JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 09 Issue 02
2025

Research Article
https://doi.org/10.56979/902/2025

A Convolutional Neural Network and Vision Transformer Based Framework for Effective Detection of Liver Cancer

Asma Zahoor[1], Erssa Arif[1], Naila Nawaz[1], Muhammad Amjad[1], Shahrukh Hamayoun[2], and Arslan Baig[1*]

[1]Riphah College of Computing, Riphah International University, Faisalabad, Pakistan.
[2]Department of Computer Science, National University of Modern Languages, Faisalabad, Pakistan
[*]Corresponding Author: Arslan Baig. Email: arslan313mughal@gmail.com

Received: July 06, 2025 **Accepted**: August 19, 2025

Abstract: Liver cancer, particularly hepatocellular carcinoma (HCC), remains one of the most prevalent and lethal malignancies worldwide, underscoring the urgent need for early and reliable diagnostic solutions. Conventional diagnostic methods using computed tomography (CT) imaging are often limited by inter-observer variability and the high cognitive burden on radiologists. To address these challenges, this study proposes a hybrid deep learning framework that leverages Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) for effective liver cancer detection. The research employs the publicly available 3D-IRCADb1 dataset of contrast-enhanced CT scans, with preprocessing and augmentation techniques applied to enhance model generalization. Three state-of-the-art architectures, EfficientNet-B0, TinyViT, and MobileViT v2, were trained and evaluated to assess their diagnostic performance. Among these, MobileViT v2 demonstrated superior performance and efficiency in classification tasks. To enhance clinical trust, Gradient-weighted Class Activation Mapping (Grad-CAM) was integrated to provide visual explanations of model predictions, highlighting regions of interest corresponding to tumor areas. The findings indicate that the proposed framework not only ensures robust diagnostic capability but also introduces interpretability and efficiency, making it suitable for deployment in clinical and resource-constrained environments. This research contributes to advancing AI-driven liver cancer diagnostics by bridging the gap between performance and transparency, ultimately supporting earlier detection and improved patient outcomes.

Keywords: Liver Cancer Detection; Hepatocellular Carcinoma (HCC); Computed Tomography (CT) Imaging; Vision Transformers (ViTs); EfficientNet-B0; TinyViT; MobileViTv2; Medical Image Analysis; Clinical Decision Support Systems; Early Cancer Diagnosis

1. Introduction

The liver is a key organ in the human body, involved in numerous vital functions that support overall physiological balance and health. These include filtering toxins and waste products from the blood, detoxifying hormones, metabolizing drugs, synthesizing essential proteins, supporting immune function, and producing bile to aid digestion [4],[7]. Given its central role in metabolic homeostasis and systemic health, diseases that impair liver function have profound consequences. Among these, liver cancer, particularly hepatocellular carcinoma (HCC), is the most prevalent and aggressive form of primary liver malignancy.

Hepatocellular carcinoma (HCC) commonly arises in individuals suffering from long-term liver conditions, including hepatitis B or C infections, liver cirrhosis, and non-alcoholic fatty liver disease. It is characterized by rapid progression and poor prognosis when diagnosed at advanced stages. Liver cancer

represents a significant global public health challenge. According to GLOBOCAN 2020 estimates, liver cancer is the second leading cause of cancer-related deaths in men and the sixth among women. Furthermore, it ranks sixth in incidence and third in cancer-related mortality across both sexes [23]. Recent projections suggest that the global incidence of liver cancer is expected to exceed one million cases by 2025 underscoring the urgent need for timely and accurate diagnosis [2].

Advancements in medical imaging technologies over the past decade have significantly improved the early detection and monitoring of liver cancer. Modalities such as computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, positron emission tomography (PET), and X-ray imaging have become integral to modern diagnostic workflow [13]. Among these, contrast-enhanced CT scans are particularly favored due to their speed, accessibility, high spatial resolution, and ability to capture dynamic contrast phases, especially in the portal venous phase (around 60 to 80 seconds after contrast injection) a phase that offers the best visibility for detecting tumors [9].

Despite these technological advancements, diagnostic workflows in clinical practice often rely on manual interpretation of imaging data by radiologists, which is inherently subjective, time-consuming, and susceptible to inter-observer variability. Studies have shown that evaluating large volumes of CT images across multiple phases requires significant expertise and cognitive effort, often resulting in diagnostic delays, missed lesions, and inconsistent results [3].

1.1. Deep Learning in Medical Imaging

The integration of deep learning (DL) techniques into medical image analysis has revolutionized the field of diagnostic radiology and pathology. Traditional manual diagnosis methods are often time-consuming, prone to inter-observer variability, and heavily reliant on clinical expertise. These limitations, coupled with the growing volume of medical imaging data, have necessitated the development of automated and intelligent systems capable of assisting clinicians in identifying, classifying, and monitoring diseases with greater efficiency and accuracy.

Deep learning, which is part of artificial intelligence, has been very effective in pattern recognition of large and complicated data sets. Inspired by human brain structure and function, DL models are composed of several layers of ANNs that are capable of directly using the raw input data without the need for specialized feature engineering. It is quite a nice feature if we talk about the healthcare field, where there is a lot of high-dimensional imaging data like CT, MRI, PET, and histopathology slides, which need to be processed and understood by the machines.

One of the defining characteristics of deep learning models in medical imaging is their ability to learn hierarchical feature representations. Unlike traditional machine learning algorithms that rely on handcrafted features, DL models autonomously discover optimal representations for a given task. This enables them to detect subtle anomalies, recognize complex anatomical structures, and adapt to various imaging modalities. For instance, in liver cancer diagnosis, deep learning models can learn to distinguish between different tissue textures and tumor boundaries, even in noisy or low-contrast images.

1.1.1. Convolutional Neural Network

Convolutional Neural Networks (CNNs) represent one of the most prominent and widely used architectures in the field of deep learning, particularly suited for analyzing grid-like data structures such as two-dimensional images or three-dimensional volumetric scans. Unlike traditional Artificial Neural Networks (ANNs), which process input data in a fully connected manner without considering spatial correlations, CNNs are explicitly designed to exploit the spatial hierarchies in data.

To manage the computational load and to control overfitting, pooling layers are typically interspersed between convolutional layers. These layers perform downsampling operations, such as max pooling or average pooling, which reduce the spatial dimensions of the feature maps while preserving the most important information. Pooling also helps the network become invariant to minor translations and distortions in the input image, improving its generalization capability. Training a CNN involves learning the optimal weights of convolutional filters and fully connected layers by minimizing a loss function (e.g., cross-entropy for classification tasks) using optimization algorithms such as stochastic gradient descent (SGD) or Adam. Modern CNNs often employ batch normalization to stabilize training and dropout layers to prevent overfitting.

With the advancement of computational resources and the availability of annotated medical datasets, CNNs have emerged as the backbone of many diagnostic and prognostic systems. Recent research has

expanded their use from standard 2D imaging to 3D volumetric data, enabling more precise modeling of anatomical structures and pathologies in computed tomography (CT) and magnetic resonance imaging (MRI) scans. In summary, CNNs are a cornerstone of deep learning for visual data, offering a robust, scalable, and efficient solution for complex image-based tasks. Their architectural innovations and proven efficacy continue to drive progress in computer vision and healthcare applications alike, particularly in early disease detection and image-guided diagnosis.

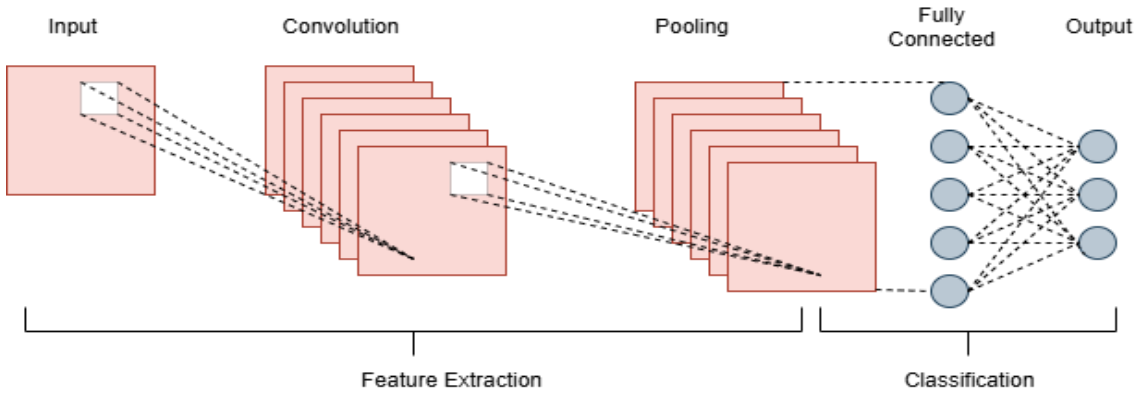

Figure 1. Architecture of CNN

2. Related Work

This section provides a structured review of existing literature relevant to the automated detection of liver cancer using deep learning techniques. It aims to examine the progression of computational methods that have been employed for analyzing liver-related medical imaging, particularly focusing on hepatocellular carcinoma (HCC), the most common and aggressive form of primary liver cancer. This is followed by a discussion on the application of Vision Transformers in the medical domain, emphasizing their capability to model global relationships across image regions using self-attention mechanisms. Recent attempts to create hybrid CNN-Transformer models are covered in the paper as well, indicating the trend towards designs that combine the local feature extraction with the long-range dependency. This chapter, along with the discussion of model architectures, considers the use of Explainable Artificial Intelligence (XAI) tools, for instance, Grad-CAM and SHAP, which are the foremost adopters of the transparency and interpretability of AI-driven diagnostic systems.

The application of deep learning and hybrid artificial intelligence models has significantly transformed the landscape of liver cancer diagnosis and classification. Recent years have witnessed an influx of research dedicated to improving lesion detection, segmentation, and classification accuracy using diverse imaging modalities and learning strategies. The studies discussed below present notable approaches, architectures, and innovations aimed at enhancing clinical decision-making in liver cancer diagnostics.

Wan et al., proposed a multi-level, multi-scale fusion convolutional neural network (MMF-CNN) designed to classify liver lesions from magnetic resonance imaging (MRI). Their architecture utilized multiple CNN branches that processed inputs at different resolutions and depths, allowing for a rich multi-scale representation of tumor features. At the decision level, the model incorporated Dempster Shafer theory for evidence fusion, effectively combining outputs from the various network branches. Additionally, the study employed Grad-CAM to provide visual heatmaps that explain which regions of the MRI were most influential in the model's decision, thereby enhancing the interpretability of predictions in clinical settings [25].

Lee et al., introduced HFS-Net, a three-stage hierarchical fusion strategy for segmenting hepatocellular carcinoma (HCC) in contrast-enhanced CT images. The framework combined the strengths of multiple specialized models: U-Net for initial liver segmentation, DenseU-Net for tumor feature enhancement, and a 3D U-Net for final tumor segmentation refinement across volumetric slices. Each model operated on a distinct image phase or tumor size range, and outputs were later merged using a tailored fusion protocol. This modular design enabled the system to adapt to diverse lesion presentations and improved performance across various phases of dynamic CT imaging [15].

Sridhar, Kavitha, Lai, & Kavin, presented a hybrid framework combining geodesic distance transformation and CNN classification for liver tumor segmentation from CT images. The process began with a semi-automated stage where radiologists marked the inner boundary of lesions, which was then refined using geodesic mapping to derive accurate tumor contours. These segmented lesions were fed into a convolutional neural network optimized using a Coot algorithm, which is a nature-inspired metaheuristic technique. The CNN classified the lesions based on extracted spatial features, producing strong segmentation and classification outcomes through a combination of manual expertise and automated processing [22].

Kriegsmann et al., employed CNN architectures such as EfficientNetV2B0 and ResNetRS50 to classify liver tissue from digitized whole-slide histopathological images. The study compiled an extensive dataset of over 200,000 tile images, extracted from more than 700 patient samples. The classification task included distinguishing between benign tissue, hepatocellular carcinoma, metastatic tumors, and other lesion types. A multiple-instance learning strategy was adopted to ensure that tile-level predictions could be aggregated into accurate case-level diagnoses. The research emphasized the value of high-throughput digital pathology combined with deep learning for objective and scalable liver cancer assessment [12].

Sarfati et al., designed a classification system that merges deep learning and radiomics to evaluate liver tumors in multiphase contrast-enhanced CT scans. Their two-stage approach involved using a CNN to extract image features related to LI-RADS scores, followed by the extraction of radiomic descriptors. These were combined and fed into classical machine learning models for final classification. The hybrid nature of the system allowed for a more nuanced interpretation of liver lesions, closely resembling the diagnostic reasoning process of experienced radiologists, while also maintaining algorithmic efficiency [20].

Kang, Ting, Ting, & Phan, introduced CAFCT-Net, a hybrid CNN-Transformer segmentation network for liver tumor detection in CT scans. The model incorporated Atrous Spatial Pyramid Pooling (ASPP), attention gates, and Attentional Feature Fusion (AFF) modules to extract multi-scale contextual features and refine tumor boundaries. The architecture leveraged the global receptive field of Transformers and the spatial sensitivity of convolutional layers to produce high-resolution segmentations. The model achieved strong Dice coefficients and intersection-over-union (IoU) scores across public datasets, reflecting its robustness in segmenting challenging liver tumors [11].

Bousabarah et al., developed an automatic segmentation model based on a U-Net architecture to detect liver and hepatocellular carcinoma regions from multiphase MRI. The dataset included sequences from arterial, portal venous, and delayed imaging phases. The model achieved high Dice similarity scores for liver segmentation and reasonable performance for HCC segmentation. Post-processing steps, such as false positive suppression and threshold optimization, further improved segmentation quality. The study demonstrated the feasibility of deep learning for fully automated liver tumor segmentation in MRI [5].

Chen, Z., Dou, Luo, & Yao, presented a multitask segmentation model based on the Swin Transformer for liver tumor analysis in contrast-enhanced MRI. The architecture incorporated self-supervised learning strategies and was trained to simultaneously predict segmentation masks and signed distance maps. Deep supervision and attention-based modules were used to refine both tasks in parallel. The use of the Swin Transformer allowed the model to capture hierarchical image structures, and its multitask formulation led to improved segmentation accuracy across multiple tumor classes and sizes [6].

Gao et al., presented STIC, a deep learning model for liver tumor classification using dynamic CT imaging and patient clinical data. The architecture integrated convolutional neural networks for spatial analysis and gated recurrent units (GRUs) for modeling temporal sequences across CT phases. The study evaluated its performance on a large dataset including hepatocellular carcinoma, cholangiocarcinoma, and liver metastases. STIC achieved diagnostic accuracy comparable to experienced radiologists and highlighted the potential of combining imaging and clinical metadata for liver cancer diagnosis [8].

Zhan et al., proposed a Transformer-based architecture for predicting early recurrence of hepatocellular carcinoma using multi-phase MRI. The model treated MRI scans from arterial, portal venous, and delayed phases as sequential inputs and processed them using a Transformer encoder to capture inter-phase relationships. Feature aggregation was applied to summarize spatial and contrast information across phases, enhancing recurrence prediction. The design demonstrated that Transformer-

based modeling can effectively utilize the temporal nature of multiphase imaging for postoperative risk stratification in liver cancer [27].

3. Proposed Methodology

The methodological framework adopted to develop a hybrid Convolutional Neural Network (CNN) and Vision Transformer (ViT) based model for the detection and classification of liver cancer using computed tomography (CT) images. The chapter elaborates on the data acquisition process, preprocessing techniques, model architecture design, training strategies, evaluation metrics, and explainability mechanisms integrated into the proposed system.

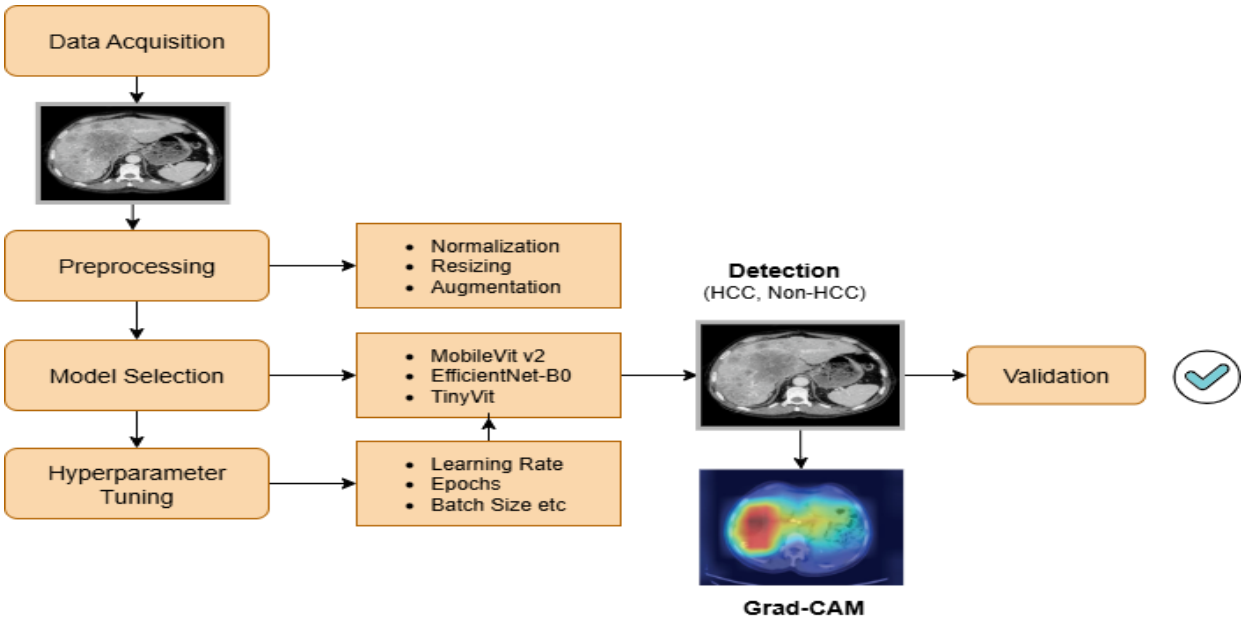

Figure 2. Research Methodology

3.1. MobileViT v2

MobileViT v2 represents a next-generation, lightweight neural network architecture combining the strengths of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) in a mobile-efficient framework. First introduced by [18] It was specifically designed to overcome the performance-efficiency trade-off found in earlier vision models. This makes MobileViT v2 particularly suitable for applications that demand real-time inference and low computational resources, such as medical image classification, where both local feature detection and global contextual understanding are essential.

SaTraditional CNNs excel at capturing local spatial patterns via convolutional kernels but struggle with modeling long-range dependencies across an image. MobileViT v2 addresses this limitation by embedding Transformer-like attention modules within a CNN backbone, enabling the model to integrate both fine-grained and broad contextual information. It refines the design of ViT and Swin Transformer components to be more resource-efficient, allowing deployment on edge devices or under memory constraints. MobileViT v2 combines the strengths of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) in a unified, efficient architecture optimized for performance and speed on mobile and embedded devices. Unlike pure Transformer models, which rely solely on self-attention, or CNNs, which focus on local feature extraction, MobileViT v2 is designed to learn both local and global representations through a structured pipeline of Mobile Inverted Residual Bottleneck (MBConv) blocks and lightweight Transformer blocks.

3.2. Experimental Setup

The proposed study was implemented and executed in a GPU-accelerated environment to ensure efficient model training and inference for liver tumor detection. An NVIDIA T4 GPU was utilized, enabling faster computation for multiple deep learning experiments. The workflow comprised four key stages: DICOM image preprocessing, dataset preparation, training of multiple deep learning models, and interpretability analysis using Gradient-weighted Class Activation Mapping (Grad-CAM).

3.2.1. DICOM Preprocessing and Conversion

The raw dataset was obtained from the publicly available 3D-IRCADb1 collection, which provides abdominal CT scans in DICOM format. To prepare the images for deep learning pipelines, the following preprocessing steps were undertaken:

- Libraries Used: pydicom was employed for reading and extracting metadata from medical images, while OpenCV (cv2) was used for image normalization and conversion.

Procedure: Patient-specific image files were recursively extracted from the 3Dircadb1.zip archive. All PATIENT_DICOM/image_* files were located for processing. Each DICOM file underwent pixel intensity normalization, after which it was saved as a JPEG image in the designated directory /content/data/HCC/. This conversion ensured compatibility with common deep learning frameworks while retaining essential diagnostic information from the CT images.

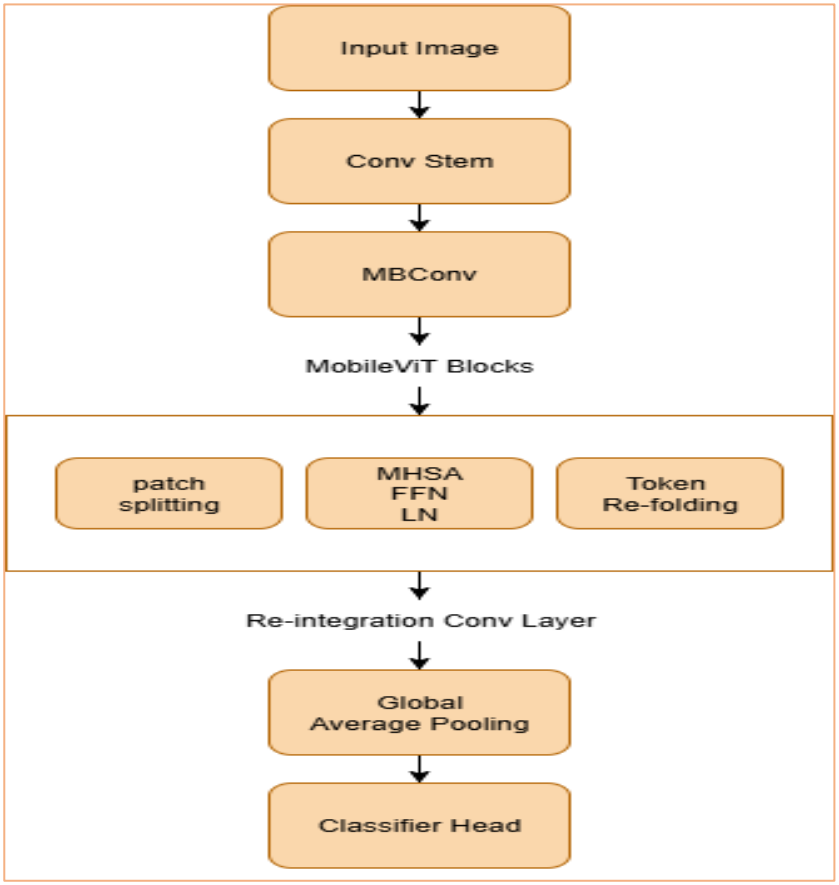

Figure 3. Architecture of MobileViTv2 Model

3.3. Results of Efficient Net B0

The implementation and training of the proposed framework were carried out using Google Colab, a cloud-based platform that provides GPU-accelerated computing resources for efficient deep learning experimentation. Among the models evaluated, EfficientNet-B0 demonstrated superior performance, achieving a test accuracy of 99.2%, highlighting its effectiveness in the liver cancer detection task. The following figures show the training and validation accuracy of this model. An efficient Net B0 model applied on the 3Dircadb1 dataset that contains liver cancer images and obtained 99.2% accuracy.

Table 1. Efficient Net-B0 results in Google Colab

Model	Test Accuracy	Precision	Recall	F1 Score	AUC
Efficient Net-B0	0.9925	1.0000	0.9843	0.9921	1.0000

3.3.1. ROC Curve and AUC Analysis of Efficient Net B0 Model

The ROC curve for the EfficientNet-B0 model is presented in Figure 4. The curve demonstrates an almost ideal trajectory, rising steeply towards the top-left corner of the graph. The model achieved an AUC of 1.000, which indicates perfect classification capability between hepatocellular carcinoma (HCC) and non-HCC cases. This flawless performance confirms EfficientNet-B0's superiority in balancing sensitivity and specificity, making it the best-performing model in this study.

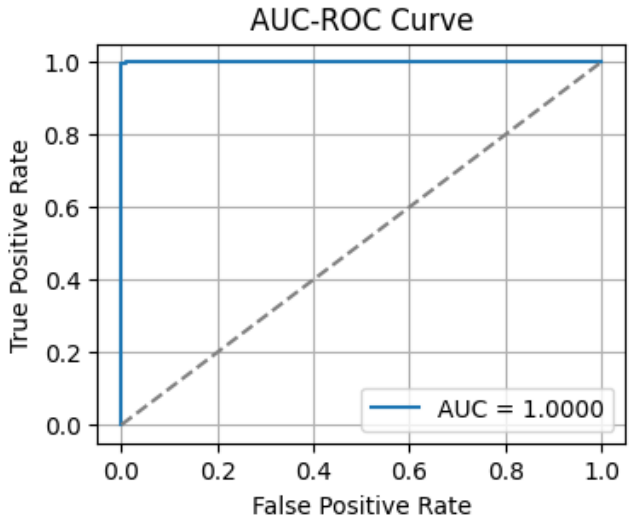

Figure 4. AUC-ROC Curve of Efficient Net B0 Model

3.3.2. Efficient Net B0 Confusion Matrix

The confusion matrix of the EfficientNet-B0 model is shown in Figure 5. The matrix provides deeper insight into the classification ability of the model beyond overall accuracy. Out of 210 hepatocellular carcinoma (HCC) cases, the model correctly classified all 210 as HCC, resulting in zero false negatives, which highlights the model's excellent sensitivity. Similarly, out of 191 non-HCC cases, the model correctly identified 188 cases and misclassified only 3 cases as HCC.

From these results, the following performance metrics can be inferred:

- True Positives (TP): 210
- True Negatives (TN): 188
- False Positives (FP): 3
- False Negatives (FN): 0

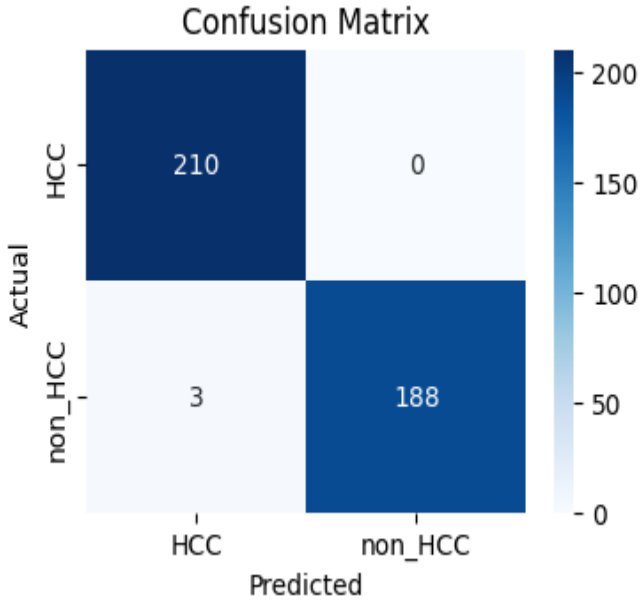

Figure 5. Efficient Net B0 Confusion Matrix

3.4. Results of TinyVit

The implementation and training of the proposed framework were carried out using Google Colab, a cloud-based platform that provides GPU-accelerated computing resources for efficient deep learning experimentation. Among the models evaluated, Tiny Vit demonstrated superior performance, achieving an accuracy of 97.8% highlighting its effectiveness in the liver cancer detection task. The following figures show the training and validation accuracy of the model.

Table 2. TinyViT results in Google Colab

Model	Test Accuracy	Precision	Recall	F1 Score	AUC
Tiny ViT	0.9700	0.9700	0.9700	0.9700	1.0000

3.4.1. TinyViT Confusion Matrix

The confusion matrix of the TinyViT model is shown in Figure 6. The matrix provides deeper insight into the classification ability of the model beyond overall accuracy. Out of 198 hepatocellular carcinoma (HCC) cases, the model correctly classified 186 as HCC, resulting in 12 false negatives. Similarly, out of 191 non-HCC cases, the model correctly classified all 191 as non-HCC, resulting in 0 false positives.

From these results, the following performance metrics can be inferred:

- True Positives (TP): 198
- True Negatives (TN): 191
- False Positives (FP): 0
- False Negatives (FN): 12

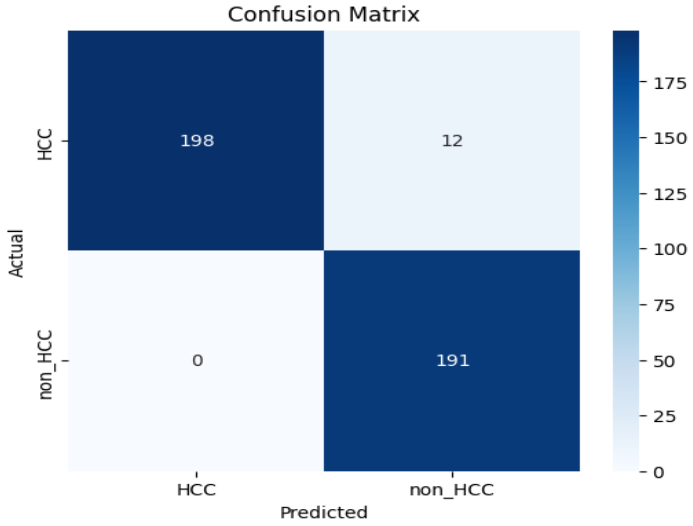

Figure 6. Confusion Matrix of TinyViT Model

3.5. Result of MobileVit v2 Model

The implementation and training of the proposed framework were carried out using Google Colab, a cloud-based platform that provides GPU-accelerated computing resources for efficient deep learning experimentation. Among the models evaluated, MobileVit v2 demonstrated superior performance, achieving an accuracy of 98.5%, highlighting its effectiveness in the liver cancer detection task. The following figures show the training and validation accuracy of this model.

Table 3. MobileViT v2 results in Google Colab

Model	Test Accuracy	Precision	Recall	F1 Score	AUC
Tiny Vit	0.9854	0.9840	1.0000	0.9026	0.9999

3.5.1. ROC Curve of MobileViT v2

The ROC curve for MobileViT v2 is displayed in Figure 7. This model achieved an AUC score of 0.9999, which reflects near-perfect classification performance. The curve remains close to the top-left border, demonstrating high sensitivity and specificity across thresholds.

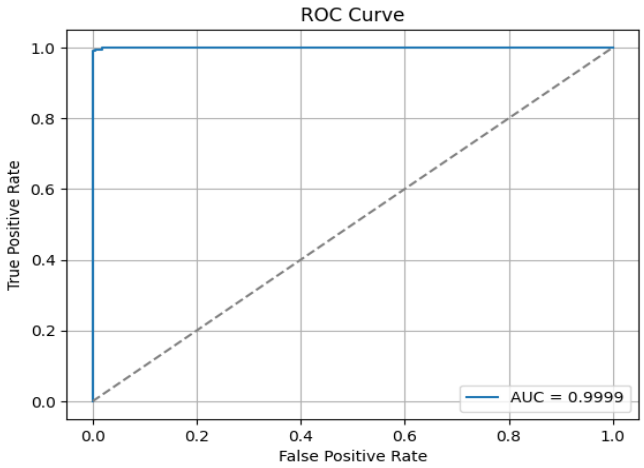

Figure 7. AUC-ROC Curve of MobileViT v2 Model

3.5.2. MobileViT v2 Confusion Matrix

The confusion matrix of the MobileViT v2 model is shown in Figure 8. The matrix provides deeper insight into the classification ability of the model beyond overall accuracy. Out of 169 hepatocellular carcinoma (HCC) cases, the model correctly classified 128 as HCC, resulting in 41 false negatives. Similarly, out of 190 non-HCC cases, the model correctly classified all 190 as non-HCC, resulting in 0 false positives.

From these results, the following performance metrics can be inferred:

- True Positives (TP): 169
- True Negatives (TN): 190
- False Positives (FP): 0
- False Negatives (FN): 41

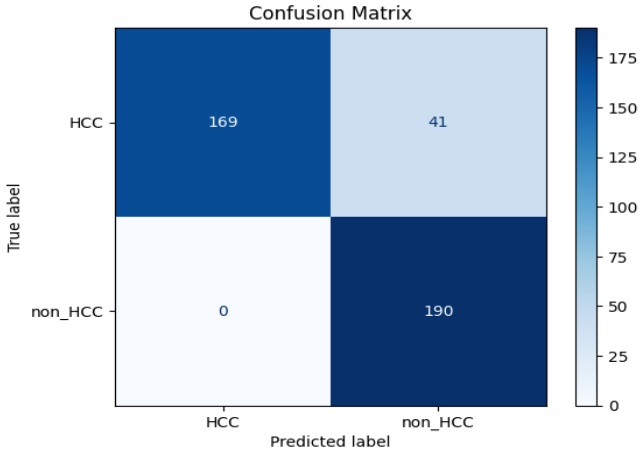

Figure 8. Results of MobileViTv2 Model

3.6. Training Accuracy vs Epoch of all Models

The following Figure shows the training accuracy vs epoch of Efficient Net B0, TinyVit and MobileVit v2 model.

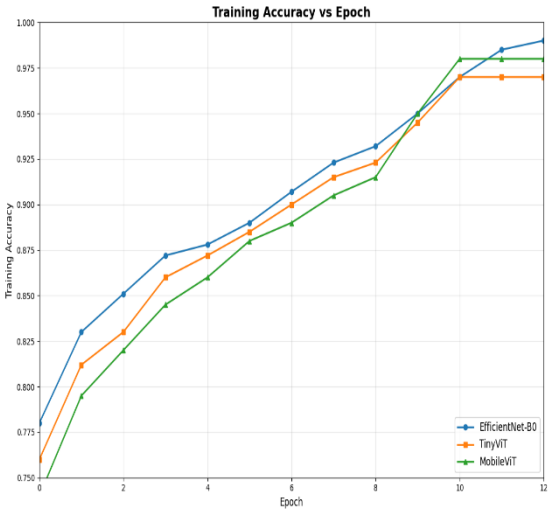

Figure 9. Results of MobileViTv2 Model

3.7. Model comparison on Test Accuracy

The following Figure shows the test accuracy comparison of Efficient Net B0, TinyVit, and MobileViT v2 models.

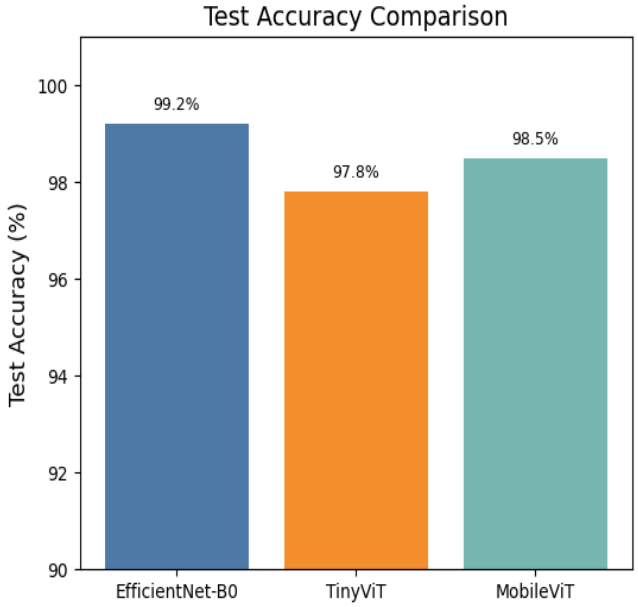

Figure 10. Results of MobileViTv2 Model

3.8. Results Comparison of Efficient Net B0, TinyVit, MobileViT v2 Models

The following table shows all the results Comparison of Efficient Net B0, TinyVit, and MobileViT v2 Model.

Table 4. All Results Comparison of Models

Models	Accuracy	Precision	Recall	F1 Score	AUC
Efficient Net B0,	0.9925	1.0000	0.9843	0.9921	1.0000
Tiny Vit	0.9700	0.9700	0.9700	0.9700	1.00
MobileViT v2	0.9854	0.9840	1.0000	0.9026	0.9999

3.9. Model Comparison on Text Metrix

The following figure shows the model comparison on test metrics.

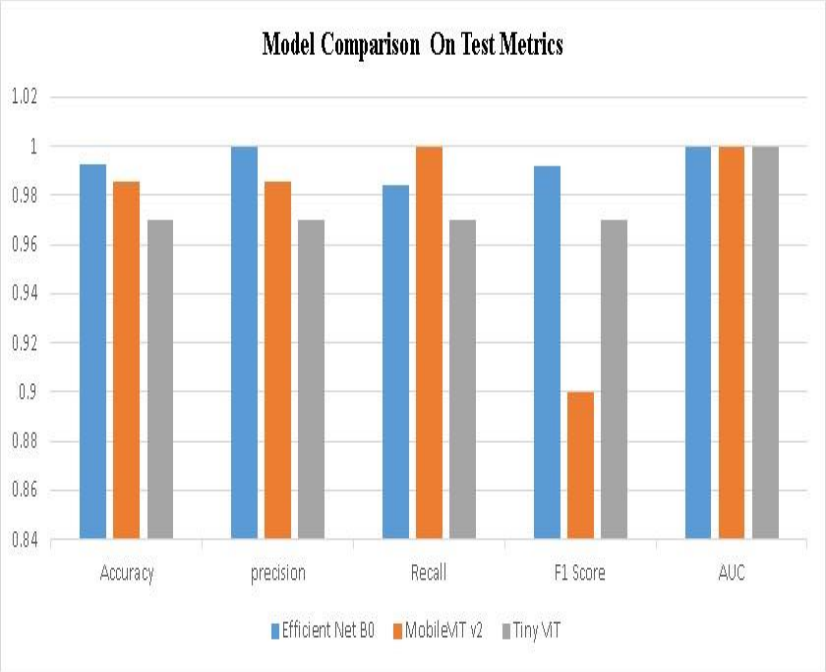

Figure 11. Model Comparison on Text Metrix

3.10. GRAD-CAM Implementation Visual

In this study Gradient-weighted Class Activation Mapping (Grad-CAM) was implemented to provide interpretability and visual explanations of the model's predictions. The primary motivation behind incorporating Grad-CAM was to ensure that the framework does not function as a "black box" but instead highlights the regions of interest within the CT scans that directly influenced the classification of liver cancer. Figure 12 presents a Grad-CAM visualization generated from one of the CT scans in the dataset. The heatmap overlay clearly demonstrates that the proposed model concentrated its attention on the liver region, specifically highlighting the area with abnormal tissue patterns. The red and yellow zones in the visualization represent regions of high activation, signifying areas most strongly correlated with the prediction of liver cancer, while the blue regions correspond to less relevant areas. This indicates that the model is correctly identifying clinically significant tumor regions rather than being distracted by irrelevant features within the scan. The successful implementation of Grad-CAM in this research adds a crucial interpretability layer to the classification framework. By providing visual justification for each prediction, the approach not only enhances trust in the model's decision-making process but also increases its potential acceptance in clinical settings. Moreover, this interpretability aspect allows radiologists and medical experts to cross-validate the predictions with their domain knowledge, ensuring greater reliability and transparency of the AI-driven diagnostic system.

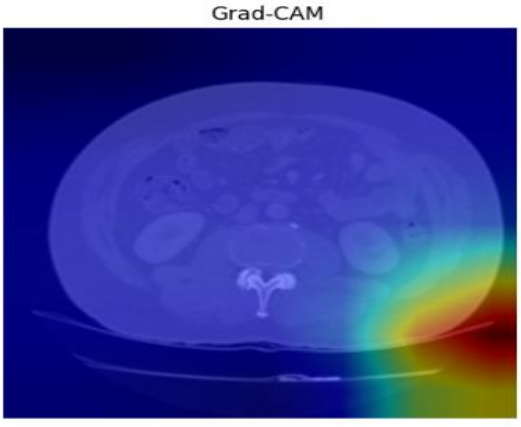

Figure 12. GRAD-CAM Implementation Visual

4. Accuracy Comparison against other work

The following table compares the accuracy results of the proposed models with those of other authors' work.

Table 5. Accuracy comparison of proposed models with existing studies

Author (s)	Method	Accuracy
(Ahmed, Rahouma, & Massoud, 2024)	U-Net, ResNet50	96%
(Ma, Gong, Qiu, Ma, & Yu, 2024)	Radiomics, SVM, KNN, MLP Random Forest, XGBoost, LightGBM	88.2%
(Slama, Sahli, Amri, & Labidi, 2025)	Hybrid V-Net, VGG16	96.52%
(Luo, et al., 2025)	UNet++	93.6%
(Lal, et al., 2025)	ML techniques (CNN, RNN, SVM, KNN)	CNN: 96.7%, RNN:94.5% KNN: 92.2%, SVM: 89.9%
(Gowda & Manjunath, 2025)	UNet70	94.58 %
Proposed Methods	Efficient -Net B0, Tiny ViT, MobileViT v2	Efficient -Net B0: 99.2% Tiny ViT: 97.8% MobileViT v2: 98.5%

5. Conclusion

The study concludes that deep learning, when applied through advanced CNN architectures, holds substantial promise in addressing the limitations of conventional diagnostic practices for liver cancer. By systematically evaluating EfficientNet-B0, TinyViT, and MobileViT v2, the research identified EfficientNet-B0 as the optimal model, with a classification accuracy of 99.2 percent on the 3Dircadb1 dataset. Its ability to combine accuracy, efficiency, and robustness demonstrates that CNN-based models, when carefully scaled and optimized, can outperform more complex architectures in specific medical imaging applications. The integration of Grad-CAM further strengthened the clinical relevance of this research by offering transparency and interpretability. Predictions were not only accurate but also explainable, addressing the critical concern of AI models functioning as "black boxes." This capability ensures that diagnostic support provided by the system can be trusted by radiologists and other healthcare professionals. Taken together, these findings suggest that EfficientNet-B0 represents a highly effective candidate for integration into AI-assisted diagnostic workflows for liver cancer detection.

Funding: This research received no external funding.

Data Availability Statement: Data will be made available on request.

Acknowledgments: Not Applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Ahmed, H. A., Rahouma, K. H., & Massoud, M. A. (2024). Automated detection of primary liver cancer using different deep learning approaches. *Journal of Advanced Engineering Trends*, *43*(1), 433-449.
2. Alawyia, B., & Constantinou, C. (2023). Hepatocellular Carcinoma: a Narrative Review on Current Knowledge and Future Prospects. Current Treatment Options in Oncology, 24(7), 711–724.
3. Anil B. C., Dayananda P., Nethravathi B., & Raisinghani, M. S. (2021). Efficient Local Cloud-Based Solution for Liver Cancer Detection Using Deep Learning. International Journal of Cloud Applications and Computing, 12(1), 1–13.
4. Bourgade, R., Rabilloud, N., Perennec, T., Pécot, T., Garrec, C., Guédon, A. F., … Loussouarn, D. (2023). Deep Learning for Detecting BRCA Mutations in High-Grade Ovarian Cancer Based on an Innovative Tumor Segmentation Method From Whole Slide Images. Modern Pathology, 36(11).
5. Bousabarah, K., Letzen, B., Tefera, J., Savic, L., Schobert, I., Schlachter, T., … Lin, M. De. (2021). Automated detection and delineation of hepatocellular carcinoma on multiphasic contrast-enhanced MRI using deep learning. Abdominal Radiology, 46(1), 216–225.
6. Chen, Z., Dou, M., Luo, X., & Yao, Y. (2025). Enhanced Liver and Tumor Segmentation Using a Self-Supervised Swin-Transformer-Based Framework with Multitask Learning and Attention Mechanisms. Applied Sciences (Switzerland), 15(7).
7. Diao, Z., Jiang, H., & Zhou, Y. (2023). Leverage prior texture information in deep learning-based liver tumor segmentation: A plug-and-play Texture-Based Auto Pseudo Label module. Computerized Medical Imaging and Graphics, 106, 102217.
8. Gao, R., Zhao, S., Aishanjiang, K., Cai, H., Wei, T., Zhang, Y., … Gu, J. (2021, December 1). Deep learning for differential diagnosis of malignant hepatic tumors based on multi-phase contrast-enhanced CT and clinical data. Journal of Hematology and Oncology. BioMed Central Ltd.
9. Ghoniem, R. M. (2020). A Novel Bio-Inspired Deep Learning Approach for Liver Cancer Diagnosis. Information (Switzerland), 11(2).
10. Gowda, Y., & Manjunath, R. V. (2025). Automatic liver tumor classification using UNet70 a deep learning model. Journal of Liver Transplantation, 18, 100260.
11. Kang, M., Ting, C. M., Ting, F. F., & Phan, C. W. R. (2024, October). Cafct-Net: A Cnn-Transformer Hybrid Network With Contextual And Attentional Feature Fusion For Liver Tumor Segmentation. In 2024 IEEE International Conference on Image Processing (ICIP) (pp. 2970-2974).
12. Kriegsmann, M., Kriegsmann, K., Steinbuss, G., Zgorzelski, C., Albrecht, T., Heinrich, S., … Gaida, M. M. (2023). Implementation of deep learning in liver pathology optimizes diagnosis of benign lesions and adenocarcinoma metastasis. Clinical and Translational Medicine, 13(7).
13. Kulkarni, N. M., Fung, A., Kambadakone, A. R., & Yeh, B. M. (2021, August 1). Computed Tomography Techniques, Protocols, Advancements, and Future Directions in Liver Diseases. Magnetic Resonance Imaging Clinics of North America. W.B. Saunders.
14. Lal, B., Prasad, K. S., Vinmathi, M. S., Purohit, P., Ashutosh, P., & Tharsanee, R. M. (2024). Applying machine learning techniques for the diagnosis and prognosis of liver cancer. In Challenges in Information, Communication and Computing Technology (pp. 33–38).
15. Lee, I. C., Tsai, Y. P., Lin, Y. C., Chen, T. C., Yen, C. H., Chiu, N. C., … Huang, Y. H. (2024). A hierarchical fusion strategy of deep learning networks for detection and segmentation of hepatocellular carcinoma from computed tomography images. Cancer Imaging, 24(1).
16. Luo, Y., Yang, Q., Hu, J., Qin, X., Jiang, S., & Liu, Y. (2025). Preliminary study on detection and diagnosis of focal liver lesions based on a deep learning model using multimodal PET/CT images. European Journal of Radiology Open, 14.
17. Ma, Y., Gong, Y., Qiu, Q. T., Ma, C., & Yu, S. (2024). Research on multi-model imaging machine learning for distinguishing early hepatocellular carcinoma. BMC Cancer, 24(1).
18. Mehta, S., & Rastegari, M. (2022). Separable self-attention for mobile vision transformers. arXiv preprint arXiv:2206.02680.
19. Rahoma, K. H., Anwar, H., Ibrahim, A., & Massoud, M. A. (2024). Automated Detection of Primary Liver Cancer using Different Deep Learning Approaches (Vol. 43).
20. Sarfati, E., Bône, A., Rohé, M.-M., Aubé, C., Ronot, M., Gori, P., & Bloch, I. (2025). Guiding the classification of hepatocellular carcinoma on 3D CT-scans using deep and handcrafted radiological features.

21. Slama, A. B., Sahli, H., Amri, Y., & Labidi, S. (2025). V-NET-VGG16: Hybrid deep learning architecture for optimal segmentation and classification of multi-differentiated liver tumors. Intelligence-Based Medicine, 11, 100210.
22. Sridhar, K., Kavitha, C., Lai, W. C., & Kavin, B. P. (2023). Detection of Liver Tumour Using Deep Learning Based Segmentation with Coot Extreme Learning Model. Biomedicines, 11(3).
23. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249.
24. Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.
25. Wan, Y., Zheng, Z., Liu, R., Zhu, Z., Zhou, H., Zhang, X., & Boumaraf, S. (2021). A multi-scale and multi-level fusion approach for deep learning-based liver lesion diagnosis in magnetic resonance images with visual explanation. Life, 11(6).
26. Wu, K., Zhang, J., Peng, H., Liu, M., Xiao, B., Fu, J., & Yuan, L. (2022, October). Tinyvit: Fast pretraining distillation for small vision transformers. In European conference on computer vision (pp. 68-85).
27. Zhan, G., Wang, F., Wang, W., Li, Y., Chen, Q., Hu, H., & Chen, Y. W. (2022, December). A transformer-based model for preoperative early recurrence prediction of hepatocellular carcinoma with muti-modality mri. In Asian Conference on Computer Vision (pp. 185-194).