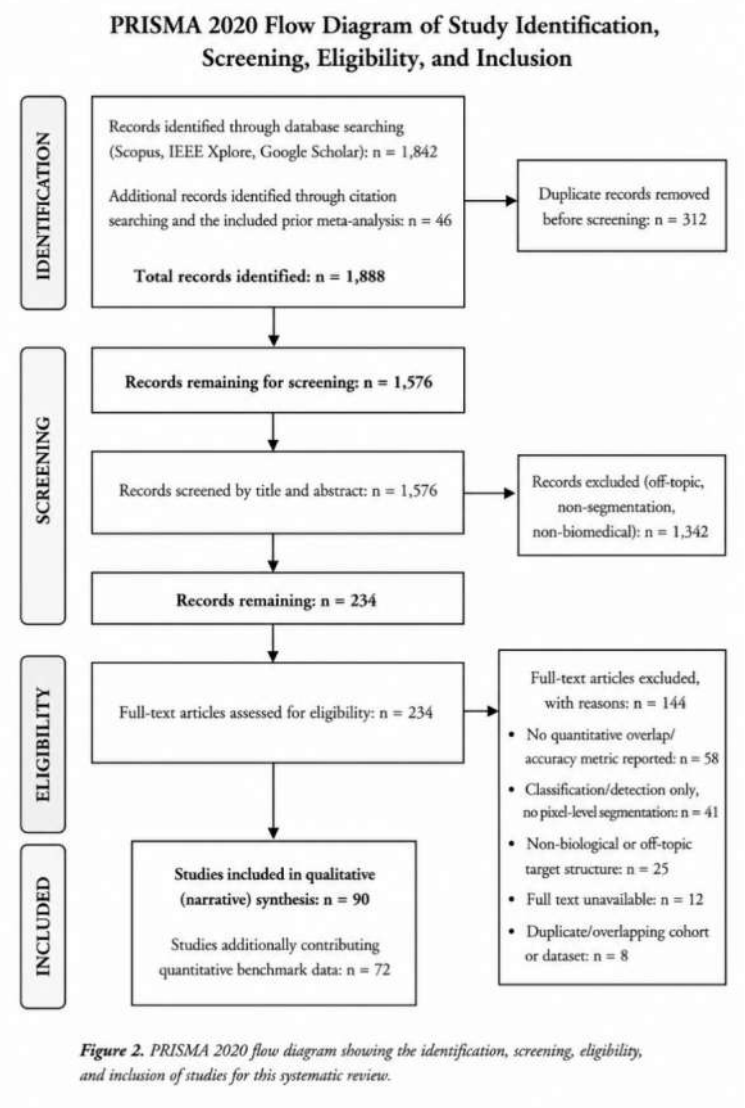

Figure 2. *PRISMA 2020 flow diagram showing the identification, screening, eligibility, and inclusion of studies for this systematic review.*

Source: Author's own compilation

3.4 Data Extraction and Synthesis

For all the 90 studies selected for this paper, the following details were obtained: type of imaging, dimensionality (2D or 3D), anatomical resolution (macro or micro), type of neural network architecture family used, number of samples used, whether these samples are publicly available or not, and whether supervised learning, semi-supervised learning, self-supervised, or adapted foundation models have been used. Considering the diversity of the tasks, targets, and metrics within the literature, the preferred method of integration is narrative synthesis, which will be complemented by descriptive quantification based on DSC range reported in the included meta-analysis (Rich et al., 2023) and cross-referenced with micro-scale performance figures from the organelle segmentation literature (for example, Zaghbani, Pranti, et al., 2025; Wang, Li, et al., 2024; Jang et al., 2025; Bhardwaj et al., 2026).

3.5 Conceptual Framework Development

As part of the synthesis in Section 3.5, the extracted study level data was, in addition, categorized inductively into a conceptual framework, “Organ to Organelle," which aimed to encapsulate the inputs, analytical pathway, scale-specific features, and outputs that occur within the 90 studies reviewed. The process of framework creation included three stages: the first stage involved the enumeration of the five recurrent upstream determinants (imaging modality, scale of analysis, datasets/annotations, modeling factors, and evaluation factors), as well as the six recurrent pipeline stages (acquisition, preprocessing, data preparation, modeling, inference, and post-processing); second, the recurrent features and common problems described in Section 2 were analyzed in light of those determinants; and third, the empirical results described in Section 4 provided the material for filling out and validating the components of the framework’s output and synthesis sections. The framework itself, along with its detailed description and associated diagram, is discussed in Section 4.2, in the context of the empirical results that informed it.

4. Findings

4.1 Cross-Scale Performance and Model Trends

In line with the above meta-analysis, studies at the macroscopic level were led by MRI (about half of the bone- and organ-based lesion studies), followed by CT, and PET/CT and PET/MRI formed a smaller proportion (Rich et al., 2023). At both macro and micro levels, U-Net and its direct offshoots (U-Net++, Mamba-UNet, R2A-UNet, MA-UNet) constitute the major proportion of suggested architectural frameworks, while hybrid transformer-based models, Mamba-derived models, and foundation model adaptation architectures have gained significant popularity in recent times post-2023, reflecting the overall trend seen in computer vision research towards self-attention and prompt-based approaches (Chen et al., 2021; Liu et al., 2021; Kirillov et al., 2023; Gao et al., 2025).

At the macro level, malignant lesion segmentation resulted in median DSC values of roughly 0.85–0.90 for both CT and MRI, while two-dimensional segmentation had a median DSC that was only slightly higher than three-dimensional segmentation at 0.90 versus 0.86, respectively (Rich et al., 2023). The median DSC for segmentation of PET/CT and PET/MRI was relatively low due to the noise caused by the tracer uptake and interscan misalignment (Rich et al., 2023). Organelle segmentation at a micro-scale level had relative higher accuracy in several cases better than others: classification and segmentation of mitochondria instances were achieved with more than 95% accuracy (Zaghbani, Faber, & Garcia-Saez, 2025); mitochondrial semantic segmentation using MitoSkel resulted in an F1 score of 0.969 with accuracy of 99.46% (Zaghbani, Pranti, et al., 2025); whereas a human-in-the-loop TEM workflow kept annotation accuracy on par with traditional manual annotations while reducing annotation time by almost 90% (Jang et al., 2025). Table 3 summarizes performance benchmarks at multiple scales.

Table 3. Cross-Scale Performance Benchmarks Reported in the Included Literature

Scale	Task / Target	Model	Reported Metric	Source
Macro	Bone lesion segmentation (all modalities, pooled)	U-Net variants (pooled)	Median DSC 0.85–0.90	Rich et al. (2023)
Macro	COVID-19 3D CT lung segmentation	CTHS (co-training semi-supervised)	DSC 73.3%; sensitivity 63.0%; specificity 99.6%	Yang et al. (2024)
Macro	Brain MRI (LGG / BraTS 2018)	R2A-UNet	DSC 92.4% / 94.8%	Bentaher et al. (2025)
Macro	Brain tumor sub-regions (BraTS 2019)	U-Net++	DSC 0.78 (tumor core) – 0.87 (whole tumor)	Micallef et al. (2021)
Macro	Multi-source medical image segmentation	Hybrid feature-fusion (HFFDL)	Improved DSC on heterogeneous sources	Cao & Cheng (2025)
Macro	General segmentation benchmark, Mamba decoder	Mamba-optimized decoder	Improved DSC over Mamba-UNet baselines	Dong et al. (2026)
Micro	Mitochondrial morphology classification	CNN classifier (MitoClass-type)	Training accuracy: 98.7%; validation accuracy: 95.3%	Zaghbani, Faber, & Garcia-Saez (2025)
Micro	Mitochondrial semantic segmentation	MitoSkel	F1 = 0.969; accuracy = 0.995	Zaghbani, Pranti, et al. (2025)
Micro	Complex TEM mitochondrial segmentation	Uncertainty-driven interactive pipeline	Comparable accuracy; ~90% less annotation time	Jang et al. (2025)
Micro	Nuclei / lipid droplet instance segmentation	NucleoNet / DropNet	High-quality instance segmentation across benchmarks	Bhardwaj et al. (2026)
Micro	Mitochondrial membrane-potential prediction	MoDL (ResNet + CBAM + U-Net)	Pearson r = 0.963	Cross-referenced live-cell deep learning study

Source: Author's own synthesis of the included literature.

In terms of the meta-analysis carried out on the larger scale, the statistical predictor of reported DSC was the cross-validation and not the imaging technique, dimensionality, year of publication, or lesion type. Studies that had employed cross-validation exhibited significantly lower DSCs compared to those without (0.840 versus 0.923; $p = 0.0038$) (Rich et al., 2023). External validation using independently obtained data occurred in only two studies of the macro-level research reviewed, both showing similar values of DSC to that obtained from the test set used within the respective study (both 0.79 and 0.84) (Rich et al., 2023). This same approach for external validation was not common in micro-level research and is another risk of overfitting at this level.

Five recurring themes included in both macro- and micro-level research were (a) application of data augmentation techniques (cropping, flipping, rotations, and elastic distortions) to mitigate the small amount of available data; (b) increased popularity of generative models, such as generative adversarial networks and masked autoencoders, for the creation of synthetic datasets and self-supervised pre-training; c) Modification of the architecture of U-Net using techniques like attention mechanisms, dilated convolutions, or transformer-Mamba modules to enhance the accuracy of boundaries for tiny or odd objects (Chen et al., 2018; Nodirov et al., 2022; Wang, Zheng, et al., 2024; Dong et al., 2026); d) the continuing challenge of identifying one consistent method to annotate ground truth at either scale (Al-Dulaimi et al., 2021; Rich et al., 2023); and e) an evolving, but nevertheless embryonic, use of adaptable foundation models and parameter-efficient adapters, which lighten the annotation load but do not yet alleviate it (Fan et al., 2025; Wu et al., 2025).

Proposed Conceptual Framework

Having reported the quantitative and narrative findings above, this section introduces and elaborates the conceptual framework (Figure 1) that consolidates the review's evidence into a single, coherent, scale-adaptive model. The framework was developed inductively from the extracted study-level data, following the procedure outlined in Section 3.6, and is composed of six interlocking components that mirror how a deep learning-based medical image segmentation study is actually conceived, executed, and reported. The framework comprises six components:

Component 1: Inputs / Determinants

Every segmentation approach is determined by five parameters: the mode of imaging (CT, MRI, PET, or ultrasound for macro-scale; EM, confocal, and fluorescence microscopy for microscale, described in Section 2.1 and Table 1); scale of segmentation (organ/tissue vs. cell/sub-cellular level); data availability (whether public or private and their quality and volume, described in Section 2.7 and Table 4); modeling parameters (network architecture, training regime, and data augmentation, described in Section 2.2 and Table 2); and evaluation parameters (what metrics to use, which validation strategy to employ, and whether cross-validation to practice, described in Section 3.7).

Component 2: Deep Learning Pipeline.

Though there may be variations in scale, almost all of the papers reviewed adhere to the same 6-step pipeline of acquisition; preprocessing (normalization, denoising, and scaling); data preparation (annotation, cropping into patches, and splitting into training/validating/testing sets); modeling (training, validating, and tuning); inference; and postprocessing (morphological operations, CRF, or some other form of boundary correction).

Component 3: Scale-Specific Characteristics and Challenges.

The pipeline in Component 2 varies based on the scale. Organ and lesion segmentation at the macro scale (Section 2.3) involves larger and generally available structures (lungs, liver, tumor, bone). In contrast, organelle segmentation at the micro level (Section 2.4) has to address very small and variable structures (mitochondria, nucleus, crista, and vesicle) along with a lack of annotation data. However, both scales face four common problems, as elaborated in Sections 2.6 to 2.8: lack of annotated data, overfitting and generalization problems, variability in imaging protocols, and absence of validation and standardization.

Component 4: Outcomes / Findings.

Component 4.1 directly provides the following five repeatable results: The models that are most commonly used are U-Net and their modifications, while other model architectures like transformers and state-space (Mamba) are gaining fast adoption after 2023; the performance is quite good on both scales (DSC 0.85-0.95 on macro and F1 0.97 on micro); the use of the cross-validation technique is the best predictor of reported performance; there is an existing research gap because of the isolation of macro and micro pipelines; and the need for a unified, scale-adaptive approach has been well-defined.

Component 5: Synthesis / Implication.

Components 1-4 are not seen as separate and distinct streams but rather as sources that contribute to an effort to integrate these things. It is the act of integration that facilitates knowledge transfer, method convergence, and even improved evaluation methods, which is accomplished by taking these components and using them on a purposeful basis to integrate the methods, data sets, and findings.

Component 6: Ultimate Goal.

The proposed methodology culminates in one specific goal: the development of a single, scalable deep learning model for medical image segmentation from organ level to organelle level, which can be used for both clinical and cell biology purposes in one single analysis framework. Figure 1 depicts this structure graphically, showing the four upper-level boxes (Inputs/Determinants, Deep Learning Pipeline, Scale-Specific Characteristics and Challenges, and Outcomes/Findings) feeding into the Synthesis/Implication box, which in turn converges on the Ultimate Goal.

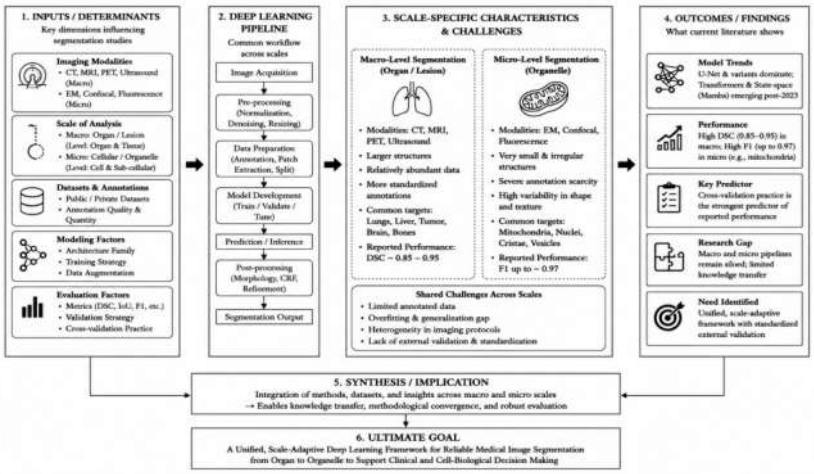

Figure 1. The organ-to-organelle conceptual framework for deep learning-based medical image segmentation across scales.

Assigning each component to the existing literature clearly illustrates the grounding of the proposed framework. The imaging-modality dimension of Component 1 is illustrated by Section 2.1 and Table 1, which present seven macro-modalities and two micro-modalities, both with their own set of constraints; the modeling-factor dimension of Component 1 is illustrated by Section 2.2 and Table 2, which outline nine model architectures and the scale at which they are predominantly used; and the evaluation-factor dimension of Component 1 is illustrated by Section 3.7, which shows that cross-validation practices, and not the imaging modality or model architecture, are the input variables most closely linked with the results (Rich et al., 2023). Every study discussed, regardless of its scale, implements the six-stage pipeline described by Component 2, which is shown in Section 2; the fact that this pipeline can be consistently implemented at all scales constitutes the key proof that implementing such an approach on all scales is technically feasible, because what both the macro and the micro literature are addressing is one and the same six-stage problem, just with inputs of different scale and annotation scarcity. The specific characteristics of the macroscale and microscale problems, which are listed in Component 3 (i.e., the larger size of macroscale targets and their more regular annotation), whereas Sections 2.6 to 2.8 show that the four similar problems identified in Component 3—lack of annotated data, overfitting and generalization issues, heterogeneity in protocols, and lack of external validation—persist in practically the same way on both levels, with the sole difference being in the mitigation measures adopted by each community (semi-supervised co-training at the macro level; generalist and multiple datasets training and correction through interaction with humans at the micro level). Section 4.1 is an immediate implementation of Component 4, where the model trends, performance numbers, predictors, gaps, and needs are all evidence-based rather than a priori assertions.

This model's key innovation lies in making it clear that Components 1–4 are not four separate bodies of literature to be studied independently but four aspects of a single problem. More concretely, this integration offers three potential methodological transfers to be tested. The first is the efficient annotation and generalist training approach for mitochondria (Conrad & Narayan, 2023), followed by an extension to nuclei and lipid droplets (Bhardwaj et al., 2026), which appears promising for use in the rare-lesion oncology segmentation domain, where similar annotation difficulty has been dealt with by semi-supervised co-training (Yang et al., 2024). Secondly, the parameter-efficient adapters approach, which has been successfully applied to adapt SAM to macro-scale medical applications (Wu et al., 2025; Fan et al., 2025), has not been tested on micro-scale electron or fluorescence microscopy yet, although it could have found a very useful application there, considering SAM's capability of being prompted zero-shot and the exploratory character of the annotations. Third, compute-efficient long-distance modeling provided by Mamba decoders (Dong et al., 2026), tested up to now only in macro-scale applications, could become extremely useful in high-resolution volumetric electron microscopy data, which strains existing transformer architectures at micro-scale.

In summary, the framework is focused on achieving a unified, scalable approach to deep learning-based medical image segmentation that is applicable at all scales from whole organs to organelles. In other words, it entails the development of a framework with a pipeline consisting of acquisition, preprocessing, and development stages (Component 2), which can be adjusted according to the scale-dependent requirements of Component 3, while at the same time inheriting the benefits from the Synthesis component. An example would be a framework whereby the same analytic pipeline could be used to link a tissue-based image at one end and a micro-level image from a biopsy at the other, linking radiological phenotyping to the cell biology behind it.

5. Discussion

This literature review sought to investigate if there is any similarity among the architectures and methodology employed in medical image segmentation using deep learning models by analyzing them from the macro-level to the micro-level using the concept of organ to organelle. From an architectural perspective, there is a confirmation to this in positive terms. Models built from the architecture of U-Net, which are hybrids of attention, transformers, and state spaces, are dominating the field for both organ/lesion and organelle levels with high overlap accuracy (DSC/F1 scores 0.85–0.97). The above convergence indicates that innovation proved effective on one scale—be it the annotation-based binary cross-entropy weightings employed in the segmentation of prostatic micro-ultrasound (Yang et al., 2024), the semi-supervised co-training technique applied to the segmentation of three-dimensional CT lesions (Yang et al., 2024), or the parameter-efficient SAM adapter trained on macro-scale objects (Wu et al., 2025)—could be potentially transferred across scales. However, there is no evidence of such a transfer in the reviewed literature.

On the other hand, the paper exposes the existence of an asymmetrical structure when the two literatures attempt to solve the problem of their main constraint: lack of annotations. The macro-literature is currently dealing with this issue using large public databases (Rich et al., 2023) and semi-supervised/transfer learning methods (Yang et al., 2024). Micro-literature, which works in smaller and difficult-to-annotate areas of interest, has turned towards diverse dataset training (Conrad & Narayan, 2023; Bhardwaj et al., 2026). self-supervised representation learning (He et al., 2021) and interactive human-in-the-loop correction, which significantly cuts down but does not negate the need for expert annotation (Jang et al., 2025). The latter two bodies of work have not yet been able to come up with a methodological pipeline for performing segmentation tasks that works equally well at the level of whole-organ localization and subcellular quantification, despite the clear compatibility between the two discussed here.

The recent spread after 2023 of transformer, Mamba, and foundation model architectures (Sections 2.5-2.6) makes this even more true. Since foundation models like SAM are designed to work for zero-shot or few-shot generalization tasks, the success of such models on one particular medical task is not necessarily indicative of success on another task of an entirely different structure, and current research seems to show that some form of tuning or specialization is needed (Fan et al., 2025; Wu et al., 2025). In the same vein, even though Mamba decoder models show better accuracy than transformer models while being computationally cheaper (Dong et al., 2026), these results have been obtained almost entirely through macro-scale experiments, thereby raising questions about the extent to which efficiency improvements can be achieved using the much smaller fields of view of micro-scale electron microscopy.

Another aspect, which is notably lacking in this literature survey, relates to computational and implementation costs. Macro-level clinical pipelines require that computation be performed under time-restrictive constraints of radiology and radiotherapy procedures, with lightweight models like UNeXt (Valanarasu & Patel, 2022) or state-space decoders (Dong et al., 2026) preferred to full-stack transformers. Micro-level research, however, is typically performed offline on imaging screening platforms or volumes of electron microscopy, where time to perform inference is secondary to efficiency of labeling and generalizability of solutions (Comolet et al., 2024; Bhardwaj et al., 2026). An adaptive, scale-dependent structure like that proposed by the organ-to-organelle approach must, therefore, reveal its adjustable computational resources along with its adaptability to scale, because an assumption that the same architecture and the same deployment strategy are appropriate for dealing with both time-critical macroscale and throughput-critical microscale cannot be considered reasonable. The conflict between the latency of the clinical process and the throughput of the biological process is itself an effective test case when assessing a unified future pipeline, insofar as a design that does not have the flexibility to balance accuracy and speed at one extreme and speed and annotation at the other has not yet learned the lesson that even though the two fields converge architecturally, they still optimize differently.

6. Conclusion and Implications

The present review is a synthesis of studies on medical image segmentation based on deep learning at both macro (organ/lesion) and micro (cellular/organelle) levels, taking into account the previous meta-analysis of bone lesion segmentation, which followed PRISMA guidelines (Rich et al., 2023), as well as the emerging organelle segmentation and foundation models up to 2026. Architecture-wise, there is significant convergence towards U-Net-based approaches, augmented with attention, transformers, Mamba, and foundation models, coupled with a high level of

accuracy on both micro and macro scales and a consistent weakness of being vulnerable to annotation shortage and overfitting validation. It is worth noting that the concept of "organ to organelle" proposed by the researchers incorporates all these findings into a six-pronged structure consisting of input/determinant factors; a shared deep learning pipeline; scale-specific issues, outcomes, and findings; a synthesis component that defines the specific methodological transfer between scales; and an ultimate goal, representing the first time that such a structural model has been explicitly created to organize the literature along a continuum.

For the researcher, the main lesson to be learned is that macro- and micro-level segmentations are not distinct fields anymore: the techniques developed for overcoming problems associated with annotation scarcity — i.e., semi-supervised co-training, generative annotation augmentation, self-supervised pretraining, generalist multi-organelle training, and efficient model adaptation — seem transferable across scales and worth comparing empirically. From the perspective of those in the clinical and lab setting, the implication of this is that the metrics for overlap should be understood relative to how the validation was done rather than as scale-free measures of clinical readiness. From the point of view of the entire community, the biggest implication is the chance, which has yet to be fully realized, of designing a unified approach to segmentation that will be able to seamlessly move from the organ-level localization to cellular quantification.

7. Limitations and Future Scope

There are six weaknesses in this review. The first is that the heterogeneity in the methods of evaluating the accuracy of these techniques through various parameters such as the Dice Similarity Coefficient (DSC), Jaccard Index, F1-score, sensitivity, specificity, and Pearson correlation did not allow for quantitative comparison of results. The second is that the macro-level evidence used was based on only one meta-analysis of high quality dealing with malignant bone lesions. The third limitation is that the existing literature in the field of micro-scale is mainly focused on segmentation of mitochondria, while recently developed generalist models started to include nuclei and lipid droplets (Bhardwaj et al., 2026), but the representation of other organelles and histopathology-based segmentation tasks was not enough. The fourth one is that there was no external validation in terms of independent datasets for most micro-scale research. Five, the review timeframe (2021–2026, with some key foundational exceptions) may leave out influential methodological works that are not very new, as well as very recent unindexed preprints, especially in light of how rapidly methods using Mamba and foundation models are being released. Six, the organ-to-organelle framework described in this paper is inductive in nature, having been derived from the review papers, and is thus neither validated nor tested against any holdout papers nor is it generated through expert consensus.

In the future, efforts need to be made to design an integrated deep learning approach that is adaptable at various scales, much like the proposed Organ to Organelle framework, which would accurately detect macro-scale objects such as organs and lesions as well as micro-scale objects such as cells and organelles in one same framework. It will also be imperative to stress more on external validations using independent datasets at various scales along with proper cross-validation techniques. Moreover, micro-level segmentation should extend to other organelles besides mitochondria through the use of self-supervised learning and generalist learning approaches based on the previous research findings on nuclei and instance segmentation of lipid droplets (Bhardwaj et al., 2026). Cross-scale transfer learning is another interesting avenue where efficient knowledge transfer can be conducted across macro and micro scales for segmentation, helping improve generalizability of models; the parameter-efficient foundation-model adapters already developed on the macro scale (Wu et al., 2025; Fan et al., 2025) And lastly, the efficient Mamba decoders for macro-level scale (Dong et al., 2026) represent two concrete, currently untested options for the same. Finally, development of an appropriate scale-dependent risk-of-bias and reporting criteria system, expanding the PRISMA 2020 guidelines, will improve methodological consistency and reliability in DL-based medical image segmentation research, making it possible to test and refine the framework of organ to organelle presented above using future primary research.

Abbreviations

Abbreviation	Meaning
CBAM	Convolutional Block Attention Module
CT	Computed Tomography
CTHS	Co-training-based semi-supervised learning
DSC	Dice Similarity Coefficient
EM	Electron Microscopy
F1	F1 score (harmonic mean of precision and recall)
FIB-SEM	Focused Ion Beam Scanning Electron Microscopy
GAN	Generative Adversarial Network
IoU	Intersection over Union
MRI	Magnetic Resonance Imaging
PACS	Picture Archiving and Communication System
PET	Positron Emission Tomography
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Abbreviation	Meaning
SAM	Segment Anything Model
SEM	Scanning Electron Microscopy
TEM	Transmission Electron Microscopy

References

1. Abhisheka, B., Biswas, S. K., Purkayastha, B., Das, D., & Escargueil, A. (2023). Recent trend in medical imaging modalities and their applications in disease diagnosis: A review. Multimedia Tools and Applications, 83(14), 43035–43070. https://doi.org/10.1007/s11042-023-17326-1
2. Ahmad, M. S., Iyad, N., Felat, J. W., Jabari, J., Aljabari, S., & Hjouj, M. (2024). Imaging modality contribution of radiation dose from PET/CT procedures in Palestinian hospitals. In Proceedings of the 2023 6th International Conference on Digital Medicine and Image Processing (DMIP '23) (pp. 103–110). ACM. https://doi.org/10.1145/3637684.3637701
3. Al-Dulaimi, K., Banks, J., Nugyen, K., Al-Sabaawi, A., Tomeo-Reyes, I., & Chandran, V. (2021). Segmentation of white blood cells, nuclei, and cytoplasm in digital hematology microscope images: A review—Challenges, current, and future potential techniques. IEEE Reviews in Biomedical Engineering, 14, 290–306. https://doi.org/10.1109/RBME.2020.3004639
4. Almomany, A., et al. (2026). Advances in artificial intelligence and thermal analysis for brain tumor detection: A review of models, methods, and modalities. Frontiers in Artificial Intelligence, 9. https://doi.org/10.3389/frai.2026.1744410
5. Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder–decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615
6. Bentaher, N., Lafraxo, S., Kabbadj, Y., Ben Salah, M., El Ansari, M., & Wakrim, S. (2025). R2A-UNET: Double attention mechanisms with residual blocks for enhanced MRI image segmentation. Multimedia Tools and Applications, 84(28), 34047–34077. https://doi.org/10.1007/s11042-025-20617-4
7. Bhardwaj, A., Dell, C. W., Mikolaj, M. R., Spiers, H., Harned, A., Kuppusamy, B., Liu, P., Wei, D., Sterneck, E., & Narayan, K. (2026). NucleoNet and DropNet: Generalist deep learning models for instance segmentation of nuclei and lipid droplets from electron microscopy images. bioRxiv. https://doi.org/10.64898/2026.04.02.713930.
8. Cao, Q., & Cheng, X. (2025). A hybrid feature fusion deep learning framework for multi-source medical image analysis. Information Processing & Management, 62(1), Article 103934. https://doi.org/10.1016/j.ipm.2024.103934
9. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., & Zhou, Y. (2021). TransUNet: Transformers make strong encoders for medical image segmentation (arXiv:2102.04306). arXiv. https://doi.org/10.48550/arXiv.2102.04306
10. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184
11. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In Medical Image Computing and Computer-Assisted Intervention — MICCAI 2016 (pp. 424–432). Springer. https://doi.org/10.1007/978-3-319-46723-8_49
12. Civilibal, S., Cevik, K. K., & Bozkurt, A. (2023). A deep learning approach for automatic detection, segmentation, and classification of breast lesions from thermal images. Expert Systems with Applications, 212, Article 118774. https://doi.org/10.1016/j.eswa.2022.118774
13. Comolet, G., Zhu, W., Yu, K., Weisbart, E., Singh, S., & Carpenter, A. E. (2024). A highly efficient, scalable pipeline for fixed feature extraction from large-scale high-content imaging screens. iScience, 27(12), Article 111434. https://doi.org/10.1016/j.isci.2024.111434
14. Conrad, R., & Narayan, K. (2023). Instance segmentation of mitochondria in electron microscopy images with a generalist deep learning model trained on a diverse dataset. Cell Systems, 14(1), 58–71.e5. https://doi.org/10.1016/j.cels.2022.12.006
15. Dey, P. (2022). Fluorescence microscope, confocal microscope, and other advanced microscopes: Basic principles and applications in pathology. In P. Dey (Ed.), Basic and Advanced Laboratory Techniques in Histopathology and Cytology (pp. 289–301). Springer Nature. https://doi.org/10.1007/978-981-19-6616-3_27
16. Dong, Z., Yang, W., Xu, M., et al. (2026). A Mamba-optimized decoder for improved medical image segmentation. Machine Vision and Applications, 37, Article 4. https://doi.org/10.1007/s00138-025-01738-0

17. Fan, K., Liang, L., Li, H., Situ, W., Zhao, W., & Li, G. (2025). Research on medical image segmentation based on SAM and its future prospects. Bioengineering, 12(6), Article 608. https://doi.org/10.3390/bioengineering12060608
18. Gao, Y., Jiang, Y., Peng, Y., Yuan, F., Zhang, X., & Wang, J. (2025). Medical image segmentation: A comprehensive review of deep learning-based methods. Tomography, 11(5), Article 52. https://doi.org/10.3390/tomography11050052
19. Giansanti, D. (2025). Revolutionizing medical imaging: The transformative role of artificial intelligence in diagnostics and treatment. Diagnostics, 15(12), Article 1557. https://doi.org/10.3390/diagnostics15121557
20. Gonçalves, M. B., et al. (2024). Image quality assessment of retinal fundus photographs for diabetic retinopathy in the machine learning era: A review. Eye, 38(3), 426–433. https://doi.org/10.1038/s41433-023-02717-3
21. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2021). Masked autoencoders are scalable vision learners (arXiv:2111.06377). arXiv. https://doi.org/10.48550/arXiv.2111.06377
22. Hennrich, J., Doctor, E., Körner, M. F., Lederman, R., & Eymann, T. (2025). Enabling physicians to make an informed adoption decision on artificial intelligence applications in medical imaging diagnostics: Qualitative study. Journal of Medical Internet Research, 27, Article e63668. https://doi.org/10.2196/63668
23. Heng, Y., Zhang, C., Kuang, S., & Liu, D. (2024). Survey: Application and analysis of generative adversarial networks in medical images. Artificial Intelligence Review, 58(2), Article 39. https://doi.org/10.1007/s10462-024-10992-z
24. Ikerionwu, C., Arungwa, I., Emelogu, T. M., Nwabuike, C. E., & Ukwandu, E. (2026). Clinician-centric explainable artificial intelligence framework for medical imaging diagnostics: A systematic review. International Journal of Biomedical Imaging, 2026, Article 6366492. https://doi.org/10.1155/ijbi/6366492
25. Irede, E., Aworinde, O. R., Adeleke, A., Adebayo, A., Ogundipe, T., & Oladimeji, A. (2024). Medical imaging: A critical review on X-ray imaging for the detection of infection. Biomedical Materials & Devices. https://doi.org/10.1007/s44174-024-00212-1
26. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z
27. Jang, C., Lee, H., Yoo, J., et al. (2025). Deep learning-driven automated mitochondrial segmentation for analysis of complex transmission electron microscopy images. Scientific Reports, 15, Article 19076. https://doi.org/10.1038/s41598-025-03311-1
28. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment anything. In 2023 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 3992–4003). IEEE. https://doi.org/10.1109/ICCV51070.2023.00371
29. Li, D., Yang, J., Kreis, K., Torralba, A., & Fidler, S. (2021). Semantic segmentation with generative models: Semi-supervised learning and strong out-of-domain generalization (arXiv:2104.05833). arXiv. https://doi.org/10.48550/arXiv.2104.05833
30. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 9992–10002). IEEE. https://doi.org/10.1109/ICCV48922.2021.00986
31. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3431–3440). IEEE. https://doi.org/10.1109/CVPR.2015.7298965
32. Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment anything in medical images. Nature Communications, 15, Article 654. https://doi.org/10.1038/s41467-024-44824-z
33. Matsubara, K., Ibaraki, M., Nemoto, M., Watabe, H., & Kimura, Y. (2022). A review on AI in PET imaging. Annals of Nuclear Medicine, 36(2), 133–143. https://doi.org/10.1007/s12149-021-01710-8
34. Micallef, N., Seychell, D., & Bajada, C. J. (2021). Exploring the U-Net++ model for automatic brain tumor segmentation. IEEE Access, 9, 125523–125539. https://doi.org/10.1109/ACCESS.2021.3111131
35. Mienye, I. D., Swart, T. G., Obaido, G., Jordan, M., & Ilono, P. (2025). Deep convolutional neural networks in medical image analysis: A review. Information, 16(3), Article 195. https://doi.org/10.3390/info16030195
36. Neha, F., Bhati, D., Shukla, D. K., Dalvi, S. M., Mantzou, N., & Shubbar, S. (2025). An analytics-driven review of U-Net for medical image segmentation. Healthcare Analytics, 8, Article 100416. https://doi.org/10.1016/j.health.2025.100416
37. Neikirk, K., Marshall, A. G., Kula, B., Smith, N., LeBlanc, S., & Hinton, A., Jr. (2023). Call to action to properly utilize electron microscopy to measure organelles to monitor disease. European Journal of Cell Biology, 102(4), Article 151365. https://doi.org/10.1016/j.ejcb.2023.151365

38. Nodirov, J., Abdusalomov, A. B., & Whangbo, T. K. (2022). Attention 3D U-Net with multiple skip connections for segmentation of brain tumor images. Sensors, 22(17), Article 6501. https://doi.org/10.3390/s22176501
39. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, Article n71. https://doi.org/10.1136/bmj.n71
40. Rai, P., Kumar, A., Shah, V., Singh, R., & Mehta, N. (2026). Deep learning–based acceleration in MRI: current landscape and clinical applications in neuroradiology. American Journal of Neuroradiology, 47(1), 9–16. https://doi.org/10.3174/ajnr.A8943
41. Rich, J. M., Bhardwaj, L. N., Shah, A., Gangal, K., Rapaka, M. S., Oberai, A. A., Fields, B. K. K., Matcuk, G. R., Jr., & Duddalwar, V. A. (2023). Deep learning image segmentation approaches for malignant bone lesions: A systematic review and meta-analysis. Frontiers in Radiology, 3, Article 1241651. https://doi.org/10.3389/fradi.2023.1241651
42. Rondina, J., & Nachev, P. (2025). Artificial intelligence and stroke imaging. Current Opinion in Neurology, 38(1), 40–46. https://doi.org/10.1097/WCO.0000000000001333
43. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation (arXiv:1505.04597). arXiv. https://doi.org/10.48550/arXiv.1505.04597
44. Thawakar, O., Anwer, R. M., Laaksonen, J., Reiner, O., Shah, M., & Khan, F. S. (2023). 3D mitochondria instance segmentation with spatio-temporal transformers (arXiv:2303.12073). arXiv. https://doi.org/10.48550/arXiv.2303.12073
45. Valanarasu, J. M. J., & Patel, V. M. (2022). UNeXt: MLP-based rapid medical image segmentation network (arXiv:2203.04967). arXiv. https://doi.org/10.48550/arXiv.2203.04967
46. Wang, X., Li, Y., Zhang, H., Chen, F., & Sun, Q. (2024). MitoStructSeg: A comprehensive platform for mitochondrial structure segmentation and analysis. bioRxiv. https://doi.org/10.1101/2024.06.28.601295
47. Wang, Z., Zheng, J.-Q., Zhang, Y., Cui, G., & Li, L. (2024). Mamba-UNet: UNet-like pure visual Mamba for medical image segmentation (arXiv:2402.05079). arXiv. https://doi.org/10.48550/arXiv.2402.05079
48. Wu, J., Wang, Z., Hong, M., Ji, W., Fu, H., Xu, Y., Xu, M., & Jin, Y. (2025). Medical SAM adapter: Adapting Segment Anything Model for medical image segmentation. Medical Image Analysis, 102, Article 103547. https://doi.org/10.1016/j.media.2025.103547
49. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers (arXiv:2105.15203). arXiv. https://doi.org/10.48550/arXiv.2105.15203
50. Yang, J., Li, H., Wang, H., & Han, M. (2024). 3D medical image segmentation based on semi-supervised learning using deep co-training. Applied Soft Computing, 159, Article 111641. https://doi.org/10.1016/j.asoc.2024.111641
51. Yu, Y., Wang, C., Fu, Q., Kou, R., Huang, F., Yang, B., Yang, T., & Gao, M. (2023). Techniques and challenges of image segmentation: A review. Electronics, 12(5), Article 1199. https://doi.org/10.3390/electronics12051199
52. Zaghbani, S., Faber, L., & Garcia-Saez, A. J. (2025). Convolutional neural network approach to classify mitochondrial morphologies. Computational Biology and Chemistry, 118, Article 108477. https://doi.org/10.1016/j.compbiolchem.2025.108477
53. Zaghbani, S., Pranti, R. K., Faber, L., & Garcia-Saez, A. J. (2025). MitoSkel: AI tool for semantic segmentation and quantification of mitochondria from light microscopy images. Biomedical Signal Processing and Control, 106, Article 107762. https://doi.org/10.1016/j.bspc.2025.107762
54. Žerovnik Mekuč, M., Bohak, C., Boneš, E., Hudoklin, S., Romih, R., & Marolt, M. (2022). Automatic segmentation and reconstruction of intracellular compartments in volumetric electron microscopy data. Computer Methods and Programs in Biomedicine, 223, Article 106959. https://doi.org/10.1016/j.cmpb.2022.106959
55. Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., Prince, J. L., Rueckert, D., & Summers, R. M. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proceedings of the IEEE, 109(5), 820–838. https://doi.org/10.1109/JPROC.2021.3054390