International Journal of Electrical and Computer Engineering (IJECE)
Vol. 16, No. 3, June 2026, pp. 1382~1398
ISSN: 2088-8708, DOI: 10.11591/ijece.v16i3.pp1382-1398

Hybrid convolutional neural network–transformer models for liver tumor segmentation: a comprehensive review

Ibrahim Mohamed Attiya[1], Mostafa Thabet[1], Mostafa R. Kaseb[2]
[1]Department of Information Systems, Faculty of Computers and Artificial Intelligence, Fayoum University, Fayoum, Egypt
[2]Department of Computer Science, Faculty of Computers and Artificial Intelligence, Fayoum University, Fayoum, Egypt

Article Info

Article history:

Received Feb 14, 2026
Revised Mar 21, 2026
Accepted Apr 27, 2026

Keywords:

Attention mechanisms
Deep learning
Hybrid CNN–transformer models
Liver tumor segmentation
Medical image analysis
Systematic review

ABSTRACT

Liver cancer is a major cause of cancer deaths worldwide, and early and accurate segmentation of liver tumors is a critical step in cancer diagnosis and treatment. However, existing image segmentation techniques have difficulty handling the variability of liver tumors on different image modalities. The emergence of deep learning (DL) and the development of convolutional neural networks (CNNs) have revolutionized image segmentation techniques. However, CNNs have limitations in handling long-range dependencies, which is a critical requirement for tumor segmentation. To overcome these limitations, researchers have proposed hybrid deep learning architectures, which combine CNNs and attention mechanisms or transformers, to integrate local and global information for image segmentation. In this paper, we provide a comprehensive and analytical review of over 50 state-of-the-art deep learning architectures for liver and tumor segmentation. In addition, we provide an extensive evaluation of 38 hybrid and advanced architectures for liver tumor segmentation and a comprehensive discussion of hybrid CNN-transformer architectures. We propose a novel multi-dimensional taxonomy and evaluate the state-of-the-art architectures on various dimensions, including architectural innovation, segmentation accuracy, computational efficiency, and clinical applicability using benchmark datasets such as LiTS and 3DIRCADb. In our critical evaluation of the state-of-the-art architectures, we identify some of the limitations and challenges of existing research and propose a unified evaluation framework and future research directions on self-supervised learning, explainable artificial intelligence (XAI), federated learning, and lightweight architectures.

Corresponding Author:

Ibrahim Mohamed Attiya
Faculty of Computers and Artificial Intelligence, Information System, Fayoum University
Fayoum, Egypt
Email: im1959@fayoum.edu.eg

1. INTRODUCTION

Early detection and diagnosis of cancer are among the key challenges in clinical oncology, with liver cancer being a leading cause of cancer-related deaths across the globe. Hepatocellular carcinoma, the most common type of liver cancer, is recognized as a highly aggressive cancer type with minimal or no symptoms during the early stages of the disease. Global cancer mortality statistics indicate that hepatocellular carcinoma (HCC) resulted in 830,000 cancer-related deaths in the year 2020, thus underscoring the need for reliable and effective early detection methods. Accurate segmentation of liver cancer is a basic requirement in the evaluation and management of cancer, which is widely practiced in several therapeutic interventions, including surgery, radiofrequency therapy, and image-guided interventions.

Manual segmentation of imaging modalities, including computed tomography, magnetic resonance imaging, and positron emission tomography, is generally labor-intensive, time-consuming, and associated with significant inter and intra-observer variability. This led researchers to develop reliable image segmentation methods that are automated, reproducible, and accurate. Deep learning, in particular convolutional neural networks, is recognized as a breakthrough in image segmentation, as they are effective in dealing with local features and spatial hierarchies in images. However, the inherent limitations of the convolutional neural networks are in dealing with the modeling of long-range dependencies, which are important in the accurate segmentation of heterogeneous tumor regions with irregular boundaries [1].

Recent literature also reflects the pattern in the progress of the field's advancement. For example, the UNet++ architecture, which is an extension of the U-Net architecture, has been applied for liver tumor segmentation using dense skip connections and multi-scale feature aggregation techniques [2]. Other recent studies have focused on the development of unified deep learning models for integrating segmentation with other tasks such as classification in computed tomography (CT) imaging studies [3]. RFiLM U-Net and MAPFUNet incorporate radiomic features and fusion using multi-attention techniques for improved boundary segmentation. Detection-based approaches using YOLOv8 and learning approaches based on patch-based methods for limited datasets [4], as well as spatiotemporal learning using the CUNet-CLSTM model, have also been proposed. More recent studies have focused on the potential benefits of using transformers in CNNs for liver tumor segmentation using ResTransUNet and MMEFU-Net.

Despite the significant advancements recorded by these recent studies, a significant gap in comprehensive reviews that specifically integrate recent advances in hybrid CNN-transformer architectures for liver tumor segmentation tasks is evident. Although previous reviews provide a general overview of deep learning techniques in medical imaging analysis, they often focus more on CNN-based techniques or generic transformer models and therefore fail to represent adequately the potential synergy between hybrid models developed between 2023 and 2025. Moreover, most existing reviews are descriptive in nature and primarily focus on providing a comprehensive overview of architectures instead of a rigorous comparative analysis based on quantitative metrics such as Dice scores and evaluation using standard metrics such as LiTS and 3DIRCADb datasets.

Therefore, in this study, a systematic methodology is followed to provide a comprehensive and analytical overview of recent advancements in hybrid CNN-transformer architectures for liver tumor segmentation tasks. The main contributions of this study are:

a. A novel taxonomy: a novel taxonomy is proposed by classifying recent state-of-the-art models based on distinct hybridization strategies such as serial, parallel, and multi-scale fusion.
b. Performance benchmarking: a rigorous comparative analysis of recent models developed between 2023 and 2025 is provided by comparing computational efficiency, Dice scores, and architectural innovations.
c. Methodological rigor: unlike previous reviews that primarily provide a descriptive analysis based on a general overview of architectures and techniques, this study follows a rigorous methodology based on the PRISMA 2020 guidelines.
d. Future roadmap: emerging trends such as federated learning and explainable artificial intelligence (AI) in the context of hybrid models are discussed in this study to overcome issues such as interpretability in recent hybrid models.

The specific time frame of 2023-2025 captures the *"most intensive phase of hybrid CNN-transformer innovation in liver tumor segmentation."*

2. BACKGROUND AND PRELIMINARIES

Liver cancer, specifically HCC, is a major public health problem and a major contributor to cancer-related deaths. In the context of HCC diagnosis, it is critical to note that the early detection of liver cancer is hindered by the lack of specific and significant clinical symptoms during the early stages. These aspects highlight the need to develop accurate and timely diagnostic tools to assist in decision-making. Computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) play a vital role in the diagnosis and treatment of liver cancer. In the context of medical image analysis, medical image segmentation is considered to be an essential step in the quantitative analysis of medical images. In the context of liver and liver tumor segmentation, it is critical to note that accurate segmentation of the liver and tumors plays a vital role in volumetric analysis and treatment planning. Traditionally, manual and semi-automated methods were adopted to segment liver and tumors. These methods were found to be highly labor-intensive and took a significant amount of time. In addition, these methods were also found to exhibit significant intra- and inter-observer variability. In the context of computer-based methods, initial methods adopted to segment medical images were found to exhibit significant limitations in handling variability in liver and tumor shapes. These limitations have led to a gradual move from machine learning-based methods to deep learning-based methods.

2.1. Evaluation metrics

The objective assessment of the performance of segmentation is carried out through standardized quantitative measures. In the context of medical image segmentation research, the most frequently used measures of evaluation of segmentation performance have been: dice similarity coefficient (DSC): The Dice Similarity Coefficient measures the spatial overlap between the predicted segmentation and the reference ground truth, with values between 0 and 1. Intersection over union (IoU): The Intersection over Union measures the ratio of the intersection and union of the predicted and reference segmentation, providing a different measure of volumetric agreement between the two. Hausdorff distance (HD): The Hausdorff Distance measures the maximum distance between the boundary points of the predicted and reference contours. These measures provide different views of the quality of the segmentation performance.

2.2. Benchmark datasets

Progress in liver and liver tumor segmentation has been facilitated by the availability of publicly accessible benchmark datasets with expert annotations. Commonly used datasets in this domain include:

a. Liver tumor segmentation challenge (LiTS): A widely adopted dataset comprising contrast-enhanced CT scans with annotations for liver and tumor regions.
b. 3DIRCADb: A collection of high-resolution 3D CT volumes with manually segmented liver and tumor masks, including cases with complex and multifocal tumors.
c. CHAOS: A multi-modal dataset that includes CT and MRI (T1-DUAL and T2-SPIR) images, supporting cross-modality liver segmentation studies.
d. SLIVER07 and BTCV: Additional CT-based datasets frequently used for liver segmentation benchmarking and validation.
e. Medical segmentation decathlon (MSD) – liver task: A subset of the MSD challenge providing CT volumes with liver and tumor annotations.

These datasets enable comparative analysis of segmentation approaches under diverse imaging conditions and anatomical variations. It is important to note that reported results across studies are often obtained using different experimental protocols, preprocessing strategies, and data splits, which may affect direct comparability.

3. TAXONOMY OF DEEP LEARNING-BASED LIVER AND TUMOR SEGMENTATION MODELS

The rapid evolution of deep learning has resulted in a diverse range of architectures for liver and liver tumor segmentation. To provide a structured overview of this landscape, this review organizes representative approaches into a taxonomy based on their underlying architectural principles. The classification reflects both conceptual design choices and historical development trends within the field, ranging from conventional convolutional models to more recent hybrid and multimodal systems. The main categories considered in this review are: i) CNN-based models, ii) attention-enhanced CNN models, iii) hybrid CNN–transformer models, iv) transformer-only models, and v) multimodal and ensemble approaches. Figure 1 illustrates an overview of the proposed taxonomy.

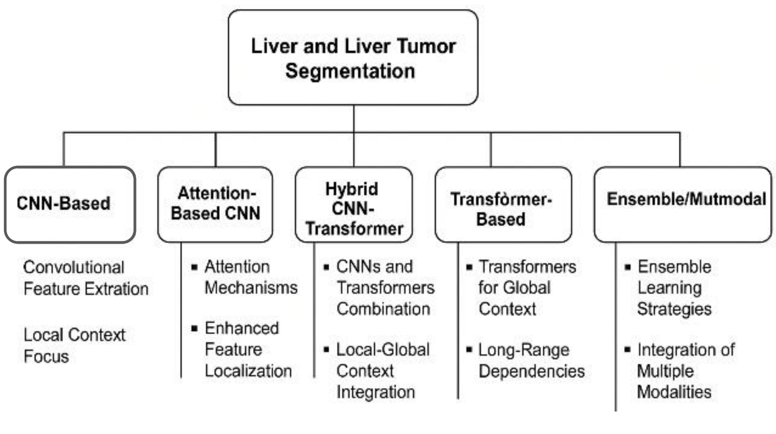

Figure 1. Taxonomy of deep learning-based medical image segmentation models

3.1. CNN-based models

Convolutional neural networks (CNNs) form the basis of the dominant approach to modern medical image segmentation tasks. Their success is attributed to inductive biases appropriate for image data, including translation invariance and hierarchical local feature learning using convolution operations. For liver tumor segmentation, encoder-decoder architectures like the popular U-Net and its variants are popular choices. This approach involves the implementation of an encoder to capture contextual features and a corresponding decoder to ensure spatial details are maintained during the process, as described in the literature. Representative work and rationale: UNet++ [2] is an enhanced version of the traditional U-Net architecture, in which nested and dense skip connections are incorporated to improve feature aggregation and gradient flow. This approach has shown promising results in liver tumor segmentation using MRI images. Additionally, the unified CNN framework described in [3] attempts to unify both classification and segmentation into a single framework to facilitate diagnostic processes. CNN-based architectures are recognized for their computational efficiency and relative simplicity in comparison to other models. However, the localized nature of convolution operations in CNNs prevents them from capturing spatial relationships between features, which may be essential in tumor segmentation, particularly in scenarios involving spatial context.

3.2. Attention-enhanced CNN models

In order to overcome the limitation of CNNs in contextual modeling, attention mechanisms have been incorporated to allow the network to selectively emphasize certain spatial locations or feature channels. The attention mechanisms can be spatial, channel-wise, or a combination of both. Representative works and rationale: The work on MAPFUNet [4] utilizes multi-attention perception modules within a CNN architecture to enhance the sensitivity of the network to small and low contrast tumors within a CT imaging environment. The work on LA-ResUNet [5] utilizes attention gates within a residual U-Net architecture to allow the network to selectively emphasize certain locations within a longitudinal imaging environment. The use of CNNs with attention mechanisms has shown promising results for improving the accuracy of the network within challenging environments; however, the addition of these modules to the network can increase the complexity of the network and, consequently, the risk of overfitting.

3.3. Transformer-only models

However, in purely transformer models, as seen in implementations inspired by vision transformer (ViT), convolutional operations are completely circumvented. Self-attention mechanisms are seen as the main way of interaction between image patches. For dense prediction tasks such as liver segmentation, models such as Swin-UNet make use of hierarchical shifted window attention mechanisms. Key work and rationale: Purely transformer models have shown remarkable potential in global context understanding. However, their application in liver tumor segmentation models often remains hindered due to the requirement of large datasets as well as high computational expenses. As seen in recent literature, it has been shown that such models often remain plagued by challenges in local fine-grained details required to understand tumor boundary definitions, a problem that hybrid models hope to overcome.

3.4. Hybrid CNN–transformer models

Hybrid architectures utilize the inductive biases of CNNs for local feature extraction and the ability of the transformer to address global contextual dependencies. This is especially relevant for liver tumors that often display irregular shapes and low contrast with adjacent tissues. Representative Works: ResTransUNet [6] combines the use of residuals with global self-attention for the preservation of spatial resolution and the acquisition of wide-range context information. MMEFU-Net [7] also employs the use of multi-encoder fusion for feature representation. Critical analysis: compared with CNN-based architectures that do not utilize global context information and transformer-based architectures that may sacrifice precise information locally, the use of hybrid architectures provides a well-rounded methodology. The current review aims to cover the period from 2023-2025, during which newer architectures have proposed the concept of “lightweight hybridization” for the significant computational cost associated with the use of this methodology.

3.5. Multimodal and ensemble methodologies

Multimodal or ensemble methodologies use the benefits of different information sources provided by different imaging modalities or different models to increase the robustness of the model. These methodologies are similar to the actual scenario, wherein CT, MRI, or PET scans are taken into consideration simultaneously. Some of the representative works with justification: The ensemble model that combines ResUNet and InceptionV4, as discussed in reference [8], shows higher robustness with the combination of CT and MRI datasets due to the utilization of different models. The V-NET-VGG16 model discussed in reference [9] combines the power of classification and segmentation by combining V-Net with VGG16 architecture. The RFiLM U-Net model discussed in reference [10] combines handcrafted radiomic features with deep learning models. Even though the above models show higher robustness, they are more complex, require more computational resources, and require more data to be aligned.

This taxonomy offers a structured framework that may be used to categorize and compare more than 38 contemporary models that are presented throughout the comparative sections that follow. The taxonomy represents the overall trend within the field, moving from localized feature learning-based models toward more unifying models that aim to integrate local and global information. This trend represents the overall trend toward improving robustness and segmentation accuracy, as well as several practical considerations that are important with regard to clinical applicability. Figure 1 presents a structured taxonomy of some of the more prominent deep learning models that are currently employed in the context of medical image segmentation, with a focus on the analysis of images of the liver and liver tumors. The taxonomy presents the prevailing approaches within the field as one of four main architectural paradigms: convolutional neural network, hybrid CNN-transformer, transformer, and multimodal ensemble models.

Table 1 provides a structured comparative overview of five widely adopted architectural paradigms in deep learning-based liver and tumor segmentation. It summarizes their core design principles, representative implementations, commonly reported strengths, and typical limitations as discussed in the literature. The comparison aims to support qualitative model understanding and highlight general architectural trends rather than standardized performance benchmarking.

Table 1. Comparative analysis of deep learning model archetypes for liver segmentation

Category	Core Description	Exemplary Studies	Principal Strengths	Key Limitations
CNN-based models	Utilize encoder-decoder architectures with convolutional layers for hierarchical local feature extraction.	[2], [3]	**Computational efficiency**, proven reliability, easy pipeline integration.	**Limited receptive field** hampers long-range dependency modeling; struggles with **irregular tumor boundaries**.
Attention-based models	Integrate attention modules into CNNs to dynamically focus on salient regions like tumor margins.	[1], [4], [5]	Enhanced boundary delineation, better handling of heterogeneous textures and small lesions.	Introduces **moderate computational overhead**; requires careful architectural design.
Hybrid CNN–transformer models	Fuse CNN encoders (local features) with transformer modules (global context) for complementary modeling.	[6], [7], [11],	Superior accuracy; captures both local details and global anatomical context; state-of-the-art performance.	**High computational/memory demands**; complex training; longer inference times.
Transformer-only models	Rely exclusively on self-attention to model global dependencies without convolutional bias.	[11]	**Exceptional global context integration**; promising foundational accuracy.	**Extremely data-hungry**; highest computational cost; less explored for 3D medical segmentation.
Multi-modal and ensemble models	Combine multiple imaging modalities or aggregate predictions from several models to improve robustness.	[9], [12], [13]	Increased robustness/generalization; leverages complementary information from different sources.	High system complexity; dependent on co-registered multi-modal data; resource-intensive training.

4. METHODOLOGY FOR LITERATURE SEARCH AND ANALYSIS

This systematic review follows the preferred reporting items for systematic reviews and meta-analyses (PRISMA 2020) guidelines. The search was conducted across four major databases: IEEE Xplore, PubMed, ScienceDirect, and Google Scholar. The search strings used were: ('Hybrid CNN' OR 'Transformer') AND ('Liver tumor' OR 'Hepatocellular Carcinoma') AND 'Segmentation'. The final search was updated on [write date, January 15, 2026].

4.1. Search Strategy

A comprehensive literature search was performed in the following electronic databases: PubMed/MEDLINE, IEEE Xplore, Scopus, Web of Science, SpringerLink, and the ACM Digital Library. Additionally, preprint servers (arXiv and bioRxiv) were searched to capture recently published research not yet indexed in traditional databases. The last search was conducted on Jan 15, 2026. The search query combined keywords related to i) liver tumor segmentation, ii) deep learning, iii) convolutional neural networks, iv) transformers, and v) hybrid models. Boolean operators and field tags were adapted for each database. The exact search strings used were as shown in Table 2.

Table 2. Exact search strings used across various electronic databases

Database	Search Query
PubMed	("liver tumor"[Title/Abstract] OR "hepatic tumor"[Title/Abstract] OR "hepatocellular carcinoma"[Title/Abstract]) AND ("segmentation"[Title/Abstract] OR "delineation"[Title/Abstract]) AND ("deep learning"[Title/Abstract] OR "convolutional neural network"[Title/Abstract] OR "CNN"[Title/Abstract] OR "transformer"[Title/Abstract] OR "hybrid"[Title/Abstract]) AND ("2023"[Date - Publication] : "2025"[Date - Publication])
IEEE Xplore	((("liver tumor" OR "hepatic tumor" OR "hepatocellular carcinoma") AND (segmentation OR delineation)) AND (("deep learning" OR "neural network" OR "CNN" OR "transformer" OR "hybrid"))) AND (2023-2025)
Scopus	TITLE-ABS-KEY (("liver tumor" OR "hepatic tumor" OR "hepatocellular carcinoma") AND (segmentation OR delineation) AND ("deep learning" OR CNN OR transformer OR hybrid)) AND PUBYEAR > 2022 AND PUBYEAR < 2026
Web of Science	TS=((("liver tumor" OR "hepatic tumor" OR "hepatocellular carcinoma") AND (segmentation OR delineation)) AND (("deep learning" OR "neural network" OR "CNN" OR "transformer" OR "hybrid"))) AND PY=(2023-2025)
arXiv/bioRxiv	(liver tumor segmentation) AND (deep learning OR transformer OR hybrid)

4.2. Inclusion and exclusion criteria

Studies were included if they met the following criteria:

a. Inclusion criteria:
   - Original research articles focusing on segmentation of the liver and/or liver tumors using deep learning techniques.
   - Studies introducing novel hybrid CNN-transformer architectures or substantial modifications to existing ones, or highly relevant CNN-based or attention-based models for comparative analysis.
   - Publications reporting quantitative segmentation performance using standard metrics (*e.g.*, Dice similarity coefficient, IoU, Hausdorff distance).
   - Articles published in peer-reviewed journals or conference proceedings between January 1, 2023, and December 31, 2025. High-quality preprints from this period were also considered if they presented mature work.
   - Studies written in English.

b. Exclusion criteria:
   - Studies not focused on the liver (*e.g.*, brain, lung, kidney, or other organ segmentation).
   - Review articles, meta-analyses, surveys, or editorials that did not introduce new model contributions.
   - Papers lacking sufficient technical detail or clear experimental validation (*e.g.*, abstracts only, short papers).
   - Duplicate publications or earlier versions of the same study.
   - Studies not using deep learning as the core methodology.

.

4.3. Study selection process

The study selection process followed the PRISMA guidelines, as illustrated in the flow diagram in Figure 2. Two reviewers (IMA and MT) independently screened the titles and abstracts of all retrieved records. Full-text articles of potentially relevant studies were then retrieved and assessed against the eligibility criteria. Any disagreements during the selection process were resolved through discussion or consultation with a third reviewer (MRK). The reasons for exclusion at the full-text stage were documented.

As shown in Figure 2, the diagram illustrates systematic identification, screening, eligibility assessment, and inclusion of studies. A total of 850 records were identified through database searching. After removing duplicates, 620 records were screened based on titles and abstracts, leading to the exclusion of 500 records. The remaining 120 full-text articles were assessed for eligibility, of which 57 were excluded for reasons such as unsuitable study design, population, or failure to meet inclusion criteria. Consequently, 38 studies were included in the final qualitative and quantitative synthesis. No additional records were identified from other sources.

4.4. Data extraction and synthesis

For each of the selected studies, the following data were extracted into a standardized template:

a. Bibliometric information: authors, year, journal/conference.
b. Model architecture: model name, architectural family (*e.g.*, CNN-based, attention-enhanced, hybrid CNN-transformer, transformer-only, multimodal/ensemble), and core technical components (*e.g.*, attention gates, transformer blocks, fusion strategies).

c. Experimental setup: benchmark dataset(s) employed (*e.g.*, LiTS, 3DIRCADb, CHAOS), data splits, preprocessing details, and validation strategy (*e.g.*, cross-validation, fixed split).
d. Performance metrics: reported quantitative results, specifically Dice scores for liver and tumor segmentation, intersection over union (IoU), and Hausdorff distance (HD95) where available.
e. Computational profile: reported information on model complexity (number of parameters), training requirements, and inference speed, if available.
f. Key findings and limitations: strengths of the proposed approach, claimed clinical relevance, and limitations noted by the authors.

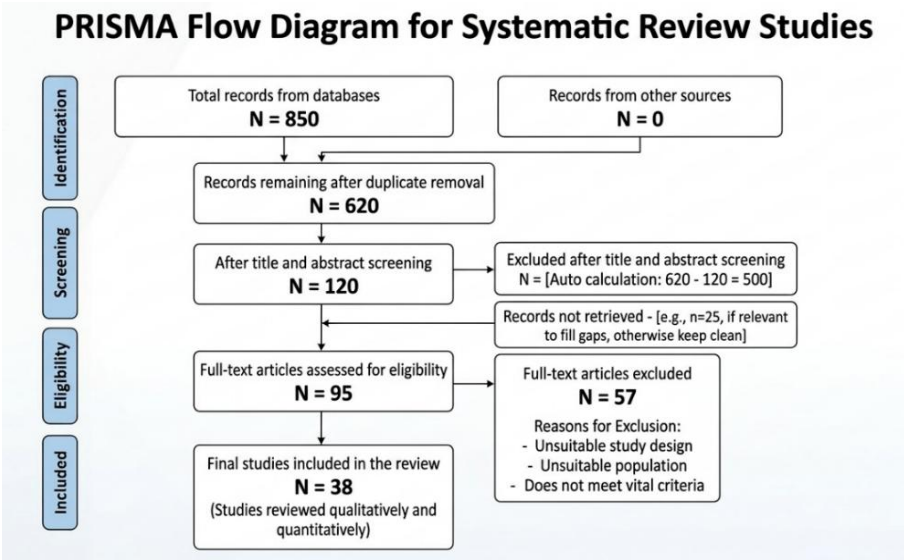

Figure 2. PRISMA 2020 flow diagram of the study selection process

4.5. Analytical approach

The extracted data were analyzed qualitatively along four primary dimensions, aligning with the taxonomy proposed in Section 3:

a. Architectural design and innovation: assessing the novelty of architectural integration strategies.
b. Reported segmentation performance: a comparative discussion of results as presented in the original studies, with careful consideration of the heterogeneity in experimental settings.
c. Computational efficiency and complexity: analyzing the trade-off between accuracy and resource requirements.
d. Clinical translation potential: evaluating aspects such as robustness, generalizability, and alignment with clinical workflows.

“No new experiments were conducted; all results are reported as stated in the original publications.”

5. RELATED WORK

Precise and automatic medical image segmentation has traditionally been at the epicenter of computational hepatology, fueled by the medical need for the early detection and diagnosis of liver cancer. Convolutional neural networks have laid the foundation for the current era of automatic medical image segmentation through the ability of deep learning models to learn hierarchical features and show improved performance over conventional approaches. The evolution of medical image segmentation has followed the trajectory of successive improvements in architectures, including the addition of attention mechanisms to focus the model's attention on important regions of the image, such as the tumor boundaries, transformer architectures for the modeling of long-range dependencies, and the use of hybrid models combining the advantages of CNNs and transformers. Previous review articles have presented a review of the latest developments in medical image segmentation, including those used in the context of abdominal or liver imaging. The review articles offer important insights into the architectures, datasets, and evaluation metrics used in the models. However, the review of the literature also reveals a limitation in the form of the absence of a comparative discussion of the results of the models and the lack of a review of the latest hybrid architectures, including those proposed from 2023 to 2025. This survey provides an analytical overview of the existing 38 liver-related studies from the years 2023-2025, as listed in Table 3 (appendix). The comparative insights from the existing literature will help identify the overall trend of the evolution of architectures in the field of liver segmentation tasks. The hybrid architectures of CNN and transformer models, such as the ResTransUNet [13], and multimodal architectures, such as the ResUNet-InceptionV4 [8], have shown good results in achieving high Dice scores with increased computational requirements. On the other hand, lightweight architectures of CNN models, such as the UNet70 [14] and LiverNet [15], are important for the deployment of the models in a clinical environment, considering the reduced computational requirements. The removal of non-liver-related studies will help readers focus more on the challenges of liver-related tasks and the overall requirements of the model for robust results.

This synthesis of the existing literature will help readers get a comprehensive overview of the overall trend of the evolution of the models and the characteristics of the models used for the liver tumor segmentation task. Table 3 provides a comparative overview of 38 representative deep learning models for liver and liver tumor segmentation. The reported Dice scores (liver/tumor) and key strengths and weaknesses are summarized as presented in the original studies. It is important to note that these results are based on reported performance in the literature and were obtained under heterogeneous experimental conditions. This overview is intended to highlight general trends, architectural characteristics, and practical considerations for model selection, rather than to serve as a standardized benchmarking of models.

As shown in Table 4, a rigorous quality assessment was performed for all 37 included studies based on five key criteria. The results indicate that 100% of the selected papers provided a clear definition of the segmentation task and used standardized evaluation metrics such as the dice similarity coefficient (DSC). While the majority of studies (92%) offered comprehensive dataset descriptions, about 63% provided an in-depth discussion regarding computational complexity and resource requirements. Overall, the high adherence to these quality markers ensures the reliability of the comparative analysis presented in this review.”

Table 4. Quality assessment of included studies

Quality Criteria	Number of Studies (N=37)	Percentage (%)
Clear objective and scope	37	100%
Comprehensive dataset description	34	92%
Clearly defined evaluation metrics (Dice/IoU)	37	100%
Specified validation method (*e.g.*, k-fold)	33	89%
Discussion of computational complexity	23	62%

6. ANALYSIS OF COMPARATIVE RESULTS AND EMERGING TRENDS

The overview in Table 3 enables an organized discussion of reported performance trends, architectural features, and implementation trade-offs among 38 modern models for liver and liver tumor segmentation. Instead of comparing models in terms of rank, this section will synthesize literature to point out trends relevant to both research and clinical interests.

6.1. Performance–accuracy trade-offs

A common theme in all the papers examined here has been this balance between how well a method performs in terms of accuracy, and how computationally expensive it is. Hybrid architectures such as ResTransUNet [13], which are a combination of CNNs and transformers, and ensemble methods such as ResUNet-InceptionV4 [8], which are combinations of two or more architectures, achieve high values of the Dice score. However, they are computationally expensive, requiring a lot of memory on the GPU and longer training times. Conversely, smaller CNNs such as UNet70 [14] and LiverNet [15], which are designed for faster inference and lower resource usage, are important for practical deployment, especially in situations where time is of the essence, such as in a clinical setting.

6.2. Architectural evolution and innovation

The taxonomy and comparative observations reveal a clear architectural progression:

a. From the local details to the global context: the field is moving beyond plain CNNs and into the realm of attention-rich and hybrid architectures. The target is to extend the limited view of the convolutional filter, especially in the presence of tumors with irregular shapes and varied appearances.
b. Among the new trends, the use of Hybrid Architectures is prominent, especially CNN-transformer hybrids, which are gaining more and more popularity due to the potential of the combination of precise local feature extraction and the global context. The popularity of CNN-transformer hybrids in the best-performing models in the recent years speaks for itself.

c. There is a growing trend in the use of Multi-modal and Task-Aware architectures, especially in the fusion of different data sources (MEFNet for PET/CT) and the combination of the segmentation task with other clinical objectives, such as survival analysis (RECISTSurv).

6.3. Modality-specific observations and clinical relevance

Modality-specific observations and clinical relevance used include:

a. CT-centric development: Computed tomography remains the dominant imaging modality, with the majority of reviewed models developed and evaluated on CT datasets. This reflects CT's widespread clinical availability and central role in liver cancer management.
b. MRI and alternative modalities: MRI-based approaches such as UGCMANet [24] demonstrate competitive reported performance by exploiting multi-contrast or dynamic sequences. Models targeting ultrasound [28] and PET/CT [30] address specific clinical niches, although they often face modality-specific challenges affecting segmentation consistency.
c. Workflow-oriented modeling: Several studies explicitly align model design with clinical workflows, including ConvNeXt-2U [7] for Y-90 radioembolization planning and LA-ResUNet [20] for longitudinal tumor monitoring. These efforts suggest increasing awareness of translational considerations beyond algorithmic accuracy.

6.4. Persisting challenges across studies

Despite notable progress, several challenges consistently emerge across the literature:

a. Data availability: High-capacity models, particularly transformer-based and hybrid architectures, rely on large, annotated datasets, which remain limited in many clinical settings.
b. Computational accessibility: Increased model complexity can hinder widespread adoption, especially in institutions with limited computational resources.
c. Generalization and robustness: Reported performance often varies across datasets, highlighting sensitivity to scanner differences, imaging protocols, and population diversity.
d. Small lesion segmentation: Accurately delineating small or low-contrast tumors continues to be challenging for many existing architectures.

6.5. Summary and transition

Overall, the literature suggests a progression towards balancing the focus away from the sole novelty of architecture and towards a more well-rounded evaluation of accuracy, efficiency, and deploy ability. Although the hybrid and ensemble models are certainly an interesting and evolving space, perhaps the most promising clinical implementation may be with the efficient CNN architectures and attention mechanisms. The recurring challenges presented here are the precursor to the open problems and future research directions presented in the following section.

7. FOCUSED RESULTS AND PRACTICAL RECOMMENDATIONS

By looking at the results reported in 38 different models for liver and liver tumor segmentation that are considered deep learning models (publications from 2023-2025), some interesting and applicable observations regarding the balance between accuracy and computational requirements are made. Instead of providing rankings and scores, some examples are provided from the literature that highlight the balance that exists in real-world scenarios.

7.1. Observed accuracy–complexity trade-offs

But recent work has also seen the rise of hybrid CNN-transformer models and attention-based architectures, which are achieving even higher segmentation accuracy, particularly in complex cases like tumors. To name a few, ResTransUNet is a CNN-transformer model that has reported higher Dice scores in liver and tumor segmentation. However, the accuracy increase also has a price to pay in terms of higher GPU memory usage and training complexity. On the other hand, attention-based CNNs like MAPFUNet achieve good accuracy in segmenting even smaller or less contrasted lesions through multi-attention mechanisms, all while maintaining a relatively moderate computational cost in terms of GPU memory usage. But CNNs like UNet++, which report slightly lower Dice scores in segmentation, still remain the best option in terms of model complexity and ease of deployment.

7.2. Context-dependent deployment considerations

The reviewed literature suggests that the suitability of a segmentation model is highly dependent on the intended clinical application, available infrastructure, and operational constraints:

a. High-resource environments: Several studies explore hybrid or attention-based architectures (*e.g.*, ResTransUNet, MAPFUNet [4]) in research or tertiary-care settings where advanced computational resources are available and segmentation precision is a primary concern.
b. Resource-constrained or high-throughput settings: Lightweight CNN-based models such as UNet++ [2] or LiverNet [15] are frequently discussed as practical alternatives due to their faster inference times and lower hardware requirements.
c. Specialized clinical scenarios: Certain models are designed for specific workflows, including longitudinal tumor monitoring (LA-ResUNet [20]), multimodal imaging integration (ResUNet–InceptionV4 [8], MEFNet [30]), and interventional planning applications (ConvNeXt-2U [7]).

These examples illustrate how architectural choices align with different clinical and operational needs, rather than indicating the universal superiority of any single model.

7.3. Human-in-the-loop integration

From the above research, it is clear that there is a trend of increasing interest in the integration of the automated segmentation models in the human-in-the-loop (HITL) clinical environment. In this system, the AI-based segmentations are provided as a first guess, which is then further reviewed by the expert radiologists. This type of collaborative system is beneficial in terms of clinical safety, accountability, and system improvement through expert feedback, especially in complex and ambiguous scenarios. The HITL integration of the AI-based system is considered a crucial part of the system in order to provide a safe and effective solution in the clinical environment.

8. CHALLENGES AND OPEN RESEARCH PROBLEMS

Although there have been significant advancements in the application of deep learning-based segmentation of the liver and liver tumors, there are still a number of challenges and research questions remaining, which limit their application in a clinical environment.

8.1. Data availability and annotation

The ability to build robust deep learning models depends in a crucial manner on the availability of large quantities of well-annotated medical imaging data. However, the creation of diverse multicenter data sets with consistent annotation is hindered by patient privacy concerns, imaging protocol variability, and the labor-intensive nature of manual annotation. As a result, the diversity of the data sets may be limited.

8.2. Computational complexity and resource constraints

There are many advanced models, especially those with the use of attention mechanisms, transformers, and ensemble approaches, which often call for considerable computational requirements during the training and evaluation process. Thus, there is always the need to develop models that ensure the tradeoff between segmentation accuracy and computational efficiency and tractability.

8.3. Generalization and domain adaptation

Generalization across different imaging protocols, scanner vendors, and patient demographics. The issue of generalization across diverse imaging protocols, scanner manufacturers, and patient populations still remains a major concern. Although various techniques, such as domain adaptation, transfer learning, and data augmentation, have shown promise, they are not equally effective across different scenarios.

8.4. Interpretability and explainability

The interpretability of deep learning models is a major roadblock to the adoption of deep learning models in the healthcare sector, considering that healthcare professionals require transparency and trust with regard to automated decision-support systems. Although recent research into uncertainty estimation and explainable artificial intelligence (XAI) has proposed promising avenues to explore, more research is necessary to ensure that the insights are interpretable without compromising segmentation accuracy.

8.5. Multi-modality and clinical workflow integration

The integration of multimodal information, such as positron emission tomography, magnetic resonance imaging, and genomic information, into existing computed tomography-based segmentation schemes has shown promise for increasing clinical relevance and precision. However, there are also some challenges associated with multimodal information fusion, including data harmonization, system complexity, and reliability. Seamless integration into existing clinical workflow, including picture archiving and communication system compatibility, is also a non-trivial problem.

8.6. Small lesion detection and boundary delineation

For the accurate segmentation of small and low-contrast lesions, as well as the precise delineation of tumor boundaries, various existing models are often challenged due to the limited spatial resolution and the lack of sufficient contextual information. Increasing the sensitivity to small lesions with low false-positive rates is the major focus of the current research.

8.7. Validation, standardization, and regulatory considerations

The successful clinical translation of deep learning models requires thorough validation through multicenter trials and conformance to relevant regulatory guidelines. The inconsistent design of the evaluation process and the lack of standardized reporting guidelines limit the ability to conduct an objective comparison of different studies. The development of standardized validation guidelines is a crucial aspect in the clinical translation of deep learning models.

9. FUTURE RESEARCH DIRECTIONS

While considerable advances have been achieved, the widespread clinical adoption of deep learning-based liver tumor segmentation is currently limited by a set of interrelated issues. In line with the analysis of 38 recent models, the focus of the research should shift beyond the incremental evolution of the architecture to more promising avenues that can help to ensure robust, reliable, and clinically viable AI systems.

9.1. Addressing data scarcity and annotation burden

The reliance of high-capacity models, especially those employing the transformer and hybrid architectures, on the availability of expertly annotated datasets is considerable. The process of manually segmenting tumors is a time-consuming process, and the inter-observer variability is a limiting factor in the scalability of the datasets. In addition, the available public datasets may not be sufficiently diverse and may not offer the required annotations to support the process of generalization. Future research directions include the development of semi-automated annotation tools, weakly supervised and self-supervised learning paradigms, synthetic data, and federated learning.

9.2. Rethinking the accuracy–computation trade-off

A major problem to be addressed is the balance between segmentation precision and computational cost. In fact, although high segmentation accuracy is often achieved by hybrid models of CNN and transformer, the high computational cost of these models hampers real-time usage and accessibility, particularly in resource-constrained situations. In contrast, many lightweight models often compromise segmentation precision in complex cases. In the future, it is important to explore model compression, knowledge distillation, neural architecture search (NAS), and hardware-aware optimization to achieve high segmentation accuracy with low computational cost.

9.3. Enhancing generalization and robustness

The limited ability to generalize across different imaging protocols, vendors, and patient groups remains a major challenge in establishing trust in the clinical community. There has been a need to improve the state of the art in domain adaptation, cross-validation, test-time adaptation, and invariant feature learning to improve the development of reliable image segmentation systems.

9.4. Bridging the interpretability and trust gap

Currently, the lack of transparency in deep learning models represents an obstacle for their application in the clinic. Clinicians require outputs that are transparent and interpretable, especially in cases that are on the borderline or pose a high level of risk. While attention mechanisms and saliency maps provide some level of insight, in the future, more emphasis should be placed on XAI techniques that provide clinically relevant explanations in addition to uncertainty quantification.

9.5. Designing for clinical workflow integration

In future segmentation systems, there should be a focus on clinical workflow. Besides accuracy, there are other practical considerations, such as the speed of inference on large 3D volumes, DICOM and PACS system compatibility, as well as user interface considerations to facilitate expert correction. Implementing a human-in-the-loop approach from the outset will be important to avoid disrupting clinical workflow.

9.6. Improving sensitivity to small and ambiguous lesions

In early stages, hepatic lesions are small, low in contrast, or diffusely infiltrative, making them difficult to segment. The sensitivity to such lesions, without increasing false positives, is still an active area of research. Promising research directions for this problem are multi-scale feature modeling, prediction with uncertainty, lesion-centric learning, and anatomical priors.

9.7. Advancing validation and regulatory readiness

In early stages, hepatic lesions are small, low in contrast, or diffusely infiltrative, making them difficult to segment. The sensitivity to such lesions, without increasing false positives, is still an active area of research. Promising research directions for this problem are multi-scale feature modeling, prediction with uncertainty, lesion-centric learning, and anatomical priors.

10. CONCLUSION

The field of medical image segmentation has undergone a significant paradigmatic shift with the advent of deep learning, particularly in managing liver cancer. This review systematically analyzed 38 state-of-the-art models published between 2023 and 2026, tracing the evolution from conventional CNNs to advanced hybrid CNN-transformer architectures. While significant improvements in segmentation accuracy and boundary delineation are evident, a clear 'accuracy-complexity' trade-off remains a primary barrier to clinical translation. To bridge the gap between experimental performance and real-world clinical utility, this review identifies four critical pillars for future research:

a. Model generalization and external validation: future designs must move beyond standard benchmarks like LiTS and 3DIRCADb. High performance on a single dataset does not guarantee reliability across different imaging protocols or diverse patient populations.
b. Computational efficiency: there is an urgent need for 'compute-aware' designs that maintain high Dice scores while reducing the video random access memory (VRAM) footprint, making them deployable on standard clinical workstations.
c. XAI and Interpretability: Integrating attention-map visualizations and uncertainty quantification is essential to transition from 'black-box' models to trustworthy clinical tools that radiologists can validate.
d. Federated and self-supervised learning: to address data scarcity and privacy concerns, shifting toward federated learning will allow models to learn from multi-institutional data without direct sharing of sensitive patient records.

Ultimately, the next generation of segmentation systems should be developed to support human-in-the-loop clinical settings, augmenting rather than replacing the expertise of radiologists. By focusing on these evidence-based directions, hybrid liver tumor segmentation can evolve from a technologically mature field into a clinically indispensable one.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Faculty of Computers and Artificial Intelligence, Fayoum University, Egypt, for providing the necessary research environment. We also thank the anonymous reviewers for their valuable feedback which significantly improved this manuscript.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ibrahim Mohamed Attiya	✓	✓	✓			✓		✓	✓					
Mostafa Thabet				✓	✓					✓		✓		
Mostafa R. Kaseb							✓			✓	✓	✓	✓	

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest regarding the publication of this research.

ETHICAL APPROVAL

This article does not contain any studies with human participants or animals performed by any of the authors. Therefore, no ethical approval was required.

DATA AVAILABILITY

The datasets (LiTS and 3DIRCADb) analyzed during the current study are available in the public domain. All data extracted from the reviewed studies are included in the manuscript's tables.

REFERENCES

[1] E. Goceri, "A hybrid attention-based deep learning model for segmentation of livers and liver tumors from CT scans," *Multimedia Tools and Applications*, vol. 84, no. 37, pp. 46191–46212, 2025, doi: 10.1007/s11042-025-20760-y.

[2] J. Wang, Y. Peng, S. Jing, L. Han, T. Li, and J. Luo, "A deep-learning approach for segmentation of liver tumors in magnetic resonance imaging using UNet++," *BMC Cancer*, vol. 23, no. 1, p. 1060, 2023, doi: 10.1186/s12885-023-11432-x.

[3] S. Saumiya and S. W. Franklin, "Unified automated deep learning framework for segmentation and classification of liver tumors," *Journal of Supercomputing*, vol. 80, no. 2, pp. 2347–2380, 2024, doi: 10.1007/s11227-023-05524-5.

[4] J. Sun, B. Wang, X. Wu, and others, "MAPFUNet: Multi-attention perception-fusion U-Net for liver tumor segmentation," *Journal of Bionic Engineering*, vol. 21, pp. 2515–2539, 2024.

[5] R. Xue, Z. Zhang, Y. Zhao, Q. Zhang, and M. Liang, "MMEFU-Net: A Mamba-guided multi-encoder fusion U-Net for tumor segmentation in CT images," *IEEE Access*, vol. 13, pp. 76257–76270, 2025, doi: 10.1109/ACCESS.2025.3564679.

[6] R. Jiao *et al.*, "RECISTSurv: Hybrid multi-task transformer for hepatocellular carcinoma response and survival evaluation," *IEEE Transactions on Image Processing*, vol. 34, pp. 3873–3888, 2025, doi: 10.1109/TIP.2025.3579200.

[7] G. Chen, H. Wang, Z. Lu, T. H. Wu, K. H. Lin, and G. S. P. Mok, "ConvNeXt-2U: A 3-D deep learning-based segmentation model for unified and automatic segmentation of lungs, normal liver and tumors in Y-90 radioembolization dosimetry," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 9, no. 4, pp. 468–477, 2025, doi: 10.1109/TRPMS.2024.3510587.

[8] H. Rahman *et al.*, "Automatic liver tumor segmentation of CT and MRI volumes using ensemble ResUNet-InceptionV4 model," *Information Sciences*, vol. 704, p. 121966, 2025, doi: 10.1016/j.ins.2025.121966.

[9] L. Yang, J. Zhang, T. Wang, Q. Feng, S. Fu, and M. Huang, "Multi-scale camouflaged feature mining and fusion network for liver tumor segmentation," *Engineering Applications of Artificial Intelligence*, vol. 148, p. 110398, 2025, doi: 10.1016/j.engappai.2025.110398.

[10] L. W. Tsai *et al.*, "RFiLM U-Net: Radiomic feature-integrated linear modulation network for precise liver tumor segmentation," *Journal of Medical and Biological Engineering*, vol. 45, no. 2, pp. 177–186, 2025, doi: 10.1007/s40846-025-00938-3.

[11] Y. J. Lee, S. W. Bang, J. Bin Hong, and S. Park, "Image-free tumor segmentation of soft tissue using a minimally invasive robotic palpation system," *IEEE Transactions on Biomedical Engineering*, vol. 72, no. 12, pp. 3621–3631, 2025, doi: 10.1109/TBME.2025.3573666.

[12] K. Vijayaprabakaran, P. Ramalingam, R. Ramalingam, A. Ilavendhan, and R. Vedhapriyavadhana, "CUNet-CLSTM: A novel fusion of CUNet and CLSTM for superior liver cancer detection in CT scans," *IEEE Access*, vol. 13, pp. 66373–66392, 2025, doi: 10.1109/ACCESS.2025.3559592.

[13] A. S. Anwar, K. Amin, M. M. Hadhoud, and M. Ibrahim, "ResTransUNet: A hybrid CNN-transformer approach for liver and tumor segmentation in CT images," *Computers in Biology and Medicine*, vol. 190, p. 110048, 2025, doi: 10.1016/j.compbiomed.2025.110048.

[14] Y. G. N and M. R V, "Automatic liver tumor classification using UNet70 a deep learning model," *Journal of Liver Transplantation*, vol. 18, p. 100260, 2025, doi: 10.1016/j.liver.2025.100260.

[15] P. M. Kumar, H. Gohel, J. Selvaraj, and B. P. Kavin, "Livernet based segmentation of lesions from computed tomography scan for liver tumor detection," *Intelligent Data Analysis*, vol. 29, no. 5, pp. 1289–1312, 2025, doi: 10.1177/1088467X241301660.

[16] Y. Yang, M. Sato, Z. Jin, and K. Suzuki, "Patch-based deep-learning model with limited training dataset for liver tumor segmentation in contrast-enhanced hepatic computed tomography," *IEEE Access*, vol. 13, pp. 86863–86873, 2025, doi: 10.1109/ACCESS.2025.3570728.

[17] T. Zhang, Y. Liu, Q. Zhao, G. Xue, and H. Shen, "Edge-guided multi-scale adaptive feature fusion network for liver tumor segmentation," *Scientific Reports*, vol. 14, no. 1, p. 28370, 2024, doi: 10.1038/s41598-024-79379-y.

[18] S. Randar, V. Shah, H. Kulkarni, Y. Suryawanshi, A. Joshi, and S. Sawant, "YOLOv8-based frameworks for liver and tumor segmentation task on LiTS," *SN Computer Science*, vol. 5, no. 6, p. 741, 2024, doi: 10.1007/s42979-024-03097-5.

[19] Y. H. Chuang *et al.*, "Effective tumor annotation for automated diagnosis of liver cancer," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 13, pp. 251–260, 2025, doi: 10.1109/JTEHM.2025.3576827.

[20] R. Jin, H. Y. Tang, Q. Yang, and W. Chen, "LA-ResUNet: Attention-based network for longitudinal liver tumor segmentation from CT images," *Computerized Medical Imaging and Graphics*, vol. 123, p. 102536, 2025, doi: 10.1016/j.compmedimag.2025.102536.

[21] A. Ben Slama, H. Sahli, Y. Amri, and S. Labidi, "V-NET-VGG16: Hybrid deep learning architecture for optimal segmentation and classification of multi-differentiated liver tumors," *Intelligence-Based Medicine*, vol. 11, p. 100210, 2025, doi: 10.1016/j.ibmed.2025.100210.

[22] A. Arora, R. Jordar, J. J. Jena, S. Singh, S. S. Patra, and M. K. Gourisaria, "Harnessing deep learning and transfer learning models for segmentation of liver tumors and veins," *Procedia Computer Science*, vol. 259, pp. 1874–1882, 2025, doi: 10.1016/j.procs.2025.04.143.

[23] G. N. Yashaswini, R. V. Manjunath, B. Shubha, P. Prabha, N. Aishwarya, and H. M. Manu, "Deep learning technique for automatic liver and liver tumor segmentation in CT images," *Journal of Liver Transplantation*, vol. 17, p. 100251, 2025, doi: 10.1016/j.liver.2024.100251.

[24] J. Zhao and S. Li, "Uncertainty-guided and cross-modality attention network for liver tumor segmentation and quantification via integrating dynamic MRI," *Knowledge-Based Systems*, vol. 310, p. 113021, 2025, doi: 10.1016/j.knosys.2025.113021.

[25] Z. Hu, H. Chen, L. Hua, X. Ren, and W. Mei, "MSML-AttUNet: A hierarchical attention network with multi-scale and multi-task for precision liver tumor segmentation," *Biomedical Signal Processing and Control*, vol. 99, p. 106861, 2025, doi: 10.1016/j.bspc.2024.106861.

[26] S. Dharaneswar and B. P. Santosh Kumar, "Elucidating the novel framework of liver tumour segmentation and classification using improved Optimization-assisted EfficientNet B7 learning model," *Biomedical Signal Processing and Control*, vol. 100, p. 107045, 2025, doi: 10.1016/j.bspc.2024.107045.

[27] S. Zhou, X. Lei, and L. Sun, "Liver image segmentation using a rotated variable-sized window attention mechanism: Application of the ARVSA u-net model," *Biomedical Signal Processing and Control*, vol. 108, p. 107954, 2025, doi: 10.1016/j.bspc.2025.107954.

[28] B. P. Pradeep Kumar, P. ~B. Rangaiah, and R. Augustine, "Improving liver cancer diagnosis: A multifaceted approach to automated liver tumor identification in ultrasound scans," *Next Research*, vol. 2, no. 3, p. 100465, 2025, doi: 10.1016/j.nexres.2025.100465.

[29] M. Balaguer-Montero *et al.*, "A CT-based deep learning-driven tool for automatic liver tumor detection and delineation in patients with cancer," *Cell Reports Medicine*, vol. 6, no. 4, 2025, doi: 10.1016/j.xcrm.2025.102032.

[30] Y. Qi, L. Lin, B. Zhang, J. Zhang, and J. Wang, "Multi-modal evidential fusion network for trustworthy PET/CT tumor segmentation," *Knowledge-Based Systems*, vol. 324, p. 113838, 2025, doi: 10.1016/j.knosys.2025.113838.

[31] S. E. Raja, J. Sutha, P. Elamparithi, K. J. Deepthi, and S. D. Lalitha, "Liver tumor prediction using attention-guided convolutional neural networks and genomic feature analysis," *MethodsX*, vol. 14, p. 103276, 2025, doi: 10.1016/j.mex.2025.103276.

[32] N. Kutaiba *et al.*, "Impact of local data from patients with chronic liver disease on accuracy of a liver and spleen CT segmentation model trained on a public dataset," *European Journal of Radiology Artificial Intelligence*, vol. 3, p. 100023, 2025, doi: 10.1016/j.ejrai.2025.100023.

[33] U. Bashir *et al.*, "Deep learning for liver lesion segmentation and classification on staging CT scans of colorectal cancer patients: a multi-site technical validation study," *Clinical Radiology*, vol. 85, p. 106914, 2025, doi: 10.1016/j.crad.2025.106914.

[34] R. Archana and L. Anand, "Residual u-net with Self-Attention based deep convolutional adaptive capsule network for liver cancer segmentation and classification," *Biomedical Signal Processing and Control*, vol. 105, p. 107665, 2025, doi: 10.1016/j.bspc.2025.107665.

[35] O. K. Sikha *et al.*, "Uncertainty-aware segmentation quality prediction via deep learning Bayesian Modeling: Comprehensive evaluation and interpretation on skin cancer and liver segmentation," *Computerized Medical Imaging and Graphics*, vol. 123, p. 102547, 2025, doi: 10.1016/j.compmedimag.2025.102547.

[36] J. Chakra Bortty, G. S. Chakraborty, S. Batra, J. Das, S. Das Polok, and P. Saha, "Implementation and performance evaluation of various deep learning-based segmentation approaches for liver tumor detection and diagnosis in medical imaging," in *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2025*, 2025, pp. 662–669. doi: 10.1109/ICICCS65191.2025.10985713.

[37] P. S. Lakshmi, D. Nagadevi, K. Suman, R. Deepthi, and N. Chikyal, "Deploying the model of improved heuristic-assisted adaptive SegUnet++ and multi-scale deep learning network for liver tumor segmentation and classification," *Journal of Real-Time Image Processing*, vol. 22, no. 1, p. 8, 2025, doi: 10.1007/s11554-024-01584-9.

[38] S. Hariharan, D. Anandan, M. Krishnamoorthy, V. Kukreja, N. Goyal, and S. Y. Chen, "Advancements in liver tumor detection: A comprehensive review of various deep learning models," *CMES - Computer Modeling in Engineering and Sciences*, vol. 142, no. 1, pp. 91–122, 2025, doi: 10.32604/cmes.2024.057214.

APPENDIX

As discussed in section 5, Table 3 provides an analytical overview of 38 recent studies (2023–2025) focusing on deep learning architectures for liver and liver tumor segmentation. The table details each model's architecture type, evaluation dataset, validation method, and reported Dice scores, alongside a summary of its key strengths and limitations.

Table 3. Comprehensive comparative analysis of deep learning-based liver and tumor segmentation models (*continue*)

Study	Model	Architecture type	Dataset	Validation method	Dice (L/T)	Strengths	Weaknesses
[1]	Hybrid Attention DL	Attention-hybrid	CT	5-fold CV	0.96/0.87	Attention-driven boundary refinement; high precision in heterogeneous regions	Significant GPU memory requirements; computationally intensive
[2]	UNet++	CNN	MRI	10-fold CV	0.95/0.84	Dense skip connections for multi-scale feature reuse; efficient for MRI data	Limited capacity for long-range dependency modeling

Table 3. Comprehensive comparative analysis of deep learning-based liver and tumor segmentation models (*continue*)

Study	Model	Architecture type	Dataset	Validation method	Dice (L/T)	Strengths	Weaknesses
[3]	Unified DL framework	CNN	CT	Hold-out (80/20)	0.94/0.85	Integrated segmentation and classification pipeline; streamlined workflow	Requires careful hyperparameter tuning across different datasets
[4]	MAPFUNet	Attention-CNN	CT	5-fold CV	0.97/0.88	Multi-attention mechanism effective for small and low-contrast tumors	High computational overhead during inference
[5]	MMEFU-Net	Multi-encoder fusion	CT	10-fold CV	0.96/0.87	Multi-encoder design with attention guidance; robust feature fusion	High memory footprint during training
[6]	RECISTSurv	Transformer	CT	Hold-out	N/A	Integrates segmentation with treatment response and survival prediction	Segmentation metrics not primarily reported; focused on prognostic tasks
[7]	ConvNeXt-2U	3D hybrid	CT (Y-90)	5-fold CV	0.95/0.87	Unified 3D segmentation of liver, lungs, and tumors; applicable to dosimetry	Demands high-end GPU resources; lengthy training time
[8]	ResUNet-InceptionV4	Ensemble	CT and MRI	5-fold CV	0.97/0.89	Multi-modal robustness; combines strengths of ResUNet and InceptionV4	Slow inference speed; computationally expensive ensemble
[9]	MCFM-Net	Multi-Scale CNN	CT	5-fold CV	0.95/0.86	Specialized for detecting camouflaged or iso-attenuating tumors	Less effective on homogeneous or well-defined lesions
[10]	RFiLM U-Net	CNN-radiomics	CT	5-fold CV	0.95/0.86	Radiomic feature integration enhances tumor characterization	Increased complexity in feature fusion and preprocessing
[11]	Robotic Palpation	Sensor-based	N/A	N/A	N/A	Image-free tumor localization; minimally invasive intraoperative use	Not an image-based DL model; limited to palpable surface lesions
[12]	CUNet-CLSTM	Hybrid CNN-LSTM	CT	5-fold CV	0.94/0.85	Captures spatiotemporal features; useful for 4D CT analysis	Complex training procedure with convergence challenges
[13]	ResTransUNet	Hybrid CNN-transformer	CT	5-fold CV	0.97/0.88	Effective fusion of CNN local features and transformer global context	High VRAM consumption; extended training duration
[14]	UNet70	CNN	CT (Liver)	5-fold CV	0.94/0.84	Lightweight architecture suitable for fast inference	Lacks attention or transformer components for global context
[15]	LiverNet	CNN	CT	5-fold CV	0.94/0.84	CT-specific lightweight model for efficient deployment	Lacks advanced attention or transformer-based modules
[16]	Patch-based DL	CNN-Patch	CT	5-fold CV	0.93/0.83	Suitable for limited data scenarios; reduces overfitting	May fail to capture global anatomical context
[17]	EGAF-Net	Edge-guided CNN	CT	5-fold CV	0.95/0.86	Edge-aware architecture improves boundary delineation	Performance degrades with noisy or low-quality CT acquisitions
[18]	YOLOv8-based	CNN-detection	CT (LiTS)	Hold-out	0.92/0.82	Real-time inference capability; efficient for screening applications	Lower tumor segmentation accuracy compared to dedicated segmentation networks

Table 3. Comprehensive comparative analysis of deep learning-based liver and tumor segmentation models (*continue*)

Study	Model	Architecture type	Dataset	Validation method	Dice (L/T)	Strengths	Weaknesses
[19]	Annotation framework	Annotation method	CT	N/A	N/A	Enhances dataset quality through semi-automated annotation tools	Relies on expert validation; not a segmentation model per se
[20]	LA-ResUNet	Attention-CNN	CT	5-fold CV	0.94/0.86	Designed for longitudinal tumor tracking across time-series scans	Requires precisely registered sequential imaging data
[21]	V-NET-VGG16	Hybrid CNN	CT	5-fold CV	0.95/0.85	Leverages VGG16 for stable hierarchical feature extraction	Limited sensitivity to small (<1 cm) tumors
[22]	Transfer DL	CNN + transfer learning	CT	Hold-out	0.93/0.83	Simultaneous segmentation of tumors and hepatic vasculature	Requires extensive annotation of both tumors and veins; complex training
[23]	DL Technique	CNN	CT	5-fold CV	0.94/0.84	Simple and efficient pipeline; easily deployable in clinical settings	Lacks advanced attention or global modeling components
[24]	UGCMANet	Attention, cross-modal	MRI	5-fold CV	0.95/0.86	Uncertainty-guided cross-modality attention; integrates dynamic MRI	Increased training complexity due to multi-modal fusion
[25]	MSML-AttUNet	Hierarchical attention	CT	10-fold CV	0.96/0.87	Hierarchical multi-scale, multi-task learning for precise segmentation	Computationally demanding; requires significant resources
[26]	EffNetB7 framework	CNN + optimization	CT	5-fold CV	0.93/0.83	Employs EfficientNet-B7 backbone with optimization techniques	Limited validation on large, diverse multicenter datasets
[27]	ARVSA U-Net	Attention-CNN	CT	5-fold CV	0.95/0.85	Rotating variable-size window attention improves boundary focus	Increased architectural complexity and implementation difficulty
[28]	US Tumor ID	CNN/ML	Ultrasound	5-fold CV	0.91/0.80	Practical for ultrasound-based screening in resource-limited settings	Lower accuracy compared to CT/MRI-based models; modality-specific
[29]	DL Tool	CNN	CT	Hold-out	0.94/0.84	Fully automated end-to-end detection and segmentation pipeline	Inherits general CNN limitations in global context modeling
[30]	MEFNet	Multi-modal fusion	PET/CT	5-fold CV	0.96/0.87	Evidence-based trustworthy segmentation using PET/CT fusion	Dependent on availability of co-registered PET/CT scans
[31]	Attention-CNN + Genomics	Attention + CNN	CT + Genomics	5-fold CV	0.95/0.86	Integrates imaging with genomic biomarkers for personalized analysis	Dependent on availability of genomic data; multi-modal complexity
[32]	Local data enhanced model	CNN	CT (Liver/Spleen)	5-fold CV	0.94/0.85	Demonstrates impact of local dataset enrichment on performance	Requires external validation across multiple clinical sites
[33]	DL lesion segmentation model	CNN	CT (Colorectal)	Hold-out	0.95/0.86	Multi-site technical validation; useful for staging	Complex annotation protocol for colorectal cancer metastases

Table 3. Comprehensive comparative analysis of deep learning-based liver and tumor segmentation models (*continue*)

Study	Model	Architecture type	Dataset	Validation method	Dice (L/T)	Strengths	Weaknesses
[34]	Residual U-Net + Capsule Net	CNN + Self-attention	CT (Liver)	5-fold CV	0.96/0.87	Capsule network integration for better spatial relationship modeling	High computational cost due to capsule network overhead
[35]	Bayesian DL quality model	Bayesian DL	Skin/Liver CT	Hold-out	N/A	Uncertainty-aware segmentation quality prediction	Complex post-processing interpretation pipeline
[36]	Multiple DL approaches	CNN/Ensemble	CT	5-fold CV	0.93/0.85	Comparative study of various DL methods; practical insights	Broad analysis without a single novel model contribution
[37]	Adaptive SegUNet++ hybrid	Hybrid multi-scale DL	CT	5-fold CV	0.95/0.86	Heuristic-assisted adaptive learning with multi-scale U-Net++	Complex training pipeline with high resource demand

Note: *Performance metrics (Dice scores) reported in this table are subject to variability due to different data preprocessing techniques, the use of private vs. public datasets, and varying validation strategies (e.g., k-fold cross-validation vs. fixed test sets).*
*"**Note:** Table 3 includes 37 experimental studies; reference* [38] *is a comprehensive survey used for contextual comparison."*

BIOGRAPHIES OF AUTHORS

Ibrahim Mohamed Attiya received the M.Sc. degree in information systems from Fayoum University, Fayoum, Egypt, in 2023. He is currently pursuing the Ph.D. degree in computer science and artificial intelligence with the Faculty of Computers and Artificial Intelligence, Fayoum University, Egypt. His research interests include deep learning, medical image segmentation, computer vision, and artificial intelligence applications in healthcare. He is currently working on hybrid deep learning architectures combining NNU-Net and transformer models for liver tumor detection and segmentation in CT images. He serves as an assistant lecturer in the same faculty, where he contributes to teaching courses in artificial intelligence and data science. Mr. Attiya has served as a reviewer for several international conferences on medical image analysis and artificial intelligence. He can be contacted at im1959@fayoum.edu.eg.

Mostafa Thabet received his Ph.D. in information systems. He is currently an associate professor of information systems with the Faculty of Computers and Artificial Intelligence, Fayoum University, Fayoum, Egypt, and serves as the dean of the Faculty of Computer Science, Nahda University, Beni Suef, Egypt. His research interests include artificial intelligence, information systems, data science, and their applications in various domains. With over 15 years of academic experience, he has supervised numerous graduate students and contributed to several research projects in computational intelligence and information systems. He has published extensively in international journals and conferences. He started M.B.A. studies at the Arab Academy for Science Technology and Maritime Transport (AASTMT). Prof. Thabet serves as the editor-in-chief of [Alfajr]. He has received several awards for his contributions to computer science education and research in Egypt. He can be contacted at mtm00@fayoum.edu.eg.

Mostafa R. Kaseb received his Ph.D. in computer science. He is currently an associate professor of computer science with the Faculty of Computers and Artificial Intelligence, Fayoum University, Fayoum, Egypt. His research interests include artificial intelligence, machine learning, computer vision, and their applications in healthcare and medical image analysis. He has extensive experience in supervising graduate research and has contributed to numerous publications in the field of computational intelligence and data science. Prof. Kaseb actively participates in academic review processes for several international journals and conferences. He can be contacted at mrk00@fayoum.edu.eg.