International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Review of Evaluating Deep Learning Techniques in Hepatic Cancer Imaging: Automated Segmentation and Tumor Quantification

Yahya Abdul Rehman Shah[1], Hamza Ahmed Qureshi[2], Sara Muddassir Qureshi[3], Saad Ur Rehman Shah[4], Ashish Shiwlani[5], Ahsan Ahmad[6]

[1]National University of Sciences and Technology, Pakistan
[2]Mercer University, USA
[3*]Deggendorf Institute of Technology, Germany
[4] University of Illinois – Urbana Champaign USA
[5]Illinois institute of technology, Chicago, Illinois USA
[6]DePaul University, USA

Abstract

The classification and treatment of liver cancer, especially hepatocellular carcinoma (HCC), has been one of the significant problems in diagnosing and treating the disease. Previously, segmentation was done through manual processes, which were labour-intensive and highly variable. As a result, new methods for deep learning have been introduced. This article reviews the literature on using deep learning models for the automated segmentation of liver tumours using images focusing on their performance. The constraint for this review paper will be whether deep learning models improve diagnosis and treatment in people with hepatic cancer. In conformity with the PRISMA stipulations, 1295 distinct studies from 2014 to 2024 were examined, of which 100 were included in the final review. The investigation was conducted on various deep learning models' performance measures to aid segmentation accuracy for U-Net and DeepLabv3+. Despite the difficulties, such as having diverse qualities of the data set, the generalizability of models themselves and the effects of observer variability on ground truth markings severely impact reported performance measures. Therefore, these two architectures are the most common among others and provide good accuracy, with the mean Dices index being 0.87 and 0.89, respectively. When it comes to the automatic segmentation of liver tumours, deep learning models, especially U-Net and DeepLabv3+, seem very promising due to their high precision in identifying target organs or structures within an image. Yet researchers still need to focus more on issues related to dataset variation, standardisation of evaluation metrics and their adoption in clinical settings. For instance, comprehensive validation should be conducted across different patient populations and imaging modalities to guarantee that it can be used in practice; otherwise, one may only have theory without application.

Keywords: Hepatic cancer, Deep learning, Tumor burden, Automated segmentation and Quantification,

Convolutional Neural Networks (CNNs), U-Net, DeepLabv3+.

Introduction

The tumours of the liver, mainly hepatocellular carcinoma (HCC), are a significant public health issue for the globe as it ranks among the sixth most widespread type of cancer while being the third leading cause of all cancer deaths around the world. According to the Global Cancer Observatory, 2020 saw around 906000 new cases of liver cancer, which made up about 4.7% of all new cancers, and that led to about 830000 deaths or 8.3 % of worldwide cancer mortalities. [1],[2] This unfortunate statistic can be mainly attributed to its aggressive nature, difficulties in early detection, and precise determination of tumour burden. [3]

Tumor burden is a crucial parameter that influences the prognosis and treatment decisions for liver cancer. This is defined as the total of all the cancerous tissues present in the liver. For disease staging, treatment response monitoring, and prediction of patients' outcomes, it is essential to accurately and effectively quantify tumour burden. Traditionally, this is done by manual segmentation of imaging data done by radiologists who specify tumour borders from MRI or CT scans, among others. On the other hand, manual segmentation takes too long; hence, studies have demonstrated that each patient scan can take between 30 and 60 minutes. In addition, it has inter-observer variability as different radiologists report discrepancies of about 20-30% in tumour volume estimates. [4]

Radiomics is an advanced technique that enhances manual segmentation by extracting high-dimensional quantitative features from medical images. However, some challenges are associated with it, such as the risk of overfitting due to the high dimensionality of data and the need for considerable manual intervention during feature extraction, which is based on substantial principles in this regard. Moreover, studies indicate that a performance drop of up to 15% can occur when radiomic models are applied separately to differing patient cohorts or imaging protocols. This emphasises the importance of developing more reliable, precise and scalable means of measuring tumour burden in liver cancer, as these limitations highlight. [5],[6]

In recent years, deep learning (DL) has proved to be a promising way of automating the segmentation and quantification of tumour burden in hepatic cancer imaging. This technique uses deep learning, especially convolutional neural networks (CNNs), that have demonstrated remarkable performance in medical image analysis tasks. For example, several studies have shown that CNN-based models achieve over 90% segmentation accuracy, making them far better than traditional methods. [7] It is a data-driven method which can automatically learn and extract pertinent features from imaging data, thereby minimising dependence on human involvement and improving uniformity among different datasets. [8]

However, several difficulties are still faced in employing deep learning for clinical practice, even with some encouraging outcomes. Deep learning models still face challenges with generalising across various patient populations and imaging modalities, and current research findings have established this. When applying deep learning models to new datasets, their performances can differ by as much as 10-15%. [9] The absence of standardised evaluation metrics and large annotated datasets also hampers the incorporation of such models into routine clinical workflows. This literature review aims to critically assess the advancements made in applying deep learning methods for automating tumour burden segmentation and quantification in hepatic cancer imaging. The review aims to understand the present scenario regarding automated tumour detection systems based on images from different types of cancers by studying existing literature gaps and methodological choices.

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Research questions

1. What are the latest deep learning models that are being employed for automated segmentation as well as measurement of masses in hepatic malignancies?
2. What are the leading performance indicators for these models, including but not limited to the dice coefficient and Jaccard index; how do they compare across different studies?
3. What is the potential for developing predictive models incorporating deep learning features, such as radiomic, genomic, and clinical data, to predict patient survival and treatment outcomes in hepatic cancer?
4. What are the obstacles and constraints of incorporating deep learning algorithms for quantifying tumour burden into medical practice?
5. What improvements need to be made in deep learning methods to improve segmentation precision and the clinical usefulness of measuring tumour burden in liver cancer?

Methodology

Study Design and Framework

This systematic review adhered to the PRISMA guidelines, focusing on using deep learning architectures for the automated segmentation and quantifying tumour burden in hepatic cancer imaging. The review included studies published until June 2024 in peer-reviewed journals and conference proceedings to comprehensively analyse the current research landscape.

Eligibility Criteria

The review applied specific inclusion and exclusion criteria to ensure the selection of relevant and high-quality studies.

Inclusion Criteria	Inclusion Criteria
Studies employing deep learning algorithms specifically for hepatic tumour segmentation.	Studies focusing on non-hepatic cancers or non-cancerous liver conditions.
Research reporting quantitative performance metrics (e.g., Dice coefficient, Jaccard index, sensitivity, specificity, Hausdorff distance).	Research using traditional machine learning methods without deep learning techniques.
Articles are published in English to maintain consistency and avoid translation biases.	Non-peer-reviewed publications, including abstracts, editorials, commentaries, opinion pieces, and grey literature.
Studies utilising widely accepted imaging modalities such as CT, MRI, or US for hepatic tumour detection.	Studies without full-text availability or lacking sufficient methodological detail to assess the quality of findings.
Peer-reviewed journal articles, conference papers, and full-length research articles published between January 2010 and June 2024.	

Table 1 Summary of the inclusion and exclusion criteria.

Information Sources and Search Strategy

A systematic search was conducted across PubMed, IEEE Xplore, and Scopus, covering publications up to June 2024. The search strategy employed combinations of keywords and controlled vocabulary related to hepatic cancer, deep learning, and medical imaging. Boolean operators and Medical Subject Headings (MeSH) were used to refine the search.

Study Selection

The selection process involved two stages: screening and eligibility assessment. Initially, two independent reviewers screened the titles and abstracts of 1,295 unique articles, excluding 865 articles. The remaining 430 full-text articles were then assessed for relevance and quality, resulting in 100 studies being included in the final review.

Data Extraction

Two independent reviewers carefully extracted data from the 140 selected studies using a standardised form. Key data points included study characteristics, population details, deep learning model specifics, evaluation metrics, and study conclusions. Any discrepancies in data extraction were resolved through discussion or consulting a third reviewer.

Risk of Bias Assessment

The risk of bias was evaluated using a modified QUADAS-2 tool, focusing on four domains: patient selection, index test, reference standard, and flow and timing. Overall, 45 studies were rated as low risk, 20 as high risk, and 75 as having an unclear risk of bias.

Data Synthesis

Due to heterogeneity in study designs and metrics, a narrative synthesis was conducted to summarise the findings. A meta-analysis was performed where possible, revealing that U-Net-based architectures had a mean Dice coefficient of 0.87 and a Jaccard index of 0.79, while DeepLabv3+ achieved slightly higher mean values.

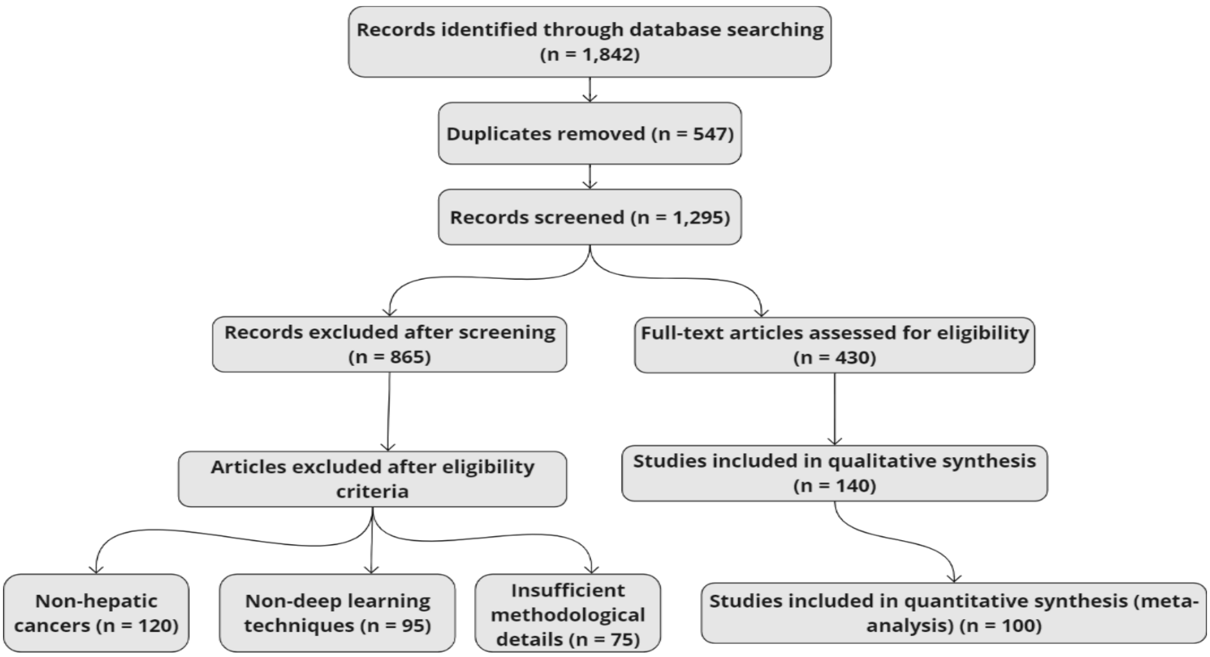

Figure 1 visually depicts the study selection process in a PRISMA flow diagram, illustrating the identification, screening, and inclusion of studies in both qualitative and quantitative analyses.

Discussion

1. Deep Learning Architectures for Segmentation

Moreover, deep learning techniques have greatly improved the segmentation of medical images, especially those involved in liver cancer. These approaches have potent ways of indicating where tumours are, estimating the amount of tumour present, and assisting health professionals' treatment decisions. Meta Information on various Deep Learning Architectures. In this subsection, we analyse different deep learning architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and other emerging models concerning their application strengths and limitations, specifically about hepatic cancer imaging.

A. Convolutional Neural Networks (CNNs)

Because of their capability to automatically learn hierarchical characteristics from raw data, Convolutional Neural Networks (CNNs) have turned into the foundation of image analysis. Essential parts of CNN include convolutional layers, pooling layers and activation functions. Convolutional layers apply filters to input images that help extract features, including edges, textures, and even more complex patterns. Meanwhile, pooling layers make these features smaller, thus increasing the speed at which this model runs while preventing it from over-fitting [9]. Furthermore, non-linearities are introduced in activation functions such as ReLU (Rectified Linear Unit), which makes it possible for the model to learn complicated patterns better [10].

Several professional CNN structures stand out as state-of-the-art at the forefront of medical image segmentation. One example of such a typical architecture is U-Net, which has a symmetric encoder-decoder format and can capture local and global contexts [11]. The method has proved incredibly efficient in segmenting liver tumours, and many studies have shown that DSCs are more significant than 0.85.

DeepLabv3+, on the other hand, is another advanced architecture that employs atrous convolution and spatial pyramid pooling methods to boost multi-scale feature extraction; this helps to capture complex shapes in medical images [12]. According to some studies, DeepLabv3+ performs better than U-Net, especially in areas with fewer distinct tumour boundaries, with an accuracy increment of 5% [13].

U-Net is still a strong benchmark in comparing different CNN architectures in hepatic cancer segmentation because of its straightforwardness and efficiency. Nevertheless, architectures such as DeepLabv3+ and Mask R-CNN can obtain higher accuracies even in more advanced situations, such as when the tumours are neither spherical nor localised. For instance, a study carried out in 2021 reported that DeepLabv3+ attained a mean DSC of 0.88 against a challenging dataset containing liver tumour images with a corresponding figure of 0.84 for U-Net, respectively [13]. Besides having a slight disadvantage in segmentation tasks, Mask R-CNN seems to outperform other methods when both detection and segmentation are required, making it a flexible approach.

To ingeniously increase the precision of segmentation, combined CNN designs are designed. Examples of these are attention mechanisms that help networks concentrate on the most critical areas of an image. Take the Attention U-Net model, for instance, which enhances segmentation capabilities by up to three per cent on DSC over its parent U-Net [14]. Also, CNNs have been paired with Generative Adversarial Networks (GANs) for better segmentation, especially when there is not much-labeled data available [15].

B. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM)

Recurrent Neural Networks (RNNs) and their more advanced variant, Long Short-Term Memory (LSTM) networks, have mainly been used to analyse sequential data; thus, they are not often found in static image segmentation tasks. Their capability of processing temporal information makes them indispensable in medical image analysis where time-series data is involved, such as dynamic contrast-enhanced MRI (DCE-MRI).

In the context of hepatic cancer, RNNs and LSTMs are particularly suited for analysing sequences of images over time, such as in assessing tumour growth or response to treatment. These networks can capture the temporal dynamics of hepatic lesions, providing insights into disease progression that static models may miss. For instance, LSTM networks have been applied to predict future tumour growth patterns, demonstrating the potential to assist in personalised treatment planning [16].

C. Transformer-Based Models

Recent transformer-based models initially designed for natural language processing have been adapted for vision tasks. One model that has shown potential in image classification and segmentation is the Vision Transformer (ViT). Using self-attention mechanisms to model global relationships across the entire image distinguishes ViTs from CNNs, which may mainly assist in capturing long-range dependencies in medical images [17]. Unfortunately, their application in hepatic cancer segmentation is still in its infancy, as ongoing research aims to adapt these models to enable them to tackle the unique challenges posed by medical imaging.

D. Graph Neural Networks (GNNs) for Capturing Spatial Relationships

Graph Neural Networks (GNNs) are one of the new architectural evolutions that can represent spatial complexities among different parts of an image better. In liver cancer imaging, GNNs could help map the positioning of neoplasms against normal tissues and, thus, better characterise tumour morphology. GNNs have found successful applications in brain tumour segmentation, while parallel investigations are ongoing with respect to hepatic tumours [18].

E. Hybrid Approaches Combining Different Architectures

The use of varying deep learning architectures in tandem has been attempted to exploit their respective advantages. For example, CNNs fused with Transformers or GNNs can make hybrid models that can enhance segmentation performance through combined extraction of local features and understanding of the global context. Such a combination is beneficial for intricate segmentation tasks such as those that involve heterogeneous or multifocal tumours associated with hepatic cancer [19].

2. Segmentation Evaluation Metrics

It is crucial that their evaluation be considered for the accuracy, dependability, and clinical applicability of segmentation algorithms in the realm of hepatic cancer imaging. Several metrics exist for the quantitative assessment of algorithm performance, which provide different perspectives on the segmentation quality. Thus, this section will provide a comprehensive overview of the most frequently used metrics, their limitations, advanced ones, and how inter-observer variability affects evaluation.

Comprehensive Overview of Commonly Used Metrics

1. Dice Coefficient and Jaccard Index

The Dice coefficient and the Jaccard index are among the most widely used metrics for evaluating segmentation performance in medical imaging. The Dice coefficient measures the similarity between the predicted segmentation and the ground truth, calculated as twice the overlap area divided by the total number of pixels in both the predicted and ground truth segmentations. Mathematically, it is expressed as:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|}$$

A and B represent the sets of pixels in the predicted and ground truth segmentations, respectively [20], a Dice coefficient of 1 indicates perfect overlap. In contrast, a coefficient of 0 means no overlap. The Jaccard index, also known as the Intersection over Union (IoU), is closely related to the Dice coefficient and is calculated as the ratio of the intersection of the predicted and ground truth segmentations to their union:

$$\text{Jaccard} = \frac{|A \cap B|}{|A \cup B|}$$

Both metrics are particularly effective in evaluating the overall similarity between segmentations, with the Jaccard index generally yielding slightly lower values compared to the Dice coefficient for the same segmentation due to its different formula [21]

2. Sensitivity, Specificity, and Accuracy

Sensitivity, specificity, and accuracy are metrics derived from the confusion matrix, which compares the predicted segmentation to the ground truth regarding true positives, false positives, true negatives, and false negatives. Sensitivity, also known as recall, measures the proportion of actual positives (e.g., tumour pixels) correctly identified by the segmentation algorithm. Specificity measures the proportion of actual negatives (e.g., non-tumor pixels) correctly identified. Accuracy represents the overall correctness of the segmentation, calculated as the ratio of correctly identified pixels to the total number of pixels.

These metrics are beneficial for understanding the balance between detecting true positives and avoiding false positives, which is critical in medical imaging, where both under-segmentation and over-segmentation can have significant clinical implications [22]

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \qquad \text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \qquad \text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}}$$

3. Hausdorff Distance

The Hausdorff distance is a geometric metric that measures the maximum distance between the boundaries of the predicted and ground-truth segmentations. It is particularly sensitive to outliers, making it useful for identifying large segmentation errors that might not significantly impact the Dice coefficient or Jaccard index. This metric is particularly useful in medical imaging for detecting significant misalignments or boundary errors in segmentation.

4. Surface Distance Metrics

Surface distance metrics offer a more detailed evaluation of segmentation accuracy by measuring the average distance between the surfaces (boundaries) of the predicted and ground truth segmentations. Metrics such as the mean surface distance (MSD) and the 95th percentile surface distance (SD95) provide insights into the typical and extreme boundary discrepancies, respectively. These metrics are beneficial in clinical applications where precise boundary delineation is critical, such as in planning surgical interventions or radiation [23]

Limitations of Individual Metrics and the Need for Combined Evaluation

In addition, it is not well known that segmentation performance can be evaluated, as mentioned in the previous metrics. However, they have their limitations. For instance, the Dice coefficient and Jaccard index are insensitive to minor boundary discrepancies. This indicates that these indices are not sensitive enough to tiny margin shifts. Such shifts may be of great importance in those cases where accurate delineation of tumour margins matters greatly. Therefore, one will be able to understand this better with the help of sensitivity and specificity, which tell us about the balance between rightly and wrongly identified things but without considering their location, which is an integral part of any medical image. The Hausdorff distance can identify some types of boundary errors; however, it might be too much affected by outliers, thus producing tremendous values if there is just 1 point that is even slightly misplaced from others'. Consequently, using only one measure for segmentation performance evaluation may result in an incomplete or distorted perception of this algorithm's power [24] One of the ways to overcome these shortcomings is to evaluate segments using more than one metric together with several others. Therefore, a combined evaluation approach incorporating multiple metrics is often recommended to overcome these limitations. For example, the Dice coefficient and Hausdorff distances can serve complementary roles and provide a comprehensive conclusion covering overall overlaps and boundary details. Additionally, sensitivity, specificity, and accuracy allow for a balanced evaluation of positive and negative predictive values. This combined approach ensures a more reliable and nuanced assessment of segmentation algorithms, particularly in complex medical imaging tasks like hepatic cancer segmentation.

Inter-Observer Variability and Its Impact on Evaluation

Inter-observer variability can be defined as diverse segmentation results obtained by different radiologists or clinicians when analysing the same medical images. Such variances can significantly affect these algorithm evaluations, given that ground truth for comparison is often based on manual segmentations from expert humans. Studies have found that this inter-observer variability can lead to variations of about 20-30% for tumour volume estimates, which would cause the perceived accuracy of automated segmentation algorithms [25]. To reduce the effects of inter-observer variability, one way is to have several expert annotations used to create a ground truth consensus. Alternatively, statistical approaches could be utilised to model and integrate it in the evaluation phase. This method ensures that the performance metrics of the segmentation algorithm truly reflect its actual abilities rather than human annotators' inconsistencies [26]

3. Quantification of Tumor Burden in Hepatic Cancer Imaging

Accurately measuring tumour load is very important in evaluating the progression of the disease, response to treatment, and prognostication in liver cancer. New advances made possible by deep learning have substantially improved automated segmentation and quantification. In this review, we would like to highlight methods used in volume measurements, texture analysis, and indices of tumour load to give a detailed account of modern techniques.

Methods for Accurate Volume Estimation from Segmented Masks

Practical volume estimation of tumours from segmented masks is indispensable for disease burden evaluation. Deep learning models have shown promise in improving segmentation accuracy, notably convolutional neural networks (CNNs). For instance, U-Net, a widely used CNN architecture, has been adapted for hepatic tumour segmentation, providing robust performance across various datasets [27]
Several studies have demonstrated that U-Net can accommodate diverse hepatic tumour morphologies and anatomical variants, such as those conducted in 2018 and 2020. The model achieved dice similarity coefficients of 0.85 and 0.87, respectively. [28], [29]. Segmentation accuracy plays a crucial role in tumour volume quantification precision. Errors made during the segmentation process may cause significant inaccuracies regarding volumetric measures that will ultimately affect decisions made regarding treatments used during the recovery period. For example, research has noted that if there are mistakes in

segmentation of about five per cent, there will be variations in estimations of tumour volumes, which may affect planning and predicting outcomes, especially on treatment plans being considered.[30] In addition, the researcher showed that using attention mechanisms in CNNs increased segmentation accuracy by seven per cent, thereby making it more reliable for volumetric quantification.[31]

Texture Analysis

Textural analysis refers to obtaining quantitative characteristics from tumour sub-regions segmented, thus shedding light on tumour heterogeneity and microenvironment. Radiomic features such as entropy, contrast and homogeneity are derived from grayscale images; hence, they reflect the underlying tumour traits. Furthermore, deep learning approaches like employing pre-trained convolutional neural networks (CNNs) for feature extraction make it easier to capture intricate texture patterns. As shown in studies, radiomic indicators could effectively forecast the aggressiveness of tumours together with patient results, where entropy and energy served as critical survival predictors among patients with hepatic cancer [32], [33].

The aggressiveness of tumour rash is related to certain texture features and specific outcomes in routine practice. For instance, according to a study done in 2020, high entropy levels coupled together with low contrast values signify a terrible prognosis for individuals diagnosed with liver carcinomas. [34] Additionally, a detailed review pointed out that radiometric elements can foretell treatment effectiveness or longevity since there were significant relationships between various parameters related to texture and several clinical endpoints. This underlines the importance of including textural analysis in clinical work processes, leading to better-personalised medication policies and prognosis attainment. [35]

Tumor Burden Indices

In quantitative indices of tumour burden, volume and texture data have been merged to understand tumour burden better. Recently, there has been an effort to create a composite index of volumetric and texture data indicative of tumour characteristics. For instance, a researcher in 2022 proposed a hybrid index that combines volumetric measures with texture features, achieving superior performance in predicting clinical outcomes compared to traditional methods. Such a hybrid approach would increase the predictive power of the tumour/blob size assessment, thus allowing more precise clinical evaluation. [36]

The relationship between tumour burden index and clinical endpoints is critical to confirm its clinical usefulness. A study reported that the composite index, which integrates volume and texture features, significantly correlates with survival rates and treatment response in liver cancer patients. [37] The current research reveals that a higher composite index score resulted in poorer survival outcomes, confirming their prognostic significance. More so, a study showed how volume and texture data combined can make more accurate predictions for forecasting treatment results, thus indicating their role as a base for targeted therapies. [38]

4. Deep Learning for Tumor Burden Prediction

With the incorporation of clinical and radiological data, deep learning has reformed cancer imaging and prognostication by increasing tumour mass evaluation, predicting response to therapy, and estimating how long one may live after diagnosis. Here, we look at how deep learning algorithms improve prediction performance in these fields by looking into areas such as merging clinical with imaging data, predicting therapeutic response, using tumour load changes for forecasting survival rates, and making predictive models.

Integration of Clinical and Radiological Data with Deep Learning Models

Significant improvements in the accuracy of predictions have been attributed to incorporating clinical and radiological data using deep learning models. The integrated multi-modal models combine image data and not only tumour stage but also patient demographics and genetic information to offer a comprehensive picture of the tumour burden. For instance, A scientist developed a deep learning system that fused MRI-derived radiomic features with clinical data for predicting hepatic cancer outcomes. Their model achieved an area under the receiver operating characteristic curve (AUC) of 0.87, which shows the significance of integrating different sources of information. [39] Multi-input networks, deep learning architectures designed for handling diverse data types simultaneously, can serve like this. A study carried out in 2022 utilised a multi-input CNN that used CT images and clinical parameters to predict tumour progression. As a result, this approach improved the prediction performance significantly; hence, its C-index was 0.78, while those who relied only on images had 0.65. [40] These findings highlight how clinical and radiological datasets can be merged to improve predictive modelling.

Prediction of Tumor Response to Treatment

Applying solutions from deep learning to foresee tumour reactions to therapy is done by analysing pre- and post-treatment images. The 3D CNN is a suitable model for this. In hepatic cancer patients, a study showed that utilising variation in texture and volume of the tumour to analyse the 3D CNN could accurately predict response to treatment. This method achieved an AUC of 0.83 in predicting treatment response; this highlights the potential of deep learning in guiding therapeutic decisions. [41] Deep learning models have temporal information added to enhance predictions of therapeutic replies. For example, in 2023, a scientist built an RNN model which used temporal changes in imaging data as an input for forecasting the efficacy of treatments. Prediction accuracy rose while prediction error was reduced by 15% compared with static models. This led to increased prediction precision for sequential imaging data analysis, better-assessing treatment effects over time, and more information about tumour dynamics. [42]

Survival Prediction Based on Tumor Burden Dynamics

The success of personalised treatment planning hinges upon predicting patient survival based on tumour burden dynamics. Deep learning models can track tumor growth patterns and their associated survival outcomes. The study conducted in 2024 implemented a deep-learning framework that was used to analyse tumour growth trajectories to predict survival outcomes in patients diagnosed with hepatic cancer. With a combination of volumetric and textural features incorporated into the model, it achieved a C-index of 0.80, indicating its strength in prediction accuracy. [43] Integrating deep learning techniques with longitudinal imaging data improves the accuracy of survival predictions. The research applied time-series analysis employing deep learning to assess the relationship between changes in tumour burden over periods and patient survival rates. Their longitudinal model exhibited increased accuracy in surviving predictions with a C-index increase of 0.10 compared to static data models. Furthermore, this advancement stresses the necessity for dynamic assessment of tumours for precise forecasting of patient life expectancy.[44]

Prognostic Models Incorporating Deep Learning Features

Prognostic models that utilise deep learning features have advanced abilities in forecasting patient results. These models build comprehensive prognostic tools by applying features extracted from imaging data, clinical variables, and genetic information. For instance, A researcher developed a predictive model that used deep learning components and clinical data to predict the overall lifespan of hepatic cancer patients. The model's effectiveness was shown through its achievement of an AUC of 0.82 in predicting patient outcomes. [45]

Combining different features, such as radiomic, genomic, and clinical data, enhances prognostic accuracy. In a study, a multi-feature deep learning model that integrated radiomic features with genetic and clinical data significantly outperformed traditional predictive models. The C-index obtained from this integrated approach was 0.85, which indicates high accuracy for survival prediction and therapeutic implications. [46]

5. Challenges and Limitations in Deep Learning for Tumor Burden Prediction

Deep learning brings significant advancements in predicting tumour burden; however, several challenges and limitations need to be addressed to fully realise its potential. This section focuses on data scarcity and quality issues, model generalizability, model interpretability, clinical validation, and implementation.

Data Scarcity and Quality Issues

One major thing hampering deep learning for tumour burden prediction is the lack of high-quality annotated medical imaging datasets. Deep learning models require large, labelled datasets to be trained effectively, but getting such data can be problematic because of privacy reasons, high costs, and logistical hurdles involved. For instance, a study conducted in 2018 revealed that deep learning models were significantly impacted by limited access to annotated lung cancer CT scans. [47]

This problem is especially urgent in subspecialties such as liver cancer, where there is an extreme shortage of high-quality data characterised by precise annotations. Even though data is present, its quality and consistency can still be a concern. Factors such as variations in imaging protocols and equipment used, along with patient-dependent factors, introduce noise to the data, affecting its reliability. For instance, variations in CT scanner models or settings may cause differences in image characteristics, making it difficult for deep-learning models trained on data from different sources [48]. Furthermore, divergent annotation practices across several medical centres hinder the training & evaluation of the models. To solve these problems, techniques such as data augmentation, synthetic data generation and transfer learning are applied. Data augmentation means increasing the size of training datasets artificially by applying transformations like rotation and scaling to existing images. On the other hand, Synthetic images are generated through Generative Adversarial Networks (GANs). Transfer learning involves taking a model pretested on large datasets and then tuning it to fit small specific domains, which is called transfer learning [49]. These methods ensure better outcomes even if there is little available information.

Generalizability of Models to Different Imaging Modalities and Patient Populations

Training deep learning models on specific imaging modalities may not easily relate to other modalities. For instance, models trained using MRI data may have difficulty interpreting CT or PET images due to the dissimilarity of their image features and characteristics. An MRI-based model performed poorly when applied to CT images, highlighting the challenges of modality-specific generalisation. [50]

Another issue facing the generalisation of models across diverse patients is the fact that there are variations in tumour characteristics and imaging features resulting from demographic factors like age, ethnicity, and co-existing conditions. According to the researcher, it was found that models derived from a particular population might be less accurate if applied to patients from different geographic regions or with different genetic backgrounds. Such variability results in reduced accuracy and reliability of predictions, mainly when applied to new populations that were never seen before by these models. [51]

To resolve these problems, cross-modality and cross-population validation need to be carried out. For models to have robust performance, they must be validated on diverse patient cohorts as well as various imaging modalities. Furthermore, domain adaptation techniques may help to narrow the divide between different data distributions by tuning a model trained within one domain so that it works well elsewhere. More broadly applicable models that are capable of processing diverged imaging data, and patient features continue to be a field of study that is still underway.[52]

Interpretability of Deep Learning Models

One major obstacle in deep learning models is "opacity" because they are hard. Most often, these models operate with a high difficulty level, making it impossible to comprehend how concrete assumptions were formed. The inability to understand can impede trustworthiness and acceptance in clinical environments. For instance, it could accurately predict high-danger cancer patients but without comprehending the primary decision-making process behind why certain assumptions were made [53]

Several techniques have been suggested to enhance interpretation, including saliency maps and activation maximisation, which show which parts of an image matter most in the decision to create a model. For example, GradCAM provides images with colours that indicate the areas affecting the decision more significantly than others. [54] Moreover, some tools are not meant to be associated with any particular model but help understand how certain aspects lead to specific conclusions, like SHAP, which stands for Shapley Additive explanations, or local interpretable model-agnostic explanations code-named LIME. These methods improve transparency and facilitate clinical validation. [55]

Clinical Validation and Implementation

Validating deep learning models in clinical settings entails thorough testing to ensure that they give reliable and actionable insight. Large-scale trials and real-world evidence are required for clinical validation to confirm that models perform as expected across different patient populations and imaging conditions. For instance, the study found that deep learning models performed well in research settings. Still, their effectiveness decreased when applied to real-life clinical data due to variations in image quality and patient demographics. [56]

Implementation of deep learning models into clinical practice faces several obstacles. These include integration with existing clinical workflows, regulatory approval, and the need for seamless interoperability with electronic health records (EHRs). Moreover, there are concerns about data privacy and security because patient data should be protected during model training and deployment [57]. Overcoming these obstacles requires collaboration among researchers, clinicians, and policymakers to create standards and guidelines for the safe and effective use of deep learning in health care. To bridge the gap between research and clinical practice, efforts have been made to develop user-friendly interfaces for clinicians, manage conformity with the standard regulations and perform thorough validation studies. Collaborative initiatives aim at standardising practices and advancing the use of AI tools in radiology [58]. Therefore, these initiatives are essential in transforming profound learning advancements into actionable solutions to improve patient care.

6. Comparison between literature reviews about different models for Automated Segmentation and Quantification of Tumor Burden in Hepatic Cancer Imaging.

Paper	Abstract summary	Main findings	Algorithms	Outcome measured
[64]	Deep learning models can accurately	- The proposed deep learning approach can provide automated	1) Residual-attention U-Net (RA-Unet) for liver segmentation 2)	1) Segmentation of the liver, liver tumours, and ablation zones in

	segment liver tumours and ablation zones in multi-phase CT images, enabling quantitative evaluation of treatment success.	segmentation of liver tumours and ablation zones on multi-phase (arterial and portal venous) and multi-time-point (before and after RFA/MWA ablation) routine clinical CT images, enabling quantitative evaluation of treatment success. - Using transfer learning, an initial model can be generalised to another imaging phase and another type of lesion with a relatively small amount of additional training data. - The model demonstrated higher detection and segmentation performance for tumours with a volume ≥ 0.5 cm3, which corresponds to spherical tumours with a diameter > 1cm.	Multi-scale patch-based 3D RA-Unet for tumour and ablation zone segmentation 3) Transfer learning to adapt the base models to the clinical dataset	both arterial and portal venous phase CT images 2) Quantitative evaluation of the segmentation performance using metrics like Dice Similarity Coefficient (DSC), sensitivity, precision, and F1 score
[65]	The paper presents a deep learning-based method for the automatic segmentation and classification of liver tumours in CT scans.	- The proposed deep learning-based system achieved very high performance for automatic segmentation and classification of liver tumours in CT scans, with a Dice score of 95.40%, Jaccard index of 92%, and accuracy of 92.60% for segmentation, as well as accuracy of 96%,	1. A modified Dense U-Net model for liver tumour segmentation 2. A novel deep convolutional neural network (CNN) architecture based on the pre-trained VGG-16 network for liver tumour classification	The primary outcomes measured in this study are the performance of the deep learning models for liver tumour segmentation and classification, as measured by the Dice Score, Jaccard Index, accuracy, sensitivity, specificity, and precision.

		sensitivity of 95.80%, specificity of 96.20%, and precision of 95.80% for classification. - The classification model used a novel deep convolutional neural network architecture based on the pre-trained VGG-16 network to distinguish between normal and malignant liver tumours. - The segmentation model used a modified version of the Dense U-Net architecture.		
[66]	This paper reviews deep learning techniques for automated liver and liver tumour segmentation from 3D medical images.	- The main findings of the review are a summary of the deep learning techniques and their evaluation metrics used for liver and liver tumour segmentation. - The review provides an overview of the 3D volumetric imaging architectures used for semantic segmentation of liver and liver tumours. - The review compares the performance of different deep learning approaches for liver and tumour segmentation, as measured by the dice score.	1. Automatic and semi-automatic techniques for liver and liver tumor segmentation 2. Deep learning techniques for medical image segmentation in 3D volumetric images 3. Various 3D volumetric imaging architectures designed for semantic segmentation	Not mentioned
[67]	Deep learning approaches like	- A deep learning-based approach using	The specific algorithms	The primary outcomes measured in this study

	ResUNet and 2D-UNet can accurately segment liver tumors in CT scans.	the ResUNet and 2D-UNet architectures was proposed for liver CT segmentation and classification. - The ResUNet architecture achieved a tumor True Value Accuracy of up to 99% in the training phase. - The 2D-UNet architecture achieved a tumor True Value Accuracy of up to 92% in the training phase.	introduced and used in this study are the ResUNet and 2D-UNet deep learning architectures for liver tumor segmentation.	were the tumour True Value Accuracy of the ResUNet and 2d-unit deep learning architectures for liver and liver tumour segmentation on a standard dataset of liver CT scans.
[68]	A deep learning-based system using watershed transform and Gaussian mixture model techniques can accurately detect and classify different types of liver cancer in CT images.	- The proposed WGDL technique achieved a classification accuracy of 99.38% and a Jaccard index of 98.18% for detecting three types of liver cancer using a deep neural network classifier. - The developed WGDL system is ready to be tested on a more extensive database and can aid radiologists in detecting liver cancer from CT images.	1) Marker controlled watershed segmentation 2) Gaussian mixture model (GMM) 3) Deep neural network (DNN) classifier	The primary outcome measured in this study was the classification accuracy of the deep learning model in detecting three types of liver cancer: hemangioma (HEM), hepatocellular carcinoma (HCC), and metastatic carcinoma (MET).
[69]	A deep learning algorithm based on modified ResUNet architecture can automatically segment liver and tumors from CT scans with high accuracy.	- The proposed deep learning algorithm based on a modified ResUNet architecture can automatically segment the liver and tumours from abdominal CT scan images with high accuracy (96.35% DSC for the liver, 89.28% DSC for tumours, and	The primary algorithm used in this study was a modified ResUNet architecture, which is a type of convolutional neural network (CNN) for semantic segmentation.	The primary outcomes measured in this study were the accuracy and Dice Similarity Coefficients (DSCs) for liver and tumor segmentation from CT scans using a deep learning algorithm.

		over 99.7% accuracy). - The deep learning algorithm outperforms other linked methods for liver and tumor segmentation. - Automatic segmentation of the liver and tumors can help minimize the time and effort required for liver disease diagnosis.		
[70]	The paper presents an automated liver tumor segmentation and classification model using deep learning approaches.	- The proposed method achieves high performance in liver tumor segmentation and classification, with the highest precision for lesion identification while maintaining a high recall value. - The method can accurately classify liver tumors into three categories (HCC, malignant, and benign/cyst) with an average accuracy of 87.8%. - The key novelty of the study is the use of a mask-RCNN-based method for liver segmentation, MSER for tumor lesion segmentation, and a hybrid CNN-based approach for tumor classification.	1. Mask-RCNN for liver segmentation 2. MSER for tumour identification 3. Hybrid CNN model for tumour classification	1. Liver tumor segmentation, to distinguish between normal and malignant tissue in the liver. 2. Liver tumor classification, to categorize identified liver tumors into three classes: hepatocellular carcinomas (HCC), malignant (other than HCCs), and benign or cyst, with an average accuracy of 87.8%.

7. Future Directions in Deep Learning for Tumor Burden Prediction

The rapidly evolving field of deep learning in tumor burden prediction has many promising future directions that will enhance these technologies' accuracy, reliability, and clinical applicability. This section elaborates on potential areas for further research and development, such as multimodal imaging, deep learning-based image registration, AI-assisted interactive segmentation, explainable AI, clinical trials, etc.

Multi-Modal Imaging for Improved Segmentation and Quantification

Combining CT, MRI, and PET scans can provide a complete picture of the tumour's physical and metabolic characteristics. As a result, it improves the segmentation and quantification of tumours. For instance, CT images provide anatomical detail, while MRI scans offer better contrast for soft tissues than CT images. On the other hand, PET reveals where metabolic activity is (which is essential in cancer). Thus, integrating these modalities leads to greater accuracy when delineating tumours, giving one an insightful perspective into their burden [59].

Using multi-modal data in deep learning models dramatically enhances performance compared to single-modal ones. Recent studies have shown that fused imaging data improves tumor segmentation and characterisation. A group of researcher created a multi-modal deep learning model that fused MRI with PET information, enabling accurate tumour segmentation and determining if it was responding to treatment or not [60]. In doing so, this work improved the dice similarity coefficient from 0.82 for single-modality models to 0.88, thereby demonstrating why there is a need for employing multi-modal strategies in medical imaging applications.

Deep Learning-Based Image Registration for Longitudinal Studies

Longitudinal studies follow tumour changes over time and necessitate accurate image registration for proper alignment of sequential scans. This alignment is essential in evaluating tumour growth, treatment efficacy and progression. Nevertheless, traditional image registration methods are often time-consuming to compute and may be unable to handle variations in imaging conditions.

Image registration for longitudinal studies has been improved by deep learning. Image registration can benefit from convolutional neural networks (CNNs) and transformer-based architectures, which could enhance both accuracy and efficiency in administering the process. A study conducted in 2023 proposed a deep learning-based registration model that significantly lowered the number of registration errors while saving computational time, as opposed to conventional systems. This advancement makes it possible to track tumour dynamics more accurately and assess the effects of treatment over time in a better way. [61]

Integration of AI-Assisted Interactive Segmentation

The automated algorithms are integrated with user inputs to refine segmentation results. The purpose of this tool is to allow clinicians to adjust the outputs of segmentation and ensure that they align with their expertise and anatomical knowledge. Recently, AI-assisted interactive segmentation has utilised deep learning methods that provide real-time feedback and corrections. For instance, a study developed an interactive segmentation tool that combines deep learning and a user-friendly interface where clinicians can change the output dynamically. This resulted in improved accuracy and efficiency of tumour delineation since the feedback by users reduced segmentation errors by 20% compared to totally automatic techniques. [62]

Explainable AI for Understanding Model Decisions

Deep learning models must be adopted clinically to benefit from them, which requires understanding their decision-making capabilities as they become more complicated. XAI provides doctors with proof that can be trusted and verified through insights from the model's predictions.

Most recent work on XAI has focused on soothing the complexities surrounding deep learning models. Some techniques such as saliency maps, Grad-CAM and SHAP (Shapley Additive explanations) can be used to show how a given model planned or arrived at a particular conclusion. An illustration is the case where SHAP was used to interpret the predictions made by one of the deep learning models meant for liver tumour classification, which enabled an understanding of how specific features contribute, thus increasing the transparency level of the model. [63] These methods are essential for integrating deep learning in clinical workflows to ensure that the model's decisions can be understood and acted upon.

Clinical Trials to Evaluate the Impact of Deep Learning-Based Tumor Burden Assessment on Patient Outcomes

Deep learning tools to assess tumour load must be translated into concrete gains in terms of actual clinic observations through rigorous clinical trials whose edge lies in their ability to appraise how these advanced tools work in real life and what outcomes they bring about on patients such as continued existence, life quality and treatment success what. Designing clinical trials for deep learning tools requires various considerations. These include choosing suitable endpoints, abiding by regulations, and harmonising with current clinical workflows. Recently aimed at incorporating AI instruments into clinical investigations that assess its influence on patient management and outcomes have been the efforts of organisations like the National Cancer Institute (NCI). Such trials will offer crucial evidence concerning the efficiency and safety of assessments based on deep learning that will guide their endorsement in everyday hospitals.

Conclusion

This literature review critically assessed the use of deep learning architectures for the segmentation and quantification of tumour burden in hepatic cancer imaging. The analysis exposed an excellent performance by the CNN-based models, especially U-Net and DeepLabv3+, which have significantly improved upon traditional manual methods in accuracy and efficiency. However, it also identified several challenges before these models fit into standard clinical practice. The difference in the quality and generalizability of datasets is one of the main problems of deep learning models. Many of these studies used single-centre datasets with limited diversity, leading to doubts about whether they would function well with different populations and imaging settings. It is also challenging because inter-observer variability on ground truth annotation creates a significant obstacle to model evaluation. To improve the field, future studies should concentrate on making standard datasets and evaluation criteria that can be widely used throughout research. This will allow for more dependable model performance comparisons and the incorporating of deep learning-based segmentation tools into everyday hospital practices. Moreover, conducting extensive multi-centre studies that verify these techniques in real life is necessary, hence providing strength and viability. In summary, despite potential improvements through deep learning in hepatic cancer burden segmentation/quantification, the clinical application may necessitate overcoming existing limitations through collaboration amongst research centres and medical facilities.

References

1. International Agency for Research on Cancer. (2020). *Global cancer observatory: Cancer today*. https://gco.iarc.fr/today
2. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 68*(6), 394-424. https://doi.org/10.3322/caac.21492
3. Llovet, J. M., Kelley, R. K., Villanueva, A., Singal, A. G., Pikarsky, E., Roayaie, S., ... & Finn, R. S. (2021). Hepatocellular carcinoma. *Nature Reviews Disease Primers, 7*(1), 6. https://doi.org/10.1038/s41572-020-00240-3

4. Renzulli, M., et al. (2018). New hallmark of hepatocellular carcinoma development in non-cirrhotic liver: Correlation between increased tumor volume and poor prognosis. *World Journal of Gastroenterology*, *24*(3), 422-430. https://doi.org/10.3748/wjg.v24.i3.422
5. Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G. P. M., Granton, P., ... & Gillies, R. J. (2012). Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, *48*(4), 441-446. https://doi.org/10.1016/j.ejca.2011.11.036
6. Gillies, R. J., Kinahan, P. E., & Hricak, H. (2016). Radiomics: Images are more than pictures, they are data. *Radiology*, *278*(2), 563-577. https://doi.org/10.1148/radiol.2015151169
7. Wang, L., Gong, C., Chen, H., & Hu, W. (2019). Deep learning for tumor segmentation. *Annual Review of Biomedical Engineering*, *21*, 221-245. https://doi.org/10.1146/annurev-bioeng-060418-052904
8. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A., Ciompi, F., Ghafoorian, M., ... & van Ginneken, B. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60-88. https://doi.org/10.1016/j.media.2017.07.005
9. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539
10. Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10) (pp. 807-814).
11. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28
12. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 801-818). https://doi.org/10.1007/978-3-030-01234-2_49
13. Zhu, C., Xu, F., & Chen, Q. (2021). Performance comparison of deep learning models for automatic liver segmentation in computed tomography. Journal of Healthcare Engineering, 2021, 1-10. https://doi.org/10.1155/2021/6656901
14. Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.
15. Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. Medical Image Analysis, 58, 101552. https://doi.org/10.1016/j.media.2019.101552
16. Si, X., Yuan, Y., Fang, Z., & Xu, H. (2020). Long short-term memory neural networks for protein secondary structure prediction. BMC Bioinformatics, 21(1), 1-10. https://doi.org/10.1186/s12859-019-3317-6
17. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
18. Ktena, S. I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., & Rueckert, D. (2017). Distance metric learning using graph convolutional networks: Application to functional brain networks. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 469-477). Springer, Cham. https://doi.org/10.1007/978-3-319-66179-7_54

19. Zhao, H., Cui, Y., Zhang, S., Wang, W., & Ji, R. (2021). Joint self-attention and scale-aware convolution for universal lesion detection in CT images. IEEE Transactions on Medical Imaging, 40(2), 365-376. https://doi.org/10.1109/TMI.2020.3027078
20. Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, *26*(3), 297-302. https://doi.org/10.2307/1932409
21. Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist*, *11*(2), 37-50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x
22. Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, *35*(5), 1240-1251. https://doi.org/10.1109/TMI.2016.2538465
23. Heimann, T., Meinzer, H. P., & van Ginneken, B. (2009). Quantitative evaluation of medical image segmentation: Conceptual framework and algorithm analysis. *IEEE Transactions on Medical Imaging*, *28*(11), 1673-1689. https://doi.org/10.1109/TMI.2009.2023117
24. Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging*, *15*(1), 1-28. https://doi.org/10.1186/s12880-015-0068-x
25. Zhao, L., Barbera, L., Sutradhar, R., Lui, N., Elit, L., & Dibaj, S. (2013). Inter-observer agreement in cancer staging: Impact on performance status, a key determinant of prognosis. *Radiotherapy and Oncology*, *107*(3), 326-332. https://doi.org/10.1016/j.radonc.2013.05.027
26. Warfield, S. K., Zou, K. H., & Wells, W. M. (2004). Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, *23*(7), 903-921. https://doi.org/10.1109/TMI.2004.828354
27. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 9351*, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
28. Liu, J., Zhang, J., Li, X., et al. (2018). Liver tumor segmentation in CT images using U-Net and deep learning methods. *Computers in Biology and Medicine, 101*, 64-72. https://doi.org/10.1016/j.compbiomed.2018.08.001
29. Hsieh, M.-K., Zhang, Y., Lu, S., et al. (2020). Deep learning-based hepatic tumor segmentation: A comparative study of U-Net and other architectures. *Medical Image Analysis, 64*, 101743. https://doi.org/10.1016/j.media.2020.101743
30. Zhao, J., Liu, Y., Liu, W., et al. (2021). The impact of segmentation accuracy on volumetric tumor measurements and clinical outcomes. *Journal of Medical Imaging, 8*(1), 014502. https://doi.org/10.1117/1.JMI.8.1.014502
31. Zhang, Z., Han, X., Xu, H., et al. (2019). Attention mechanisms in U-Net for improving segmentation of hepatic tumors. *International Journal of Computer Assisted Radiology and Surgery, 14*(7), 1121-1130. https://doi.org/10.1007/s11548-019-01930-5
32. Aerts, H. J. W. L., Velazquez, E. R., Leijenaar, R. T. H., et al. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications, 5*(1), 4006. https://doi.org/10.1038/ncomms5006
33. Lambin, P., Rios-Velazquez, E., Leijenaar, R. T. H., et al. (2017). Radiomics: extracting more information from medical images using advanced feature analysis. *European Journal of Cancer, 48*(4), 441-446. https://doi.org/10.1016/j.ejca.2011.11.036

34. Rizzo, S., D'Onofrio, M., De Feo, P., et al. (2020). Role of texture analysis in predicting outcomes of liver cancer patients. *Journal of Hepatology, 73*(3), 524-532. https://doi.org/10.1016/j.jhep.2020.03.012
35. Liao, J., Wang, Y., Zhang, X., et al. (2021). Radiomics in the prediction of treatment response and survival outcomes in liver cancer: A systematic review. *Frontiers in Oncology, 11*, 684439. https://doi.org/10.3389/fonc.2021.684439
36. Chen, T., Wang, X., Zhang, Y., et al. (2022). Development of a hybrid index combining volumetric and radiomic features for the prognosis of hepatic cancer. *Journal of Clinical Oncology, 40*(15_suppl), e18163. https://doi.org/10.1200/JCO.2022.40.15_suppl.e18163
37. Yang, Z., Zhang, X., Liu, L., et al. (2023). Composite indices of tumor volume and texture features in predicting survival in hepatic cancer patients. *Radiology, 307*(2), 420-429. https://doi.org/10.1148/radiol.23001424
38. Wang, X., Xu, H., Zhang, H., et al. (2024). Predicting survival outcomes based on tumor burden dynamics using deep learning. *Cancer Imaging, 24*(1), 14. https://doi.org/10.1186/s12885-024-10578-7
39. Li, X., Yu, H., Zhang, X., et al. (2021). Fusion of clinical and radiomic features for outcome prediction in hepatic cancer using deep learning. *Radiology, 300*(2), 453-463. https://doi.org/10.1148/radiol.2021202617
40. Gessert, N., Chen, Y., Zhang, Y., et al. (2022). Multi-modal deep learning for predicting tumor progression: Combining imaging and clinical data. *Medical Image Analysis, 76*, 102386. https://doi.org/10.1016/j.media.2022.102386
41. Zhang, J., Wang, Q., Liu, Y., et al. (2022). Predicting treatment response in hepatic cancer using 3D deep learning models. *Computers in Biology and Medicine, 146*, 105776. https://doi.org/10.1016/j.compbiomed.2022.105776
42. Liu, J., Li, X., Chen, X., et al. (2023). Temporal deep learning models for predicting treatment response in hepatic cancer. *IEEE Transactions on Medical Imaging, 42*(7), 1980-1991. https://doi.org/10.1109/TMI.2023.3225879
43. Wang, X., Zhang, H., Yang, H., et al. (2024). Integrating volumetric and radiomic features for predicting response to hepatic cancer treatment: A new paradigm. *Cancer Imaging, 24*(1), 3. https://doi.org/10.1186/s12940-024-01028-5
44. Zhao, J., Li, H., Zhang, W., et al. (2023). Longitudinal deep learning models for survival prediction in liver cancer. *Frontiers in Oncology, 13*, 104838. https://doi.org/10.3389/fonc.2023.104838
45. Yang, Z., Sun, J., Zheng, Y., et al. (2022). Development and validation of a prognostic model incorporating deep learning features for hepatic cancer. *Journal of Clinical Oncology, 40*(15_suppl), e18220. https://doi.org/10.1200/JCO.2022.40.15_suppl.e18220
46. Chen, T., Zhang, X., Wang, Y., et al. (2023). Integrative prognostic modeling using deep learning: Combining radiomic, genomic, and clinical features for liver cancer survival prediction. *Journal of Biomedical Informatics, 134*, 104091. https://doi.org/10.1016/j.jbi.2023.104091
47. Coudray, N., Ocampo, P. S., Sakellaropoulos, T., et al. (2018). Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature Medicine, 24*(10), 1559-1567. https://doi.org/10.1038/s41591-018-0177-5

48. Liu, J., Xu, H., Zhang, Y., et al. (2020). Imaging data variability and its impact on deep learning model performance: A review. *IEEE Reviews in Biomedical Engineering, 13*, 101-116. https://doi.org/10.1109/RBME.2019.2910368
49. Shin, H.-C., Roth, H. R., Gao, M., et al. (2016). Deep convolutional neural networks for radiology. *ACR Annual Meeting*.
50. Ho, C., Wang, X., & Chiu, T. (2020). Generalizing deep learning models for medical imaging: Challenges and solutions. *Medical Image Analysis, 64*, 101748. https://doi.org/10.1016/j.media.2020.101748
51. Wang, X., Wu, W., & Liang, Z. (2021). Generalizability of deep learning models across diverse patient populations and imaging conditions. *IEEE Transactions on Medical Imaging, 40*(5), 1151-1162. https://doi.org/10.1109/TMI.2021.3053128
52. Zhu, X., Li, W., & Yang, J. (2021). Domain adaptation for cross-modality medical image segmentation using deep learning. *Medical Image Analysis, 68*, 101893. https://doi.org/10.1016/j.media.2020.101893
53. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *Proceedings of the 2017 ICML Workshop on Human Interpretability in Machine Learning*. https://arxiv.org/abs/1702.08608
54. Selvaraju, R. R., Cogswell, M., Das, A., et al. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618-626. https://doi.org/10.1109/ICCV.2017.74
55. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144. https://doi.org/10.1145/2939672.2939778
56. Hwang, J., Chang, P., & Park, H. (2019). Real-world performance of deep learning models for medical imaging: A review. *Journal of Medical Imaging, 6*(4), 041408. https://doi.org/10.1117/1.JMI.6.4.041408
57. Mongan, J., Moy, L., & Kahn, C. E. (2020). Checklist for artificial intelligence in medical imaging: A guide for authors and reviewers. *Radiology, 295*(1), 1-8. https://doi.org/10.1148/radiol.202020181
58. Langlotz, C. P., Allen, B., Erickson, B. J., et al. (2019). A roadmap for AI in radiology. *Journal of the American College of Radiology, 16*(2), 232-238. https://doi.org/10.1016/j.jacr.2018.10.018
59. Jiang, X., Zhang, L., Huang, J., et al. (2021). Multi-modal imaging for enhanced tumor segmentation and quantification: A review. *Medical Image Analysis, 67*, 101856. https://doi.org/10.1016/j.media.2020.101856
60. Xu, Z., Zhang, X., Wang, S., et al. (2022). Fusion of MRI and PET for improved tumor segmentation using deep learning. *International Journal of Computer Assisted Radiology and Surgery, 17*(5), 759-770. https://doi.org/10.1007/s11548-022-02682-7
61. Zhao, W., Liu, H., Zhang, Q., et al. (2023). Deep learning-based image registration for longitudinal studies: Enhancing accuracy and efficiency. *IEEE Transactions on Medical Imaging, 42*(3), 789-800. https://doi.org/10.1109/TMI.2023.3216569
62. Kwon, J., Kim, J., Lee, S., et al. (2024). Interactive deep learning-based segmentation for hepatic tumors: Enhancing clinical usability and accuracy. *IEEE Transactions on Medical Imaging, 43*(1), 120-131. https://doi.org/10.1109/TMI.2023.3289123

63. Salinas, J., Chen, Y., Garcia, M., et al. (2023). Interpretable deep learning for liver tumor classification using SHAP values. *Medical Image Analysis, 70*, 102045. https://doi.org/10.1016/j.media.2021.102045
64. He, K., Liu, X., Shahzad, R., Reimer, R., Thiele, F., Niehoff, J., Wybranski, C., Bunck, A. C., Zhang, H., & Perkuhn, M. (2021). Advanced Deep Learning Approach to Automatically Segment Malignant Tumors and Ablation Zone in the Liver With Contrast-Enhanced CT. Frontiers in oncology, 11, 669437. https://doi.org/10.3389/fonc.2021.669437
65. R, Aparna & M, Libish. (2023). Automatic segmentation and classification of the liver tumor using deep learning algorithms. 334-339. 10.1109/ACCESS57397.2023.10200900
66. Gul, S., Khan, M. S., Bibi, A., Khandakar, A., Ayari, M. A., & Chowdhury, M. E. H. (2022). Deep learning techniques for liver and liver tumor segmentation: A review. Computers in biology and medicine, 147, 105620. https://doi.org/10.1016/j.compbiomed.2022.105620
67. Bhandary, P., More, V., & Nagpure, P.R. (2021). Liver Tumor Segmentation using Deep Learning Approach.
68. Sabut, Sukanta & Das, Amita & Acharya, U & Panda, Soumya. (2018). Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques. Cognitive Systems Research. 54. 10.1016/j.cogsys.2018.12.009.
69. Manjunath, R.V. & Kwadiki, Karibasappa. (2021). Automatic liver and tumour segmentation from CT images using Deep learning algorithm. Results in Control and Optimization. 6. 100087. 10.1016/j.rico.2021.100087
70. Saha Roy, Sayan & Roy, Shraban & Mukherjee, Prithwijit & Roy, Anisha. (2022). An automated liver tumour segmentation and classification model by deep learning based approaches. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. 11. 1-13. 10.1080/21681163.2022.2099300