ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 1 2026

CLASSIFICATION OF LIVER CANCER TYPE HCC AND NON-HCC USING DEEP LEARNING

Sonia Jamil
soniamalikmj@gmail.com
Department of Computer Science & Information Technology, University of Southern Punjab (USP), Multan, Punjab, Pakistan soniamalikmj@gmail.com
Hamid Ghous
hamidghous@usp.edu.pk
Department of Computer Science & Information Technology, University of Southern Punjab (USP), Multan, Punjab, Pakistan
Majid Khawar
majidkhawar76@gmail.com
Department of Computer Science & Information Technology, University of Southern Punjab (USP), Multan, Punjab, Pakistan
Muhammad Talah Zubair
talhazubair009@gmail.com
Department of Computer Science, NCBA&E Lahore Sub Campus Multan, Punjab, Pakistan

ABSTRACT
Liver hepatocellular carcinoma (HCC) is the most prevalent form of liver cancer and a leading cause of cancer-related mortality worldwide. Early and accurate diagnosis is essential for improving patient outcomes. This study aims to develop an automated deep learning-based classification model to distinguish between HCC and non-HCC cases using portal-venous phase CT scans. Acurated dataset comprising approximately 36,000 axial slices from 390 patients (176 HCC-positive and 214 HCC-negative) was utilized, with expert-verified slice-level annotations achieving a Cohen's κ of ≥ 0.95. To prevent data leakage, an 80/20 patient-level split was employed. Preprocessing involved converting DICOM images to PNG format, applying liver-specific windowing (W=350, L=50), resizing to 128×128 pixels, and normalization. Data augmentation techniques rotation, translation, zoom, shear, and flipping were applied to enhance model generalization. A sparse 9-layer convolutional neural network (CNN), comprising approximately 3.3 million parameters, was designed with convolutional, batch normalization, ReLU, max pooling, dense, dropout, and sigmoid layers. The model achieved a validation accuracy of 97.35%, surpassing ResNet50 (90.59%) and closely matching VGG16 (97.64%) and InceptionV3 (97.08%). It also demonstrated high diagnostic metrics, with precision, recall, and F1-score values of 0.96, 1.00, and 0.98, respectively, indicating a reduction in false negatives. These results suggest that a well-tuned CNN can achieve performance comparable to larger transfer learning models with similar computational efficiency. The findings highlight the potential of this model as a clinical decision-support tool for early HCC detection. Future work will focus on multi-center validation and integration into clinical workflows to enhance real-world applicability.

Key words: CNN, Deep Learning, HCC, Non-HCC, Transfer Learning Models

INTRODUCTION

In 2023, liver cancer remains a major global health challenge and is among the leading causes of cancer-related mortality worldwide. (Sarfati et al., 2025) . Recent epidemiological data indicate that it ranks among the most frequently diagnosed cancers and is a leading contributor to cancer deaths, particularly in regions with high prevalence of chronic liver diseases such as hepatitis B and C infections and cirrhosis. Hepatocellular carcinoma (HCC) accounts for approximately 75–85% of primary liver cancer cases, followed by cholangiocarcinoma and other rare subtypes. In addition to primary malignancies, the liver is a common site for metastatic tumors originating from organs such as the colon, pancreas, and lungs.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 1 2026

Early detection of HCC is critical for improving patient survival; however, it remains challenging in clinical practice. Imaging modalities such as computed tomography (CT) and magnetic resonance imaging (MRI) play a central role in diagnosis, staging, and treatment planning. (Saeed et al., 2025a), (Shan et al., 2025). Despite their effectiveness, these techniques have limitations, including high cost, limited accessibility in low-resource settings, and reduced sensitivity in early-stage detection (Hussain et al., 2022a). Artificial Intelligence (AI), particularly deep learning, has emerged as a promising approach to address these challenges. Convolutional Neural Networks (CNNs) have demonstrated strong capability in automatically extracting complex features from medical images, enabling improved detection, classification, and prognostic assessment of liver cancer. Recent studies have explored various architectures, including feedforward networks, recurrent neural networks (RNNs), and CNN-based models, with CNNs showing superior performance in image-based tasks. Additionally, deep learning models applied to histopathological data, such as Hematoxylin and Eosin (H&E) stained slides, have shown potential in predicting patient survival outcomes .However, a significant research gap remains**.** Existing studies often focus either on imaging or histopathological data in isolation, lack robust validation on diverse clinical datasets, and frequently do not address challenges such as class imbalance, model interpretability, and real-world clinical integration. Moreover, limited attention has been given to developing cost-effective, scalable AI solutions suitable for resource-constrained healthcare settings (P. L. H. Yu et al., 2025a).

This study aims to address these limitations by developing and evaluating a CNN-based framework for accurate detection and classification of hepatocellular carcinoma using CT imaging data. The specific objectives are:

1. To design a robust CNN model for feature extraction and classification of liver lesions.
2. To evaluate the model's performance using clinically relevant metrics.
3. To assess its potential for early detection and prognosis support in HCC patients.

We hypothesize that the proposed deep learning approach will improve diagnostic accuracy compared to conventional methods while maintaining computational efficiency suitable for broader clinical adoption.

Ethical Considerations

This study was conducted using a retrospective clinical dataset and received ethical clearance from the Institutional Human Ethics Committee of The Islamia University of Bahawalpur (Ref. No. 326/ORIC, dated 01/04/2025). The study complies with the principles outlined in the Helsinki Declaration. Patient data were anonymized prior to analysis to ensure confidentiality. The requirement for informed consent was waived by the ethics committee due to the retrospective nature of the study.

Novelty

The proposed research presents a novel deep learning-driven architecture that will be used to group CT images as hepatocellular carcinoma (HCC) and non-HCC. The suggested methodology will be based on CNN architecture and multiple pre-trained models, such as VGG16, ResNet50, and InceptionV3, that will be exploited to use the domain-specific feature extraction and the transfer learning benefits of the models (Khaled et al., 2024) at the same time.

Contribution

This study presents a deep learning–based framework for the classification of hepatocellular carcinoma (HCC) and non-HCC cases using contrast-enhanced CT images and associated clinical data. Rather than proposing a novel network architecture, the work focuses on the systematic

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

evaluation and optimization of established convolutional neural network (CNN) and transfer learning models for improved diagnostic performance in a clinically relevant setting.

The proposed workflow consists of four stages. First, CT images were preprocessed using standardized techniques, including intensity windowing, spatial resizing, and normalization, to ensure consistency across the dataset. Second, feature representations were extracted using both a baseline CNN and pre-trained models (ResNet50, VGG16, and InceptionV3), enabling comparative analysis of architecture performance on liver imaging data. Third, model generalization was improved through regularization strategies such as data augmentation, dropout, and early stopping to mitigate overfitting. Finally, the models were trained and evaluated using appropriate performance metrics to assess their effectiveness in distinguishing HCC from non-HCC cases.

A key contribution of this study lies in the curation and validation of a clinically grounded dataset**,** where all CT images were manually annotated and verified by experienced radiologists from BVH Hospital, ensuring high-quality ground truth labels. Additionally, the study provides a comparative benchmark of widely used deep learning architectures under consistent preprocessing and training conditions, highlighting their relative strengths and limitations for HCC classification.

Related Work

The use of deep learning (DL) in diagnosing liver cancer especially hepatocellular carcinoma (HCC) has received considerable focus for its capacity to automate clinical decisions and improve diagnostic precision. In contrast to conventional machine learning methods that depend largely on manually designed features, convolutional neural networks (CNNs) facilitate automatic feature extraction from medical images, rendering them very efficient for intricate tasks like tumor detection and classification. Initial research, including works by (Chen et al., 2020), highlighted the practicality of employing CNN frameworks such as Inception V3 on histopathological H&E images from the Genomic Data Commons for the classification of HCC. Although these methods demonstrated encouraging classification results, they were constrained to single-modality datasets and did not possess generalizability across different imaging types. In a similar vein, (Azer, 2019) emphasized the advantages of CNN-based models compared to traditional methods but omitted quantitative comparisons across various datasets. Later studies focused on hybrid and transfer learning frameworks to tackle limited data and enhance performance. (Mizouri, 2022) suggested a computer-aided diagnosis (CAD) system that utilizes pre-trained VGG-16 and MobileNet-V1 models, employing transfer learning to enhance classification precision. Nonetheless, these models frequently experience overfitting when trained on small, single-center datasets, and their performance is seldom assessed across multi-institutional groups. Recent developments have concentrated on segmentation and multi-task learning, which are essential for accurate tumor localization. For example, (Dong et al., 2020) presented a hybrid fully convolutional neural network (FCNN) that integrates Inception modules and residual connections, resulting in enhanced segmentation and classification effectiveness. In a similar vein, (Rahman et al., 2025) integrated ResUNet with Inception v4 for 3D CT liver and tumor segmentation, showing improved spatial feature learning. The UNet70 model enhanced classification by utilizing an encoder–decoder structure to differentiate hepatocellular, metastatic, and normal liver tissues. Despite these advancements, many studies present findings based on restricted benchmark datasets like 3D-IRCADb-01, CHAOS, and MSD, lacking adequate evaluation across varied clinical populations. Alongside CNN-based techniques, methods driven by optimization and hybrid machine learning strategies have been investigated (Sridhar et al., 2023) employed an Extreme Learning Machine (ELM) combined with Coot Optimization Algorithm (COA) for optimizing parameters, resulting

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 1 2026

in effective classification. Likewise, (Mostafa et al., 2024) showed that implementing feature reduction methods greatly enhances classification accuracy, with models like Naïve Bayes, SVM, and KNN reaching as high as 97.33%. Nonetheless, these methods rely on manual feature engineering and do not possess the scalability of deep learning systems.

Generative models like GANs have been utilized for CT image enhancement (Khan et al., 2023) to improve image quality and robustness, resulting in better performance in downstream classification tasks. Additionally, federated learning methods (Naga Srinivasu et al., 2024) have been proposed to tackle data privacy issues by facilitating joint model training among organizations without the exchange of raw data. Although they show potential, these techniques remain in the initial phases and do not have standardized assessment procedures

Materials and methods

Using a large slice-level dataset of liver CT images, the proposed framework classifies patients into *HCC-positive* and *non-HCC* categories. The framework consists of four major steps. In the first step, the CT dataset was preprocessed, including windowing, resizing, and normalization. In the second step, feature representations were automatically extracted using convolutional and transfer learning models. Thirdly, model parameters were optimized using data augmentation, dropout, and early stopping techniques to minimize overfitting. Finally, deep learning classifiers such as CNN, ResNet50, VGG16, and InceptionV3 were trained and evaluated to improve classification accuracy. To ensure ground-truth validation of the dataset, all CT images were manually annotated and verified by experienced radiologists from BVH hospital.

The Classification of HCC and Non-HCC (COHAN) framework is shown below section.

The proposed algorithm for the deep learning–based paradigm for **Hepatocellular Carcinoma (HCC) detection using CT scans**

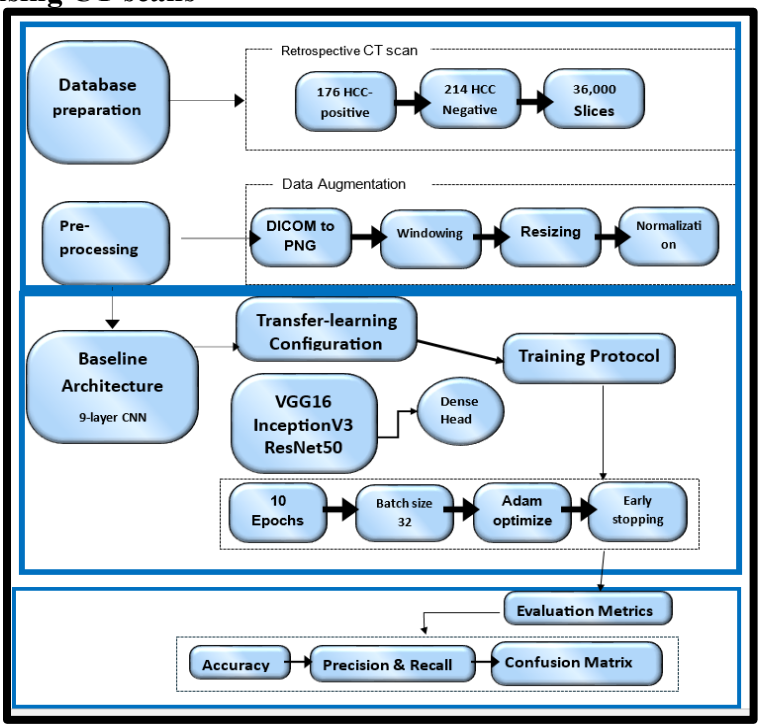

Figure 3.1 Proposed Framework of Classification of HCC and Non-HCC

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

Data Acquisition and Annotation

The study utilized a retrospectively collected dataset of portal-venous-phase CT scans from 390 patients obtained at the Radiology Department of Bahawalpur Victoria Hospital, Pakistan. Among these patients, 176 were HCC-positive and 214 were HCC-negative, generating approximately 36,000 axial CT slices. All images were annotated at the slice level based on radiology reports and validated by two hepatobiliary specialists, achieving a high inter-observer agreement (Cohen's $\kappa \geq 0.95$). The CT scans were acquired using a Toshiba Aquilion Prime TSX-303A scanner with a slice thickness of 0.39–0.45 mm. To avoid data leakage and ensure clinical robustness, the dataset was split at the patient level into training (80%) and validation (20%) sets while preserving class balance. The portal-venous phase was selected due to its superior ability to visualize liver lesions and vascular structures for accurate HCC classification. CT imaging is a common diagnostic modality in the detection of tumors in organs like lungs, kidneys, and liver compared to other modalities (Ling et al., 2022; Tajima et al., 2007). Computer-enhanced tomography (CECT) is conducted on an invasive basis whereas unenhanced CT is conducted on a non-invasive basis. The CT scans are also better in spatial resolution and visualization of vascular structures and allow radiologists to define the affected parts of liver more accurately (Singh et al., 2023).

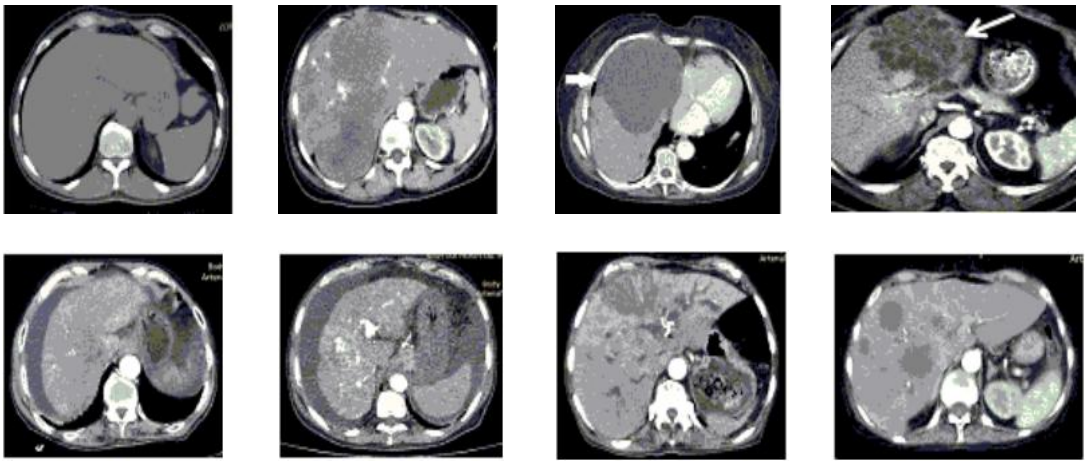

Figure 3.2 HCC and Non- HCC CT images Dataset

Pre-processing Pipeline

The steps of the pipeline were the following (Figure 3.3):

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

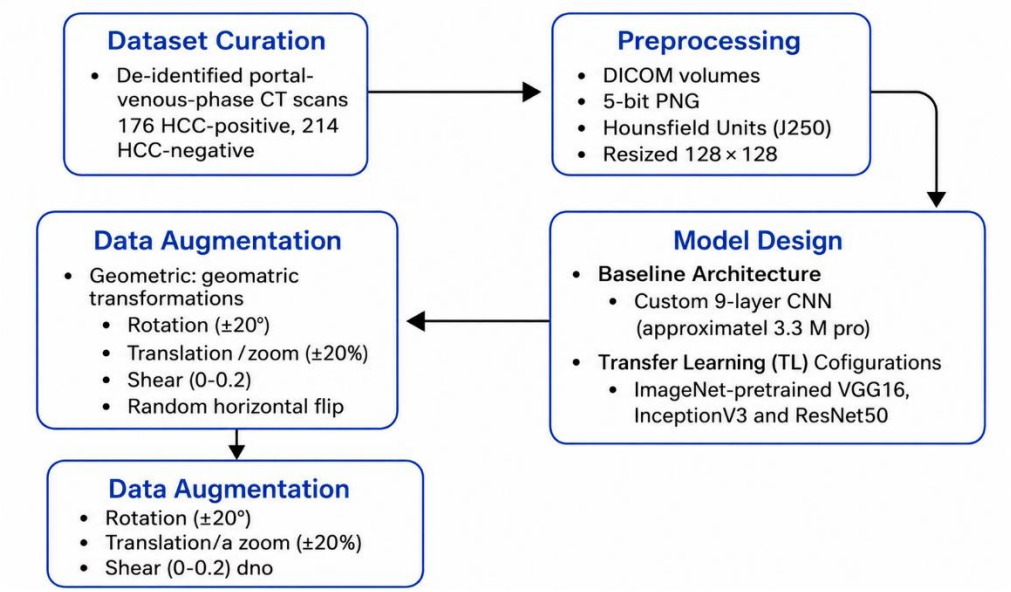

Figure 3. 3 Pre-Processing Steps of COHAN Framework

The CT preprocessing pipeline began with converting DICOM images to PNG format using appropriate software and Python libraries, including intensity rescaling via slope and intercept values. Liver-specific windowing (WL = 50 HU, WW = 350 HU) was applied to enhance soft-tissue contrast and improve lesion visibility. All slices were resized from 512 × 512 to 128 × 128 using bilinear interpolation to reduce computational cost while preserving critical anatomical details. Min–Max normalization scaled pixel intensities to the range [0, 1], ensuring uniform intensity distribution and stable model training. To reduce overfitting and improve generalization, data augmentation techniques such as rotation, translation, zooming, shearing, and horizontal flipping were applied randomly during training. This standardized preprocessing improved consistency, efficiency, and robustness of the deep learning models for HCC classification.

Baseline CNN Architecture

A lightweight 9-layer CNN was designed to achieve an effective balance between classification accuracy, interpretability, and computational efficiency for HCC detection. The network consists of an input layer, feature extraction blocks, and a classification head, making it suitable for potential real-time clinical use. The input layer accepts preprocessed liver CT slices of size 128 × 128 × 1, obtained after windowing, resizing, and Min–Max normalization. Grayscale images were used because CT scans inherently contain diagnostic information only in intensity values, not color. Using a single grayscale channel reduces computational complexity, minimizes noise, and prevents learning redundant features. Overall, this input design ensures standardized, efficient, and diagnostically rich data for robust HCC classification.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

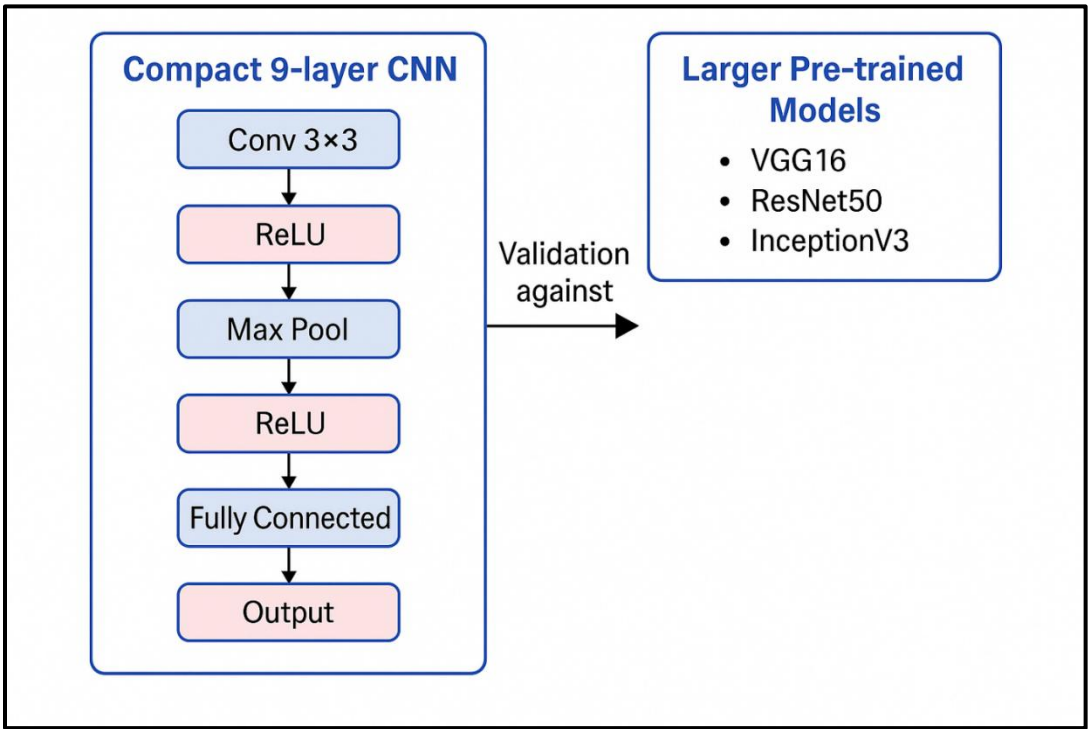

Figure 3.4 Baseline 9-Layer CNN Architecture

Feature Extraction Block

The feature extraction stage of the CNN is designed to automatically learn meaningful patterns from liver CT images using three sequential convolutional units. Each unit follows a Conv2D → Batch Normalization → ReLU → MaxPooling structure, enabling stable and efficient learning. The convolutional layers use progressively increasing filters (32, 64, and 128 with 3×3 kernels) to capture hierarchical features, ranging from fine textures and edges to complex tumor morphology and global anatomical patterns. Batch Normalization stabilizes training, accelerates convergence, and improves generalization across heterogeneous CT data. ReLU activation introduces non-linearity, enhances feature sparsity, and helps capture subtle intensity variations critical for early HCC detection. Overall, this design balances clinical interpretability, computational efficiency, and robust tumor feature learning.

Classification Head

The classification head converts the extracted convolutional features into a final HCC prediction using a sequence of fully connected operations. The Flatten layer reshapes multi-dimensional feature maps into a single vector, enabling global interpretation of learned tumor-related patterns. A dense layer with 128 units and ReLU activation learns high-level feature interactions, while a 50% Dropout layer reduces overfitting by preventing reliance on specific neurons. The final output layer consists of a single neuron with a sigmoid activation, producing a probability score between 0 and 1. This score represents the likelihood of HCC presence and is thresholded (typically at 0.5) to classify slices as HCC or non-HCC. Overall, this design ensures robust, generalizable, and clinically interpretable predictions.

Parameter Efficiency

The baseline CNN was designed to be computationally efficient, containing about 3.3 million parameters while maintaining strong predictive performance. To enhance feature learning and benchmark performance, transfer learning was applied using pretrained models VGG16,

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

InceptionV3, and ResNet50, which leverage knowledge from large ImageNet datasets. CNNs are specialized deep learning models for image analysis, consisting of feature extraction layers (convolution and pooling) and classification layers (dense and output layers). VGG16 uses uniform 3×3 convolutions and deep hierarchical feature learning, making it effective for texture and shape detection. InceptionV3 improves efficiency using multi-scale parallel convolutions and global average pooling, enabling strong feature extraction with fewer parameters. ResNet50 uses residual (skip) connections to train very deep networks effectively, and when fine-tuned with custom classification heads, all transfer learning models improve HCC detection performance on CT images.

Evaluation Matric

Slice-level accuracy was used as the primary evaluation metric to measure overall model performance on the validation dataset. Additional metrics including precision, recall, and F1-score were calculated to evaluate how well the model distinguishes between HCC-positive and HCC-negative cases. A confusion matrix was used to analyze prediction outcomes (True Positive, True Negative, False Positive, False Negative), with special focus on minimizing false negatives due to their high clinical risk in cancer detection. Accuracy measures overall correct predictions, precision measures correctness of positive predictions, recall measures ability to detect actual positives, and F1-score balances precision and recall. Error-based metrics such as MAE and MSE were also computed to quantify prediction error between actual and predicted outputs. Together, these metrics ensured comprehensive evaluation of model performance and generalization to unseen clinical data.

RESULTS & DISCUSSION

The experimental results of the proposed convolutional neural network (CNN)-based type of approach used in the automatic classification of liver CT images into a cancerous and non-cancerous category. The outputs are obtained through systematic experiments which have dataset preparation, preprocessing, training and validation. A lot of quantitative data has been presented with a critical analysis that places the data into perspective as far as current research is concerned. The chapter also covers clinical significance, methodological strengths and limitations of the study and gives some recommendations on potential future research directions.

Dataset Prepossessing Outcomes

The data set included 28,284 training data and 8,457 validation data that were categorized into two distinct classes, which are Cancer and Non-Cancer. To secure strength, a dataset was divided on the level of patients, not to leak the information between the training and validation sets. The pixel intensity normalization, resizing, and augmentation (rotation, flipping, scaling, and translation) in the preprocessing methods enhanced the generality of the model and eliminated the risk of overfitting.

Baseline Architecture of Customized CNN

The proposed CNN model consisted of three convolutional pooling layers, a fully connected dense layer, and an output layer, with approximately 3.3 million trainable parameters to learn complex medical image features. The model used Dropout for regularization and was trained using the Adam optimizer with binary cross-entropy loss for efficient HCC classification. The proposed CNN included three convolutional blocks with increasing filter depths (32, 64, and 128) followed by max-pooling layers, a 128-neuron dense layer, dropout regularization, and a final sigmoid classifier. The network contained approximately 3.3 million trainable parameters, providing strong feature extraction with computational efficiency. The model was trained for 10 epochs with

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 1 2026

optimized batch size and learning rate, while monitoring training and validation metrics to ensure good convergence and generalization.

Table 5.1 Summary of the CNN architecture used in this study

Layer (Type)	Output Shape	Parameters
Conv2D (32 filters)	(None, 126, 126, 32)	896
MaxPooling2D	(None, 63 × 63 × 32)	0
Conv2D (64 filters)	(None, 61 × 61 × 64)	18,496
MaxPooling2D	(None, 30 × 30 × 64)	0
Conv2D (128 filters)	(None, 28 × 28 × 128)	73,856
MaxPooling2D	(None, 14 × 14 × 128)	0
Flatten	(None, 25,088)	0
Dense (128 neurons)	(None, 128)	3,211,392
Dropout (0.5)	(None,128)	0
Dense (Output)	(None,1)	129

Quantitative Result of CNN

The proposed CNN consisted of three convolutional blocks with increasing filter depths (32, 64, and 128) followed by max-pooling layers, a 128-neuron fully connected dense layer, dropout regularization, and a final sigmoid classifier. The network contained 3,304,769 trainable parameters, enabling effective feature extraction while remaining computationally efficient. The model was trained for 10 epochs with optimized batch size and learning rate, while training and validation metrics were monitored to ensure convergence, good generalization, and to detect potential overfitting.

The proposed CNN showed strong learning performance over 10 epochs, with training accuracy improving steadily while training loss continuously decreased. The model achieved high early performance with 94.55% training accuracy and 98.44% validation accuracy in the first epoch, later reaching around 98.60% training accuracy by epoch five. Validation accuracy remained stable above 97% across all epochs, indicating strong generalization capability. Although validation loss fluctuated and slightly increased in later epochs, suggesting mild overfitting, overall results confirmed the model's robustness and reliability for liver CT image classification.

Table 5.2 Epoch-wise Training and Validation Performance

Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation loss
1	0.9455	0.9844	0.1372	0.0593
2	0.9698	0.9713	0.0845	0.0839
3	0.9804	0.9797	0.0550	0.0709
4	0.9845	0.9786	0.0421	0.0953
5	0.9860	0.9791	0.0381	0.1014
6	0.9892	0.9816	0.0295	0.0587
7	0.9903	0.9752	0.0292	0.1089
8	0.9902	0.9818	0.0275	0.0627
9	0.9921	0.9726	0.0242	0.1298
10	0.9907	0.9735	0.0274	0.1481

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 1 2026

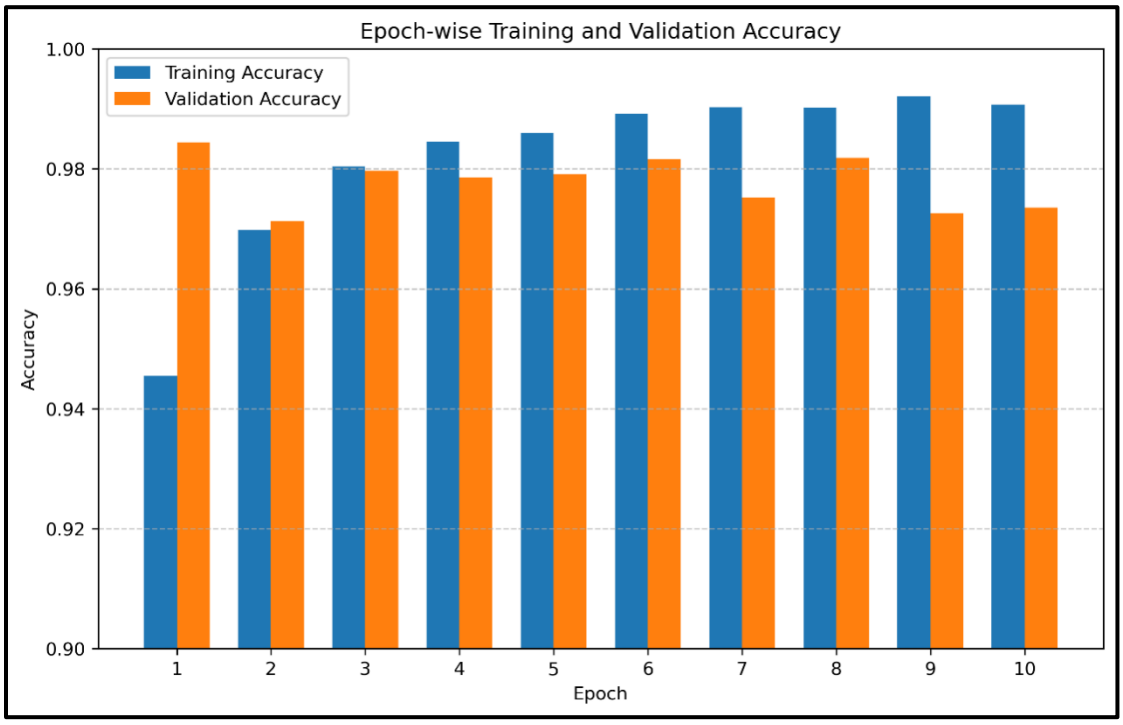

Figure 4.1 Bar chart Epoch- wise Training and Validation Accuracy

Figure 4.2 Bar chart Epoch- wise Training and Validation Loss

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

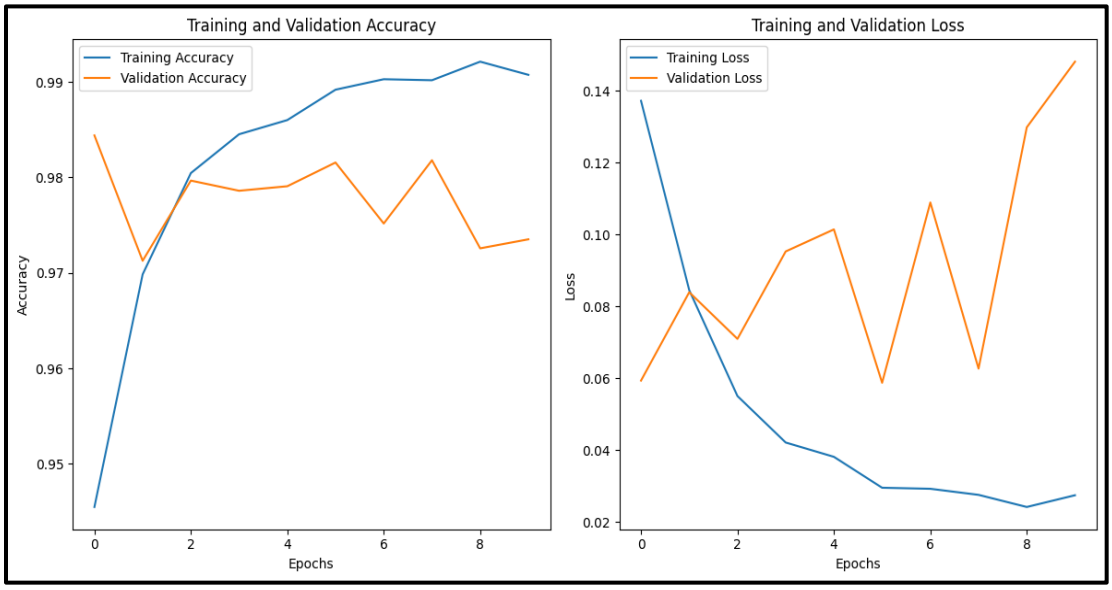

Figure 4.3 Training and Validation Accuracy and Loss

Transfer Learning (TL) Configuration

Transfer learning was applied using ImageNet-pretrained models (VGG16, InceptionV3, and ResNet50) to improve classification performance by leveraging previously learned visual features. The convolutional backbones were frozen, and a lightweight dense classification head (Flatten → Dense 128 + ReLU → 50% Dropout → Sigmoid) with about 1 million trainable parameters was added to adapt the models to liver cancer detection. In the VGG16-based model, CT images were resized and processed through five convolutional blocks that progressively learned low-level edges, mid-level textures, and high-level tumor morphology features. The frozen convolutional base (≈14.7M non-trainable parameters) preserved strong generic representations, while only the dense layers (≈1M parameters) were fine-tuned for liver CT classification. This design reduced training time, minimized overfitting, and improved generalization, making it suitable for medical datasets with limited samples.

Table 5.3 Summary of the VGG16 used in this study

Layer (Type)	Output Shape	Parameters (#)
Input Layer	(None, 128, 128, 3)	0
Block1_Conv1 (Conv2D)	(None, 128, 128, 64)	1,792
Block1_Conv2 (Conv2D)	(None, 128, 128, 64)	36,928
Block1_Pool (MaxPooling2D)	(None, 64, 64, 64)	0
Block2_Conv1 (Conv2D)	(None, 64, 64, 128)	73,856
Block2_Conv2 (Conv2D)	(None, 64, 64, 128)	147,584
Block2_Pool (MaxPooling2D)	(None, 32, 32, 128)	0
Block3_Conv1 (Conv2D)	(None, 32, 32, 256)	295,168

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

Block3_Conv2 (Conv2D)	(None, 32, 32, 256)	590,080
Block3_Conv3 (Conv2D)	(None, 32, 32, 256)	590,080
Block3_Pool (MaxPooling2D)	(None, 16, 16, 256)	0
Block4_Conv1 (Conv2D)	(None, 16, 16, 512)	1,180,160
Block4_Conv2 (Conv2D)	(None, 16, 16, 512)	2,359,808
Block4_Conv3 (Conv2D)	(None, 16, 16, 512)	2,359,808
Block4_Pool (MaxPooling2D)	(None, 8, 8, 512)	0
Block5_Conv1 (Conv2D)	(None, 8, 8, 512)	2,359,808
Block5_Conv2 (Conv2D)	(None, 8, 8, 512)	2,359,808
Block5_Conv3 (Conv2D)	(None, 8, 8, 512)	2,359,808
Block5_Pool (MaxPooling2D)	(None, 4, 4, 512)	0
Flatten (Flatten)	(None, 8192)	0
Dense (Fully Connected)	(None, 128)	1,048,704
Dropout (Regularization)	(None, 128)	0
Output Dense Layer	(None, 1)	129

The VGG16 transfer learning model demonstrated stable and strong performance across 10 training epochs, showing fast convergence due to pretrained feature extraction. The model achieved high initial performance and steadily improved, reaching peak validation accuracy of about 98.12% around epoch eight. Training and validation metrics remained closely aligned, indicating minimal overfitting and strong generalization. Slight validation loss fluctuations in later epochs reflected sensitivity to subtle medical image variations. Overall, the VGG16-based model outperformed the baseline CNN in stability, convergence speed, and validation accuracy for liver CT classification.

Quantitative Result OF VGG16

Table 5.4 Epoch-wise Training and Validation Performance of VGG16 Model

Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
1	0.9331	0.9707	0.1608	0.0836
2	0.9536	0.9697	0.1170	0.0914
3	0.9582	0.9713	0.1067	0.0744

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

4	0.9590	0.9800	0.1048	0.0647
5	0.9636	0.9786	0.0936	0.0631
6	0.9634	0.9766	0.0914	0.0634
7	0.9678	0.9801	0.0858	0.0541
8	0.9689	0.9812	0.0799	0.0625
9	0.9691	0.9734	0.0812	0.0699
10	0.9702	0.9764	0.0767	0.0663

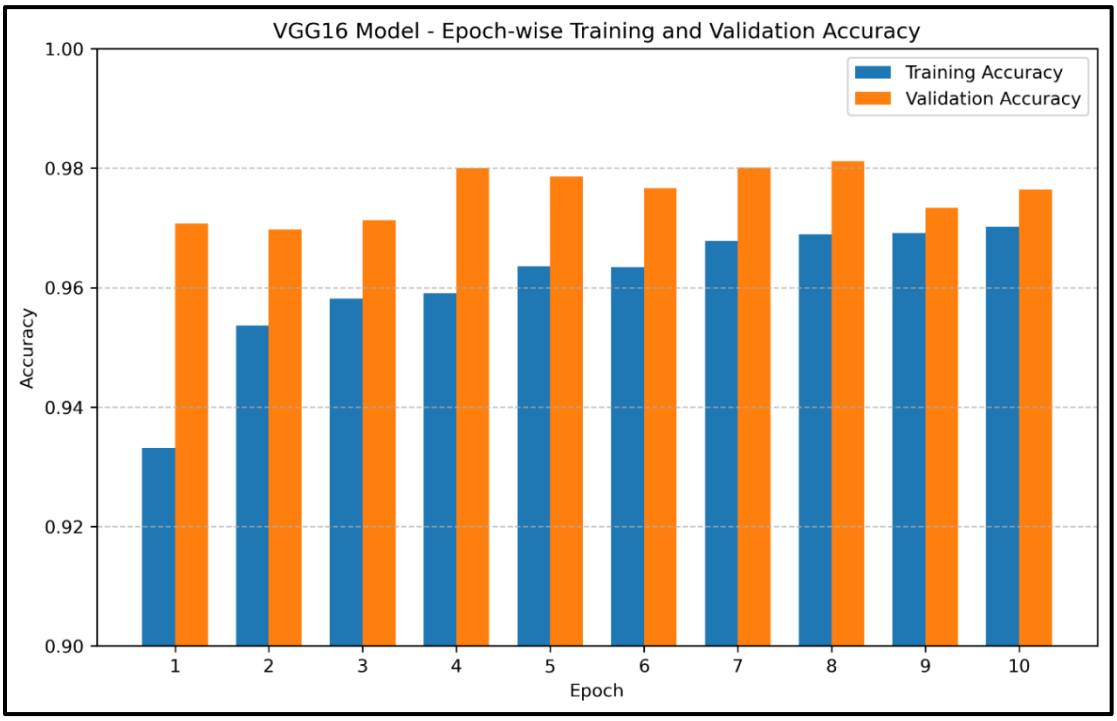

Figure 4.4 Bar graph Epoch-wise Training and Validation Accuracy VGG16 model

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 1 2026

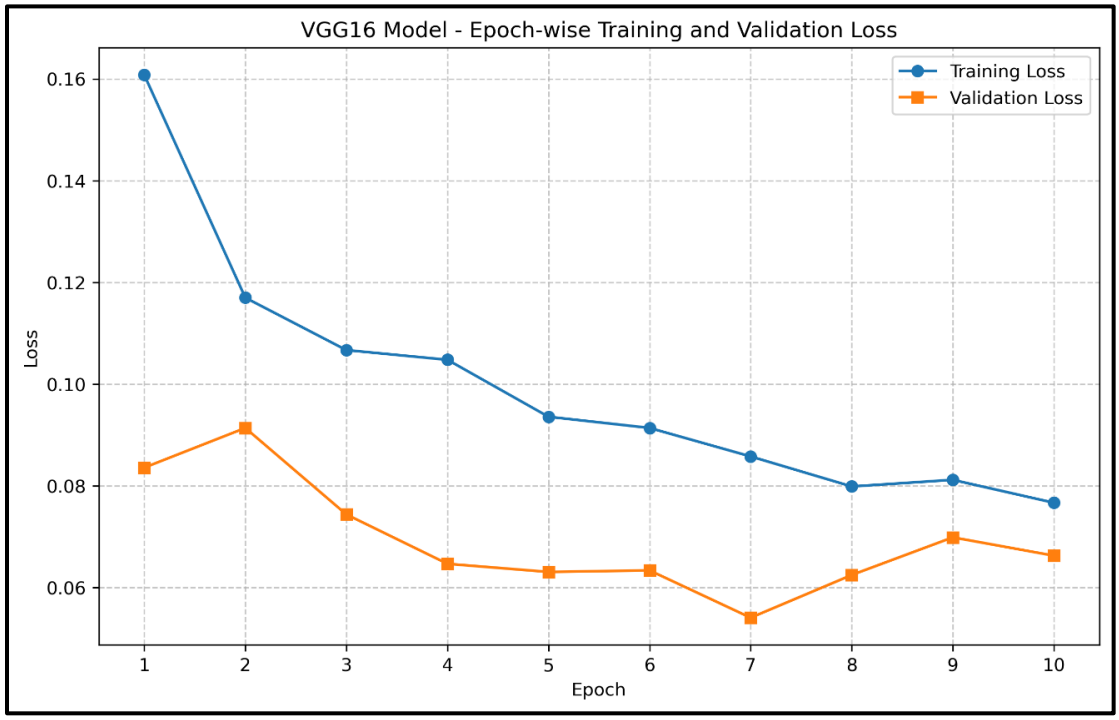

Figure 4.5 Line graph Epoch-wise Training and Validation Loss VGG16 model

The ResNet-based model uses a deep residual convolutional backbone for powerful feature extraction, followed by a lightweight fully connected classification head for liver CT classification. Residual blocks with shortcut connections enable deep hierarchical learning while preventing vanishing gradients and preserving low- and high-level features. The backbone generates high-dimensional feature maps that are flattened and passed to a 128-unit dense layer, dropout regularization, and a final sigmoid output for binary classification. Most trainable parameters are concentrated in the dense head, while the pretrained backbone provides strong generic image representations. Overall, this architecture balances computational efficiency, fast convergence, and strong detection of liver tumor features.

Table 5.5 Summary of the ResNet-based model

Layer / Block (selected)	Output Shape	Parameter Count
Input (input_2)	(None, 128, 128, 3)	0
conv1_pad / conv1_conv	(None, 64, 64, 64)	9,472
conv1_bn (BatchNormalization)	(None, 64, 64, 64)	256
pool1_pool (MaxPooling2D)	(None, 32, 32, 64)	0
conv2_block1_1_conv	(None, 32, 32, 64)	4,160
conv2_block1_2_conv	(None, 32, 32, 64)	36,928
conv2_block1_0_conv (projection)	(None, 32, 32, 256)	16,640
conv2_block1_3_conv	(None, 32, 32, 256)	16,640
conv3_block1_1_conv	(None, 16, 16, 128)	32,896
conv3_block1_2_conv	(None, 16, 16, 128)	147,584
conv3_block1_0_conv (projection)	(None, 16, 16, 512)	131,584
conv4_block1_1_conv	(None, 8, 8, 256)	131,328
conv4_block1_2_conv	(None, 8, 8, 256)	590,080

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 1 2026

conv4_block1_0_conv (projection)	(None, 8, 8, 1024)	525,312
conv5_block1_1_conv	(None, 4, 4, 512)	524,800
conv5_block1_2_conv	(None, 4, 4, 512)	2,359,808
conv5_block1_0_conv (projection)	(None, 4, 4, 2048)	2,099,200
conv5_block3_out (end of backbone)	(None, 4, 4, 2048)	(see above blocks)
flatten_1	(None, 32,768)	0
dense_2 (Dense, 128 units)	(None, 128)	4,194,432
dropout_1	(None, 128)	0
dense_3 (Dense, 1 output)	(None, 1)	129

Quantitative Result of ResNet

The ResNet transfer learning model showed steady improvement over 10 training epochs, benefiting from pretrained ImageNet weights that provided strong initial feature representations. The model started with moderate training accuracy but relatively high validation accuracy, indicating good early generalization due to frozen pretrained layers. As training progressed, both accuracy improved and loss decreased consistently, showing effective adaptation to liver CT data. Residual connections helped maintain stable gradient flow, enabling reliable convergence without major overfitting. Overall, the ResNet model demonstrated strong generalization and effective transfer learning for liver cancer classification.

Table 5.6 Epoch-wise training & validation metrics

Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
1	0.7634	0.8948	0.4659	0.2776
2	0.8234	0.8527	0.3661	0.2786
3	0.8308	0.8718	0.3435	0.2765
4	0.8392	0.9056	0.3267	0.2310
5	0.8361	0.8839	0.3292	0.2669
6	0.8384	0.8982	0.3227	0.2685
7	0.8392	0.9087	0.3178	0.2574
8	0.8494	0.8964	0.3055	0.2602
9	0.8512	0.9047	0.3050	0.2224
10	0.8548	0.9059	0.2946	0.2193

These results indicate that the ResNet model effectively learned discriminative visual features relevant to liver cancer classification while maintaining strong generalization and reduced overfitting compared to baseline CNN and VGG-based models. The achieved validation accuracy

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 1 2026

of around 90–91% highlights the model's potential suitability for real-world clinical classification and diagnostic support tasks where reliability and accuracy are critical.

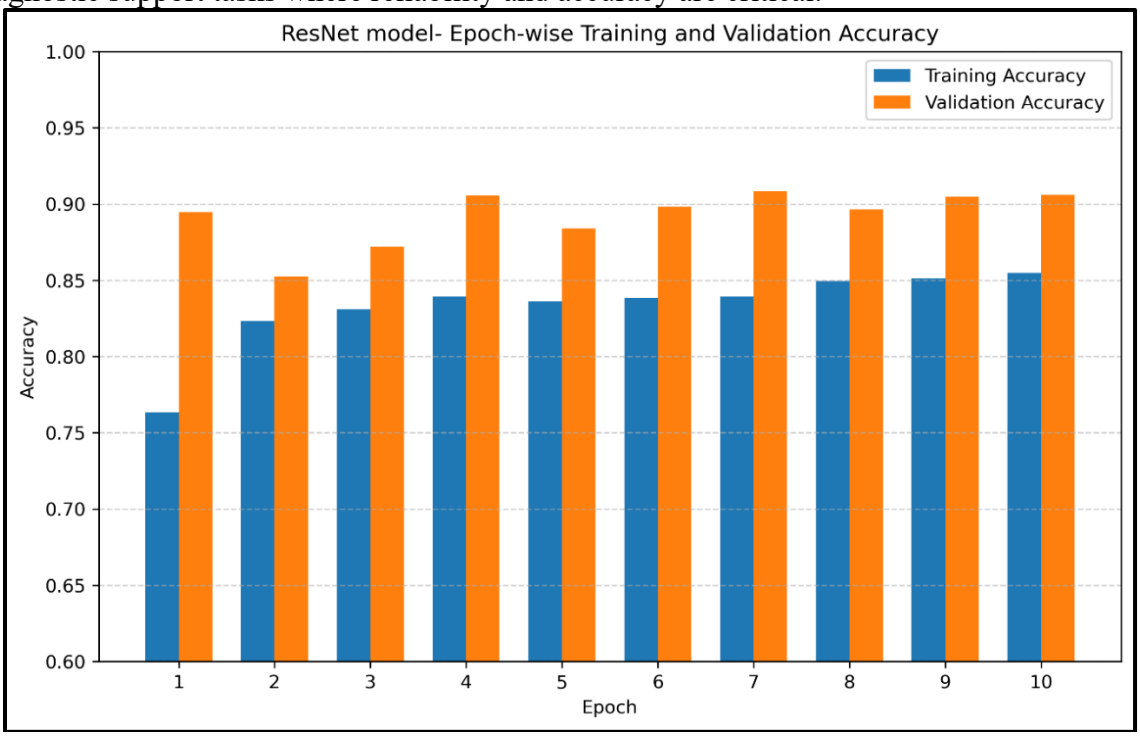

Figure 4.6 Bar Chart of ResNet Model Epoch-wise Training and Validation Accuracy

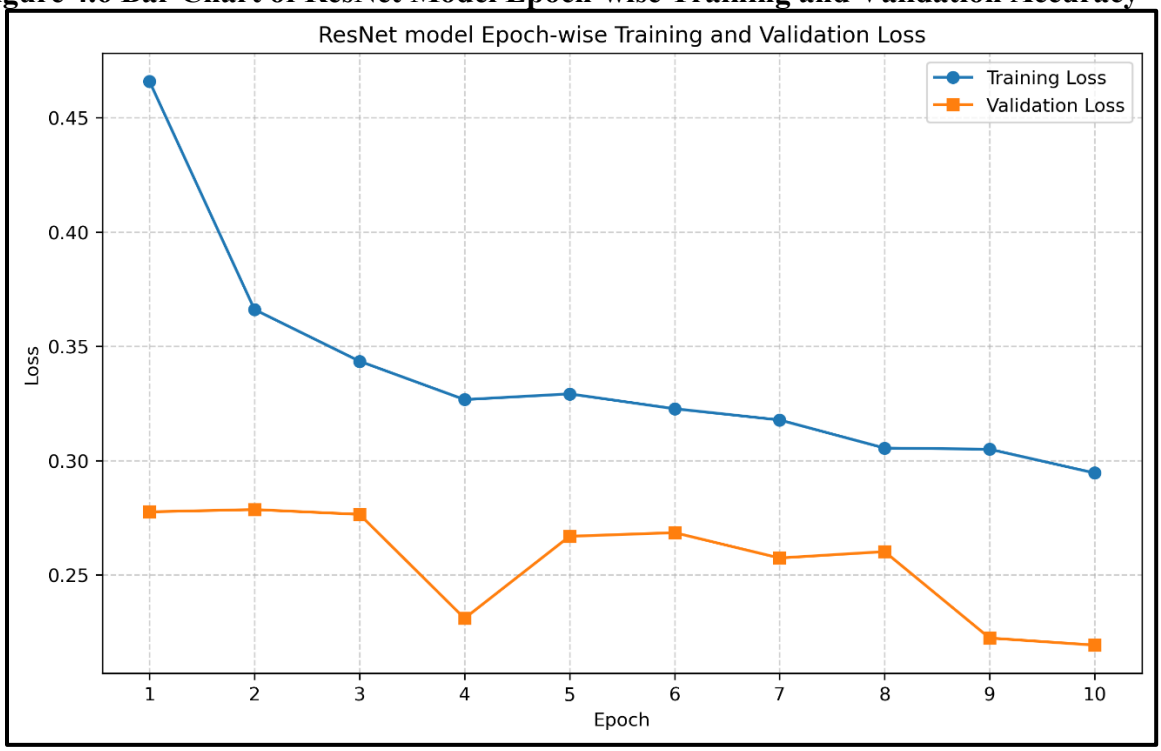

Figure 4.7 Bar Chart of ResNet model Epoch-wise Training and Validation Loss

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

The InceptionV3-based model was used as the main backbone due to its computational efficiency, multi-scale feature extraction, and strong performance in image recognition tasks. It processes liver CT images through parallel Inception modules that capture features at multiple spatial scales, making it effective for detecting lesions of varying size, shape, and texture. To adapt the pretrained model for binary liver cancer classification, a custom classification head (Flatten → Dense 128 → Dropout → Sigmoid) was added. Dropout improved generalization by reducing overfitting, while the sigmoid output produced probability-based cancer predictions. Overall, the InceptionV3 transfer learning model provided powerful feature representation with efficient computation for liver CT classification.

Quantitative Results of Inception V3

Table 5.7 Epoch-wise training & validation metrics

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	0.2115	0.9208	0.1218	0.9500
2	0.1568	0.9359	0.0919	0.9635
3	0.1415	0.9420	0.1120	0.9533
4	0.1375	0.9421	0.0807	0.9671
5	0.1341	0.9465	0.0968	0.9630
6	0.1330	0.9458	0.0972	0.9650
7	0.1222	0.9501	0.0817	0.9688
8	0.1233	0.9513	0.0727	0.9709
9	0.1225	0.9500	0.0863	0.9633
10	0.1232	0.9507	0.0830	0.9708

The InceptionV3 transfer learning model showed strong and stable performance across 10 epochs, achieving high training and validation accuracy from the early stages due to pretrained ImageNet weights. Both training and validation losses decreased steadily, indicating effective learning and strong generalization to unseen liver CT data. The model achieved peak validation accuracy of about 97.09%, demonstrating excellent ability to distinguish cancerous and non-cancerous liver images. The small gap between training and validation accuracy confirmed minimal overfitting. Overall, InceptionV3 proved to be an efficient and reliable model for automated liver cancer classification with good accuracy and moderate computational cost.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 1 2026

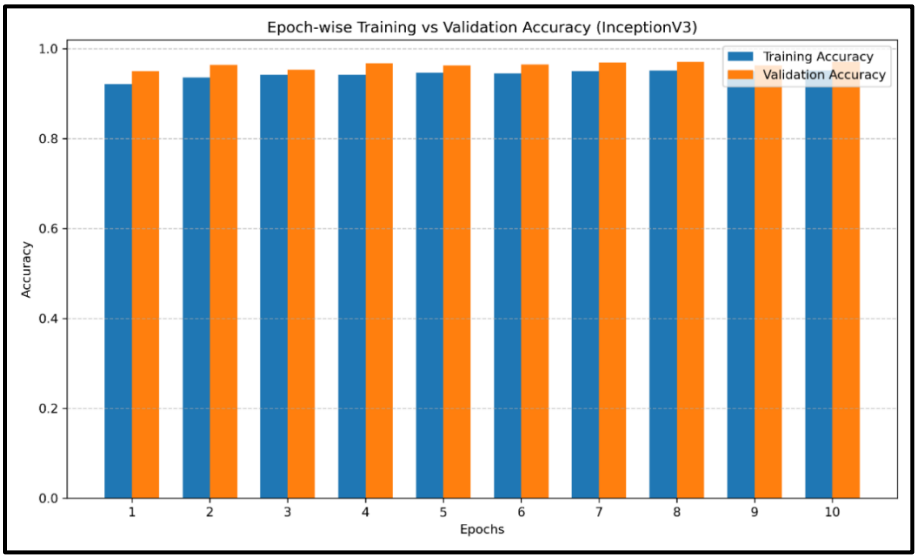

Figure 4.8 Bar Chart of Epoch-wise Training and Validation Accuracy

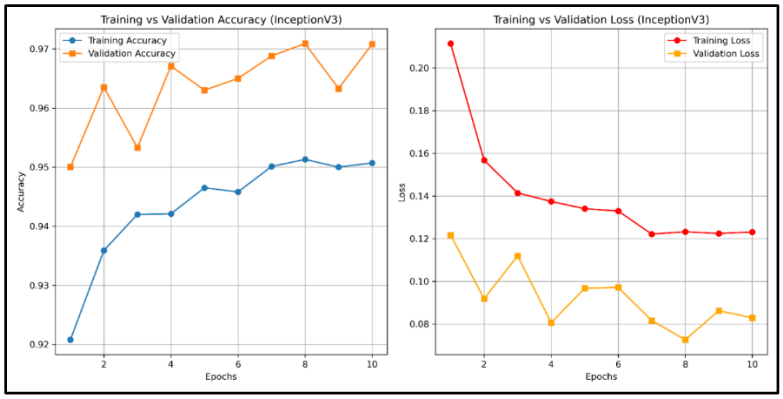

Figure 4.9 Line Chart of Training and Validation Accuracy and Loss

Validation Against of Transfer Learning Model

The baseline CNN used native 128 × 128 input CT images and was trained from scratch using the Adam optimizer with a learning rate of 1e-4. In contrast, transfer learning models (VGG16, ResNet50, InceptionV3) required resizing images to 224 × 224 to match ImageNet input standards. These TL models used frozen pretrained backbones with the same custom dense classification head (Flatten → Dense 128 + ReLU → Dropout → Sigmoid) to ensure fair performance comparison.

Table 5.8 Performance Metrics

Model	Val Accuracy	HCC Recall	Parameters	Inference Speed (ms/img)
Custom CNN	97.35%	1.00	3.3M	8.2
VGG16	97.64%	0.98	138M	12.7
InceptionV3	97.08%	0.97	23.9M	10.1
ResNet50	90.59%	0.89	25.6M	9.8

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 1 2026

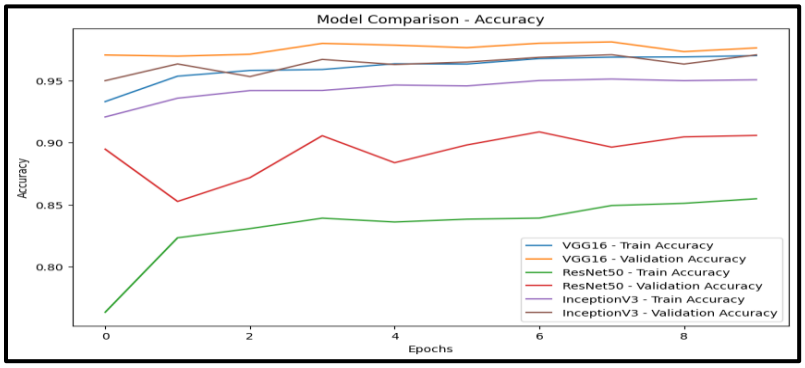

Figure 4.10 Model Comparison Accuracy

Evaluation Metrics

Slice-level accuracy was used as the primary metric to measure overall model performance on the validation dataset. Additional metrics such as precision, recall, and F1-score were calculated to evaluate class-wise discrimination between HCC-positive and HCC-negative cases. A confusion matrix was used to analyze prediction errors, with special emphasis on minimizing false negatives due to their high clinical risk in cancer detection. All evaluations were performed on a held-out validation set to simulate real-world deployment and ensure model generalization to unseen data.

Confusion Metric

Any classification model that is used for performance evaluation must include evaluation metrics. Confusion matrix is an important component of evaluation metrics that indicate the prediction outcomes of any testing and is frequently used to describe the classifier's performance (Hussain et al., 2022).

Each prediction can result in one of four outcomes, depending on how well it fits the actual value: Such as True Positive, False Positive, True Negative and False Negative. If a prediction becomes true in actual is known as True Positive. If prediction showed false and false value is described as True Negative. If prediction showed true and true value called False Positive and if prediction showed false and true values called False Negative (Flach, Peter, 2019).

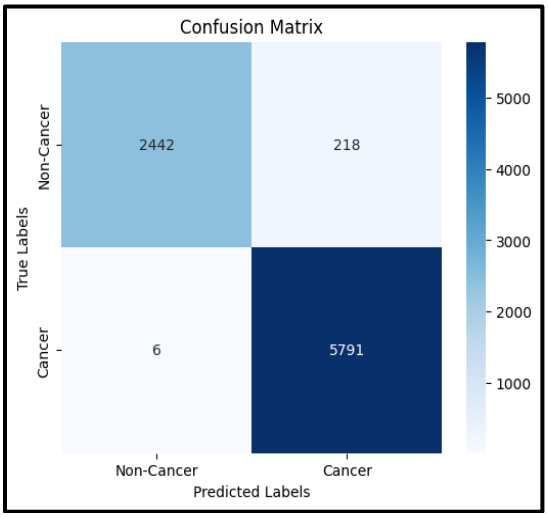

Figure 4.11 Confusion Matrix

Accuracy

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

Accuracy is defined as the number of correct predictions produced as a percentage of all predictions made. The Equation can be used to calculate accuracy (Hussain et al., 2022).

$$\textbf{Accuracy} = \frac{\textbf{correct classifications}}{\textbf{total classifications}} = \frac{\textbf{TP+TN}}{\textbf{TP+TN+FP+FN}} \qquad \textbf{14}$$

Precision

The total number of correct outputs returned by our ML model is known as precision. As shown in Equation, precision can be computed (Hussain et al., 2022)..

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{(\boldsymbol{TP+FP})} \qquad \textbf{15}$$

Recall

Our ML model's recall is the number of positives it returns. As shown in Equation, recall can be determined (Hussain et al., 2022).

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{(\boldsymbol{TP+FN})} \qquad \textbf{16}$$

F1 Score

The F1-score represents the harmonic mean of precision and recall. F1 score is calculated as the weighted average of precision and recall. F1's best and worst values are 1 and 0. The F1 score can be determined as shown in Equation (Hussain et al., 2022).

$$\boldsymbol{F-score} = 2\frac{(\boldsymbol{precision \times recall})}{(\boldsymbol{precision+recall})} \qquad \textbf{17}$$

Table 5.9

	Precision	**Recall**	**F1-Score**	**Support**
Non-Cancer	1.00	0.92	0.96	2660
Cancer	0.96	1.00	0.98	5797
Accuracy			**0.97**	**8457**
Macro Avg	0.98	0.96	0.97	8457
Weighted Avg	0.97	0.97	0.97	8457

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JOURNAL OF APPLIED LINGUISTICS AND TESOL (JALT)
Vol.9. No. 1 2026

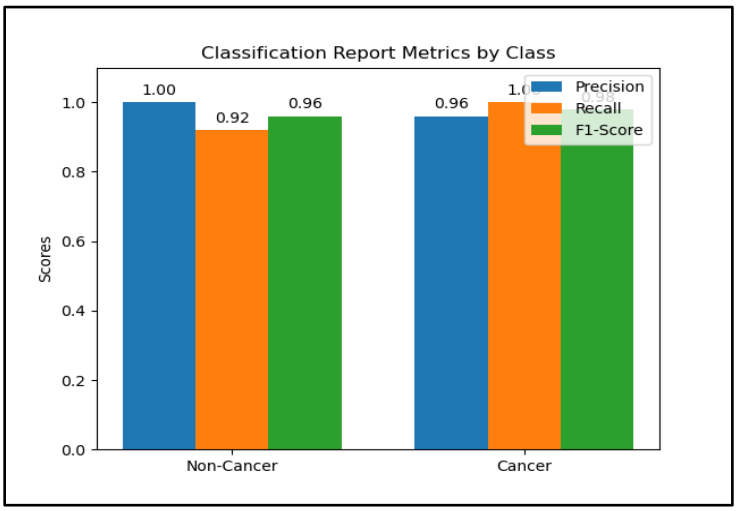

Figure 4.12 Bar Chart of classification report by classes

Discussion

This evaluated a CNN-based framework for classifying liver CT slices into cancerous and non-cancerous categories using baseline and transfer learning models. VGG16 achieved the highest validation accuracy (~97.6%), followed by the baseline CNN (~97.3%) and InceptionV3 (~97.1%), while ResNet50 showed lower but stable performance (~90.6%). Transfer learning reduced training time and overfitting, while the compact baseline CNN offered efficient and competitive performance. Future work should focus on patient-level analysis, external validation, and integration of volumetric and clinical data for real-world deployment.

Conclusion

This research developed a deep learning system that classifies liver CT slices into HCC and non-HCC with high accuracy (~97.35%) using a lightweight 9-layer CNN trained on over 36,000 labeled images. The model achieved excellent cancer detection performance while maintaining \fast speed and low computational cost, performing competitively with complex transfer learning models. Careful patient-level data splitting, preprocessing, and augmentation improved real-world reliability and generalization. Although currently limited to slice-level and single-center data, the system shows strong potential for real clinical deployment to support early liver cancer detection.

Limitations: The study used only binary classification (HCC vs. non-HCC), slice-level 2D analysis, and single-center data, which limits real-world clinical complexity, 3D tumor understanding, and generalization across hospitals.

Future Work: Future studies should use multi-center datasets, 3D volumetric deep learning models, and multi-class lesion classification to improve clinical reliability and diagnostic usefulness.

REFERENCES

Hussain, M., Saher, N., & Qadri, S. (2022). Computer Vision Approach for Liver Tumor Classification Using CT Dataset. *Applied Artificial Intelligence*, *36*(1). https://doi.org/10.1080/08839514.2022.2055395

Ioannou, G. N., Tang, W., Beste, L. A., Tincopa, M. A., Su, G. L., Van, T., Tapper, E. B., Singal, A. G., Zhu, J., & Waljee, A. K. (2020). Assessment of a Deep Learning Model to Predict Hepatocellular Carcinoma in Patients with Hepatitis C Cirrhosis. *JAMA Network Open*, *3*(9). https://doi.org/10.1001/jamanetworkopen.2020.15626

Saeed, F., Shiwlani, A., Umar, M., Jahangir, Z., Tahir, A., & Shiwlani, S. (2025). Hepatocellular Carcinoma Prediction in HCV Patients using Machine Learning and Deep Learning Techniques. *Jurnal Ilmiah Computer Science*, *3*(2), 120–134. https://doi.org/10.58602/jics.v3i2.48

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

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*. http://arxiv.org/abs/2501.08097

Shan, R., Pei, C., Fan, Q., Liu, J., Wang, D., Yang, S., & Wang, X. (2025). Artificial intelligence-assisted platform performs high detection ability of hepatocellular carcinoma in CT images: an external clinical validation study. *BMC Cancer*, *25*(1), 154. https://doi.org/10.1186/s12885-025-13529-x

Xia, Y., Zhou, J., Xun, X., Zhang, J., Wei, T., Gao, R., Reddy, B., Liu, C., Kim, G., & Yu, Z. (2024). CT-based multimodal deep learning for non-invasive overall survival prediction in advanced hepatocellular carcinoma patients treated with immunotherapy. *Insights into Imaging*, *15*(1). https://doi.org/10.1186/s13244-024-01784-8

Yu, P. L. H., Chiu, K. W. H., Lu, J., Lui, G. C. S., Zhou, J., Cheng, H. M., Mao, X., Wu, J., Shen, X. P., Kwok, K. M., Kan, W. K., Ho, Y. C., Chan, H. T., Xiao, P., Mak, L. Y., Tsui, V. W. M., Hui, C., Lam, P. M., Deng, Z., … Seto, W. K. (2025). Application of a deep learning algorithm for the diagnosis of HCC. *JHEP Reports*, *7*(1). *https://doi.org/10.1016/j.jhepr.2024.101219*

Azer, S. A. (2019). Deep learning with convolutional neural networks for identification of liver masses and hepatocellular carcinoma: A systematic review. *World Journal of Gastrointestinal Oncology*, *11*(12), 1218–1230. https://doi.org/10.4251/wjgo.v11.i12.1218

Dong, X., Zhou, Y., Wang, L., Peng, J., Lou, Y., & Fan, Y. (2020). Liver Cancer Detection Using Hybridized Fully Convolutional Neural Network Based on Deep Learning Framework. *IEEE Access*, *8*, 129889–129898. https://doi.org/10.1109/ACCESS.2020.3006362

Fenech, E., Briffa, J. A., & David SCICLUNA, I. (2025). *Analysis and optimisation of deep-learning liver segmentation techniques on abdominal computed tomography scans*. https://www.um.edu.mt/library/oar/handle/123456789/131282

Galicia-Moreno, M., Silva-Gomez, J. A., Lucano-Landeros, S., Santos, A., Monroy-Ramirez, H. C., & Armendariz-Borunda, J. (2021). Liver Cancer: Therapeutic Challenges and the Importance of Experimental Models. *Canadian Journal of Gastroenterology and Hepatology*, *2021*(1), 8837811. https://doi.org/10.1155/2021/8837811

Hussain, M., Saher, N., & Qadri, S. (2022). Computer Vision Approach for Liver Tumor Classification Using CT Dataset. *Applied Artificial Intelligence*, *36*(1). https://doi.org/10.1080/08839514.2022.2055395

Khan, R. A., Luo, Y., & Wu, F. X. (2023). Multi-level GAN based enhanced CT scans for liver cancer diagnosis. *Biomedical Signal Processing and Control*, *81*(February 2022). https://doi.org/10.1016/j.bspc.2022.104450

Malik, M. H., Ghous, H., Rashid, T., Maryum, B., Hao, Z., & Umer, Q. (2024). Feature extraction-based liver tumor classification using Machine Learning and Deep Learning methods of computed tomography images. *Cogent Engineering*, *11*(1). https://doi.org/10.1080/23311916.2024.2338994;PAGE:STRING:ARTICLE/CHAPTER

Mizouri, N. (2022). *Deep learning neural network with transferlearning for liver cancer classi cation*. 0–18.

Naga Srinivasu, P., Jaya Lakshmi, G., Canavoy Narahari, S., Shafi, J., Choi, J., & Fazal Ijaz, M. (2024). Enhancing medical image classification via federated learning and pre-trained model. *Egyptian Informatics Journal*, *27*, 100530. https://doi.org/10.1016/J.EIJ.2024.100530

N, Y. G., & R V, M. (2025). Automatic liver tumor classification using UNet70 a deep learning model. *Journal of Liver Transplantation*, *18*, 100260. https://doi.org/10.1016/J.LIVER.2025.100260

Okimoto, N., Yasaka, K., Kaiume, M., Kanemaru, N., Suzuki, Y., & Abe, O. (2023). Improving detection performance of hepatocellular carcinoma and interobserver agreement for liver imaging reporting and data system on CT using deep learning reconstruction. *Abdominal Radiology*, *48*(4), 1280–1289. https://doi.org/10.1007/s00261-023-03834-z

Rahman, H., Ben Aoun, N., Bukht, T. F. N., Ahmad, S., Tadeusiewicz, R., Pławiak, P., & Hammad, M. (2025). Automatic liver tumor segmentation of CT and MRI volumes using ensemble ResUNet-InceptionV4 model. *Information Sciences*, *704*, 121966. https://doi.org/10.1016/J.INS.2025.121966

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), 1–23. https://doi.org/10.3390/biomedicines11030800