Received: January 15, 2026. Revised: February 13, 2026.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Automated Multi-class Liver Cancer Classification from MRI Using a Cross-gated NASNet–vision Transformer Architecture

Jisha V[1]* **P. Sujatha[2]**

[1]*Department of Computer Science, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai, India*
[2]*Department of Computer Applications, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai, India*
* Corresponding author's Email: jishavasu@gmail.com

Abstract: Liver cancer is a predominant cause of cancer-related death globally. Timely and accurate diagnosis is required for improving survival and treatment results. Magnetic resonance imaging (MRI) is pivotal in the identification and clinical evaluation of liver cancer. Reliable liver cancer classification is challenging due to heterogeneous tumor morphology, subtle inter-class variations among different cancer subtypes, and low contrast between malignant, benign, and healthy liver tissues. Conventional diagnostic approaches rely heavily on handcrafted radiomic features and radiologist-driven interpretation, which are limited by inter-observer variability, imaging complexity, and scalability in large-scale screening scenarios. This study presents a hybrid Neural Architecture Search Network (NASNet) –Vision Transformer (ViT) model with Cross-Gating Fusion for automated multi-class liver cancer classification from MRI images. The NASNet component extracts discriminative local spatial and textural features, while the vision transformer modules capture long-range global contextual and inter-channel dependencies. The cross-gating fusion mechanism adaptively integrates complementary representations, enhancing feature discrimination and classification robustness. The proposed model was evaluated on a liver MRI dataset comprising 11,002 images across five classes. Experimental results demonstrate excellent performance, achieving an overall classification accuracy of 98.54%, along with good precision, F1-score and recall values across all liver cancer categories. The model exhibits stable convergence, robust generalization capability, and minimal misclassification, outperforming recent CNN-based and radiomics-driven approaches. These findings underscore the efficacy of integrating convolutional and transformer-based representations via cross-gated fusion, providing a robust and scalable solution for clinical decision support systems and MRI-based liver cancer diagnosis.

Keywords: Magnetic resonance imaging, Liver cancer, Deep learning, Neural architecture search network, Convolutional neural network, Vision transformer, Cross-gating fusion.

1. Introduction

The liver is a vital, essential organ located immediately below the ribs in the upper right quadrant of the abdomen. It is one of the body's most essential organs and is essential to preserve general health. The liver performs metabolism of nutrients, detoxification of harmful substances, bile production for fat digestion, production of clotting factors and blood proteins, storage of vitamins and minerals, regulation of blood sugar and cholesterol levels, and support of immune defense. Liver diseases are caused by viral illnesses including hepatitis B and C, excessive alcohol usage, fatty liver associated with obesity and diabetes, long-term drug or toxin exposure, autoimmune disorders, genetic conditions, metabolic abnormalities, bile duct disorders, and liver tumors [1]. These diseases affect the liver by damaging liver cells, blocking bile flow, reducing blood supply, causing inflammation and scarring, which disrupts its ability to detoxify blood, metabolize nutrients, produce bile and proteins, regulate glucose and cholesterol, and maintain normal immune and clotting functions. Among these

conditions, liver cancer represents the most prevalent malignant disease. Liver cancer is distinguished by the uncontrolled spread of aberrant cells in the liver, forming malignant tumors that interfere with normal liver function, most commonly arising as hepatocellular carcinoma (HCC) from liver cells.

Early identification of liver cancer is crucial because the disease is frequently asymptomatic in its initial stages and is typically detected at an advanced stage with few treatment choices [2]. Early detection of liver cancer enables possibly curative interventions such as liver transplantation, surgical removal or local ablative therapy. Early-stage tumors are typically smaller, localized, and less likely to exhibit vascular invasion or distant metastasis, resulting in improved survival outcomes. Better liver function preservation is also made possible by prompt diagnosis, especially in patients with cirrhosis or underlying chronic liver disease. In addition, early detection facilitates accurate staging and individualized treatment planning. Overall, early diagnosis significantly reduces disease-related morbidity and mortality [3]. Traditional approaches for detecting and evaluating liver cancer include clinical assessment and blood tests, imaging techniques like ultrasound, contrast-enhanced CT and MRI to visualize liver lesions, and histopathological examination through biopsy to confirm malignancy and assess tumor grade and stage. Despite these advantages, several limitations remain, including difficulty in identifying cancer at an early stage, dependence on operator expertise, restricted precision in differentiating benign from cancerous tumors, the invasive nature of biopsy procedures, variability in interpretation among clinicians, and an inability to fully assess tumor complexity and prognosis.

Advancements in deep learning (DL) and artificial intelligence (AI) methodologies significantly enhance the diagnosis and analysis of liver cancer, overcoming several limitations of traditional approaches. Artificial intelligence improves the precision and effectiveness of liver cancer evaluation by facilitating automated analysis of medical imaging data. Machine learning techniques help by analysing images and data in a more consistent and objective manner [4]. These methods assist in classifying liver lesions as benign or malignant and support predictive modelling to estimate disease progression and treatment response. However, this approach often relies on handcrafted features and require large, well-labelled datasets and poor-quality data can lead to inaccurate predictions. In addition, many machine learning models function as black boxes, making clinical interpretation challenging. Deep learning further improves liver cancer evaluation by automatically learning intricate patterns from MRI, CT and ultrasound images without human feature design [5, 6]. They also support accurate tumor segmentation, enabling precise assessment of tumor size and extent. Predictive modeling supports treatment planning and prognostic evaluation. Diagnostic errors related to human fatigue and cognitive bias are reduced. Overall, artificial intelligence and deep learning significantly improve the reliability and clinical utility of liver cancer evaluation. The key objectives of the current study are outlined below:

- To create a hybrid NASNet–ViT model with cross-gating fusion that effectively integrates convolutional feature extraction with transformer-based global contextual modelling for multi-class liver MRI classification.
- To boost the accuracy and robustness of liver carcinoma detection by collecting both fine-grained local spatial features and long-range inter-channel dependencies across different liver tumor types.
- To comprehensively evaluate the developed model on a multi-class liver MRI dataset using standard performance metrics in order to demonstrate its suitability for clinical decision support applications.

The remaining portions of this work are organized in the following sequence: Recent studies on the categorization of liver cancer using deep learning and medical imaging methods are reviewed in Section 2. The design and methodological specifics of the created hybrid model are described in Section 3. Section 4 presents the outcomes of the experiments and performance evaluation. Ultimately, Section 5 concludes the investigation and suggests possible lines of inquiry for further research.

2. Related works

Gowda N et al. [7] created a study based on deep learning on liver disease diagnosis using CT images, addressing challenges in radiological assessment of tumor type, size, and severity. CT scans were categorized into liver and liver carcinoma categories, including metastatic and hepatocellular carcinoma, utilizing a UNet70 architecture. The study used 1520 training images and 240 testing images from the 3Dircadbs dataset, with confusion matrix–based evaluation. The accuracy was 94.6%, the sensitivity was 97.5% and the dice score was 94.73%, outperforming several comparative classifiers. Limitations included binary class formulation,

limited sample diversity, and exclusive use of CT imaging.

Song et al. [8] established a deep learning approach rooted on intratumoral heterogeneity for MRI based preoperative histopathologic grade prediction in hepatocellular carcinoma. 858 people from two external and one primary cohort were included in the study, utilizing axial portal venous phase images acquired at 1.5T and 3.0T. K-means clustering was used to group voxel-level radiomics features to produce three subregions of the tumor, followed by feature extraction through 2.5D and 3D ResNet-based models. Selected features were integrated using a Random Forest classifier. The Intratumoral heterogeneity-based 2.5D feature-fusion model achieved improved discrimination compared to the whole-tumor approach, with AUC values of 0.82 versus 0.72, and maintained AUCs between 0.78 and 0.83 on validation cohorts. Limitations included retrospective design, voxel resampling affecting spatial detail, and reliance on a single MRI phase.

Stollmayer et al. [9] presented an approach based on deep learning for the automation of hepatocellular carcinoma risk assessment using gadoxetate disodium–enhanced MRI in accordance with LI-RADS v2018. 602 at-risk patients were included in this single-center retrospective analysis, and their manual semantic segmentations of LR-3 to LR-5 and LR-M lesions serving as reference. A collection of U-Net architectures with 14 source paths were trained using the nnU-Net platform and post-processing was then applied for lesion detection, classification, and instance segmentation. In the training, internal and external test cohorts, LR-5 lesions ≥10 mm had sensitivities between 0.83 and 0.90 and favorable predictive values between 0.88 and 0.94. Dice coefficients ranged from 0.52 to 0.84. Limitations involved single-reader annotation, partial LI-RADS coverage, absence of histopathologic correlation, limited etiologic analysis, and reliance on a single contrast agent.

Jia et al. [10] presented an MRI based deep learning technique that can predict hepatocellular cancer with a P53 mutation. 312 pathologically verified patients who had gadolinium-enhanced MRI were included in this retrospective analysis. An 8:2 ratio was used to divide them into training and testing groups. Images of the portal venous phase, arterial phase, T2-weighted and hepatobiliary phase were all processed using the EfficientNetV2 architecture, followed by construction of single-sequence and multiphase combined models. Among single-sequence approaches, the hepatobiliary phase had the best discrimination with an AUC of 0.715. The combined T2-weighted, arterial, and portal venous phase model achieved better results on the testing dataset, with an AUC of 0.914. Limitations included retrospective design, limited sample size, predominance of hepatitis B–related disease, restricted sequence selection, and absence of multicenter validation.

Wei et al. [11] created an automated three-dimensional tumor volume derived from MRI with deep learning assistance to forecast postoperative early recurrence in hepatocellular cancer. This one-center retrospective research included 592 cases with stage A and B illness of Barcelona Clinic Liver Cancer (BCLC) who had surgical resection and preoperative contrast-enhanced MRI. Automated segmentation was applied to obtain total tumor volume and total tumor burden, with radiologist review for quality control. Cox regression analysis was performed to assess prognostic relevance. Results showed higher 2-year early recurrence rates in BCLC B patients compared to BCLC A patients (45.3% vs. 30.0%; HR = 1.8). Total tumor burden emerged as the effective predictor (HR = 2.2), enabling risk stratification using a 6.84% threshold. Limitations included retrospective single-center architecture, lack of external validation, limited BCLC B cases, inclusion of only surgically treated patients, and partial radiologist involvement in segmentation.

Slama et al. [12] established a study using deep learning to automatically recognize and categorize liver tumor using CT images. A hybrid framework that used V-Net for liver and tumor segmentation and VGG16 for classification was applied to an improved CT dataset that had been divided into training and testing subsets. Segmentation improved classification reliability by providing focused tumor regions. The model's tumor segmentation Dice score was 97.34%, and its classification accuracy was 96.52% outperforming classification performed on unsegmented images, which yielded 89.34% accuracy. Comparative analysis indicated competitive performance with existing CT-based studies. Limitations included a relatively small dataset, constrained generalization capability, reliance on a simple hybrid architecture, and evaluation restricted to CT images without multi-organ or multi-disease validation.

Yin et al. [13] created a deep learning radiomics study using CT to predict response to treatment and survival without progression in patients with incurable liver cancer undergoing hepatic arterial infusion chemotherapy in conjunction with tyrosine kinase inhibitors and PD-1 along with trans arterial chemoembolization. There were 172 patients from

two centers in the retrospective analysis, divided into training, testing, and external validation cohorts. Pre-treatment contrast-enhanced CT images were analyzed using a ResNet-based feature extraction framework combined with clinical variables. Predictive models were evaluated for treatment response and survival outcomes. The AUC scores for the response prediction model were 0.96 in training, 0.87 in testing, and 0.85 in external validation, with accuracies ranging from 79.1% to 89.5%. Progression-free survival prediction demonstrated AUCs between 0.76 and 0.87 across cohorts, and the combined model achieved a C-index of 0.75. The single-center design, small sample size, brief follow-up period, and fluctuating treatment frequency were among the limitations.

Li et al. [14] introduced a work that used gadoxetic acid-enhanced MRI and supervised deep learning to help diagnose hepatocellular cancer. This retrospective analysis included 839 patients with 1023 focal liver lesions confirmed by pathology or imaging follow-up. Multiphase MRI sequences were manually annotated with bounding boxes and used to train a convolutional neural network classifier to differentiate Hepato Cellular Carcinoma (HCC) from non-HCC lesions, while a post hoc feature inference model identified LI-RADS imaging features. Model outputs were compared with radiologist assessments. The AI system's sensitivity and specificity were 90.6% and 90.7%, respectively exceeding LI-RADS category 5 sensitivity. Radiologist performance improved with AI assistance, achieving sensitivities of 85.7% and 89.1% without loss of specificity. Limitations included retrospective design, selective second-read assessment, exclusion of transitional phase imaging, restriction to high-risk patients, contrast-related variability, and limited population diversity.

He et al. [15] created a supervised deep learning system that predicts microvascular invasion in hepatocellular carcinoma in three classes before surgery using clinical biomarkers and MRI radiomics. This retrospective study included 437 pathologically confirmed patients from two campuses of a single institution, divided into training, internal testing, and external validation cohorts. A two-stage feature selection technique was used to extract radiomic features from Gadolinium benzyl oxy propionic tetra acetate (Gd-BOPTA)-enhanced MRI and combine them with laboratory markers. A Transformer-based architecture was then used for classification. The model's corresponding macro-average AUC values were 0.880 and 0.886, while the internal test cohort's accuracy was 0.733 whereas the external cohort's was 0.758. Sensitivity for high-risk M2 cases reached 0.73 with specificity of 0.91. Limitations included retrospective design, single-institution data sources, manual tumor segmentation with potential observer bias, and restricted population diversity.

Wu et al. [16] conducted a multicenter retrospective study using radiomics, deep transfer learning, and multimodal fusion learning for distinguishing hepatocellular dual-phenotypecarcinoma, hepatocellular carcinoma, and intrahepatic cholangiocarcinoma from contrast-enhanced MRI. The analysis included 381 patients from four centers, divided into training, internal testing, and external testing cohorts. Quantitative radiomic features and deep features extracted using VGG19 were analyzed individually and in combination through supervised classification frameworks. Model evaluation relied on confusion matrices and ROC analysis. Radiomics achieved macro-AUC values above 0.90 with accuracy and F1-scores exceeding 0.75 across validation cohorts. Deep transfer learning models showed improved discrimination, while fusion learning achieved the most stable performance, with macro-AUC values below 0.98 and accuracy above 0.85 in external testing. Limitations included retrospective design and limited biological interpretability of combined features.

Jeong et al. [17] conducted a retrospective study utilizing supervised deep learning-based analysis and logistic regression to predict post-hepatectomy liver complications in hepatocellular cancer.1760 individuals from two referral hospitals who had gadoxetic acid-enhanced hepatobiliary MRI prior to surgery were included in the cohort. Deep learning algorithms that evaluated the volumes and signal strengths of the liver and spleen were combined with clinical considerations to construct complete-liver and remaining-liver nomograms. Liver malfunctioning associated with hepatectomy happened to 7.8% of patients. The optimism-corrected AUC values for the whole-liver and remnant-liver models were 0.78 and 0.81 for total liver failure, and 0.81 and 0.84 for cases with symptoms. Limitations included retrospective design, lack of external validation, potential selection bias, and restricted generalizability due to data from two centers within a single country.

Zhang et al. [18] performed a multicenter retrospective analysis using deep learning and diffusion-based generative learning to enable gadolinium-free abbreviated MRI for hepatocellular carcinoma detection. A total of 1,769 patients from three institutions were split up into testing, training and external validation cohorts. From non-contrast sequences, stable diffusion models produced artificial

contrast-enhanced MRI phases, which were combined to form deep learning–based abbreviated MRI protocols and compared with conventional MRI. Image quality was assessed using non-inferiority testing, and diagnostic evaluation relied on sensitivity, specificity, and inter-method agreement. Abbreviated MRI reduced acquisition time from 28.1 to 4.1 minutes. Image quality was non-inferior, with excellent agreement in tumor size measurement (R2 > 0.94). The values that correspond to sensitivity and specificity are 0.866 and 0. 922.Limitations included retrospective design, limited external datasets, cohort imbalance, lack of ultrasound comparison, and restricted model comparisons.

Gu et al. [19] conducted a retrospective study using deep learning–based multilayer perceptron learning combined with radiomics and clinicoradiological analysis to forecast which vessels in hepatocellular carcinoma will enclose tumor clusters. Gadobenate-enhanced MRI prior to surgery, from 230 patients across three hospitals was divided into training, testing, and validation cohorts. Clinicoradiological variables were identified through regression analysis, while radiomics features were extracted from intratumoral and peritumoral regions and selected using LASSO-based filtering. Multiple machine learning classifiers were evaluated, and an optimal radiomics signature was integrated with clinical predictors to form a fusion model. The final model demonstrated improved discrimination compared with clinicoradiological and radiomics models alone, with AUC values exceeding 0.80 across datasets. Limitations included retrospective design, absence of survival analysis, lack of prospective validation, and relatively low sensitivity of the radiology-based multilayer perceptron model.

Dai et al. [20] introduced a supervised deep learning study that classified liver pictures for the early identification of hepatocellular cancer using convolutional neural network learning. The LiverCompactNet framework classified images into normal, benign and malignant categories using 5,000 liver images divided into training, test and validation sets. Image preprocessing included normalization, class balancing, and exploratory principal component analysis on derived tabular features. Model training demonstrated stable convergence across epochs, with progressive improvement in classification accuracy and reduced loss. The model achieved high diagnostic performance, with accuracy, sensitivity, specificity, and precision values remaining below 98%, alongside an AUC-ROC close to 0.99. Despite effective discrimination, limitations included reliance on a controlled dataset, limited generalizability, labor-intensive data annotation, and decreased interpretability of model since deep learning systems are black-boxed.

Yashaswini et al. [21] presented a deep learning project focusing on semantic liver segmentation and tumors from CT scans using fully convolutional neural networks. UNet and a modified Residual UNet architecture were applied to the 3Dircadb dataset to perform pixel-wise liver and tumor segmentation while preserving spatial features. The Residual UNet achieved dice similarity coefficient scores of 75.84% for tumor segmentation and 91.44% for liver segmentation on 128 × 128 images. Comparative discussion included cascaded UNet frameworks and squeeze-and-excitation–based 3D networks that enhanced feature calibration and segmentation accuracy across benchmark datasets. The study demonstrated reliable tumor delineation despite challenges posed by variable tumor shapes and sizes. A key limitation involved dependence on manually annotated datasets, which increased time and expert effort and restricted large-scale dataset availability.

2.1 Research gap

Despite significant progress, several research gaps persist across existing deep learning–based liver cancer studies. Most works rely on retrospective datasets with limited population diversity, restricting model generalizability [8–11, 14–18]. Several studies focus on binary or narrowly defined classification tasks, such as HCC vs. non-HCC, without addressing comprehensive multi-class tumor differentiation or disease spectrum modeling [7, 12, 14]. Manual or semi-automatic segmentation remains common, introducing observer bias and scalability issues [9, 11, 15, 21]. Additionally, limited prospective validation, lack of standardized benchmarks, and insufficient interpretability of deep models hinder clinical translation [13, 16, 20]. These gaps underscore the need for unified, interpretable, multimodal, and externally validated learning frameworks with reduced annotation burden and enhanced clinical applicability.

3. Materials and methods

Accurate liver cancer classification is essential for early disease diagnosis, as timely and reliable detection supports clinical decision-making and improves diagnostic accuracy. The workflow of the developed hybrid NASNet-ViT with cross gating fusion model for liver cancer classification is illustrated in Fig. 1. The input liver MRI image is processed in two parallel branches: a pre-trained NASNet-Large CNN extracts local texture features,

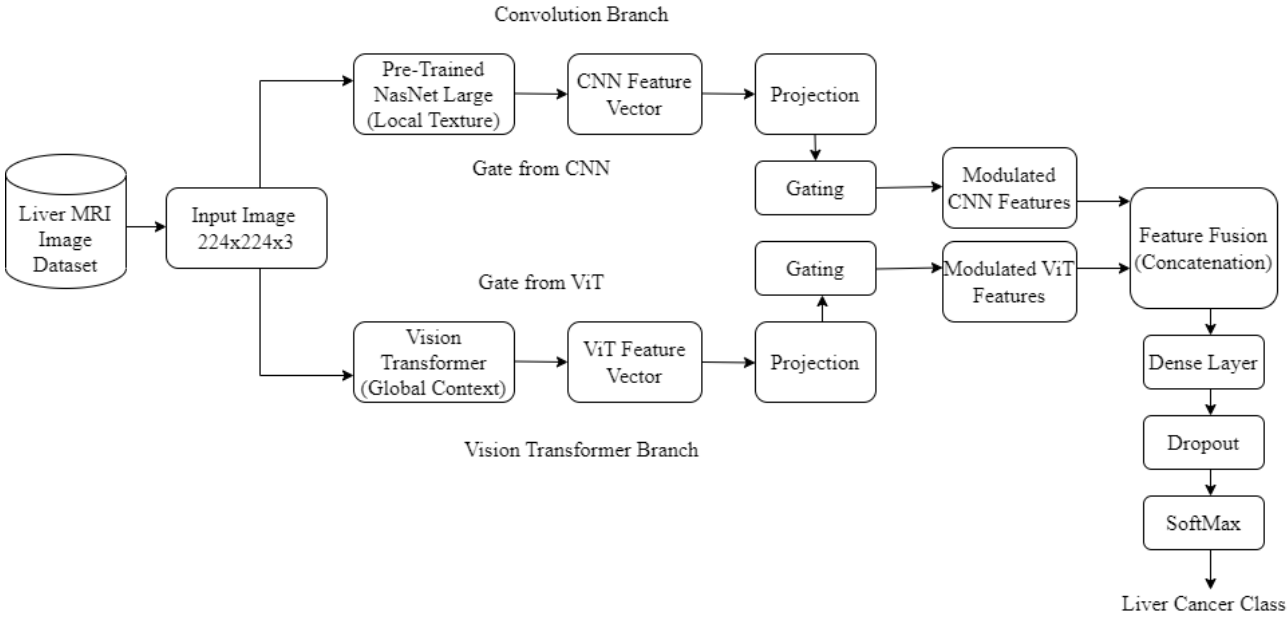

Figure. 1 Block schematic of the proposed model

while a Vision Transformer captures global contextual information. Each branch produces a feature vector that is projected to a common space and then passed through gating modules, where features from one branch modulate the other to emphasize complementary information. The gated CNN and ViT features are then concatenated and fused, followed by a dense layer and dropout for classification. A softmax layer ultimately produces the anticipated class for liver cancer.

3.1 Dataset description

The liver MRI imaging dataset utilized in this investigation comprises 11,002 MRI images obtained from a publicly available Kaggle repository [22]. Representative sample images from the five classes are shown in Fig. 2, illustrating the visual characteristics of healthy liver tissue and different liver tumor types observed in MRI scans. The dataset consists of liver MRI images distributed across five clinically relevant categories, namely Angiosarcoma, Cholangiocarcinoma, Hemangioma, Hepatocellular Carcinoma, and Healthy cases. The class-wise distribution includes 2376 Angiosarcoma, 2512 Cholangiocarcinoma, 2376 Hemangioma, 2328 Hepatocellular Carcinoma, and 2400 Healthy images, indicating a relatively balanced dataset that helps minimize class imbalance–induced bias during model training and evaluation, as depicted in Fig. 3. All images are provided in standard digital image formats and represent domain-specific medical visual data capturing liver anatomical and pathological characteristics from MRI scans. The dataset is released at the image level and does not include patient identifiers, subject-level annotations, or slice-to-subject association metadata. Consequently, subject-wise data partitioning could not be enforced, and each image was treated as an independent sample during model training and evaluation. For supervised multi-class learning, each category was assigned a unique numerical label corresponding to its diagnostic class, enabling effective learning of discriminative features for accurate liver cancer classification.

3.2 Data pre-processing and augmentation

Prior to model training, all liver MRI images are subjected to a series of preprocessing operations to ensure spatial uniformity, numerical stability, and effective feature learning. Each image is first resized to a fixed resolution of 224 × 224 × 3 to satisfy the input requirements of the NASNet and Vision Transformer backbones. Pixel intensity values are then normalized to the [0, 1] range using min–max normalization, defined as in Eq. (1).

$$I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}} \quad (1)$$

where I is the original pixel intensity value, $I_{\min}$ represents the minimum intensity value within the image, $I_{\max}$ represents the highest intensity value in the image and I_{norm} is the normalized pixel value.

Image resizing operation applied to the image is defined as Eq. (2).

Figure. 2 Sample dataset images: (a) Angiosarcoma, (b) Cholangiocarcinoma, (c) Hemangioma, (d) Hepatocellular Carcinoma (HCC), and (e) Healthy.

$$I'(x', y') = I(x, y) \quad (2)$$

where $I(x, y)$ denotes the intensity value associated with the original image at coordinates (x, y) and $I'(x', y')$ denotes the intensity value of the resized image at position (x', y').

Following preprocessing, the dataset is partitioned into training and testing subsets using a stratified 80:20 split, and identical preprocessing steps are applied across all folds in cross-validation experiments. To improve generalization performance and mitigate overfitting, data augmentation is applied exclusively to the training set.

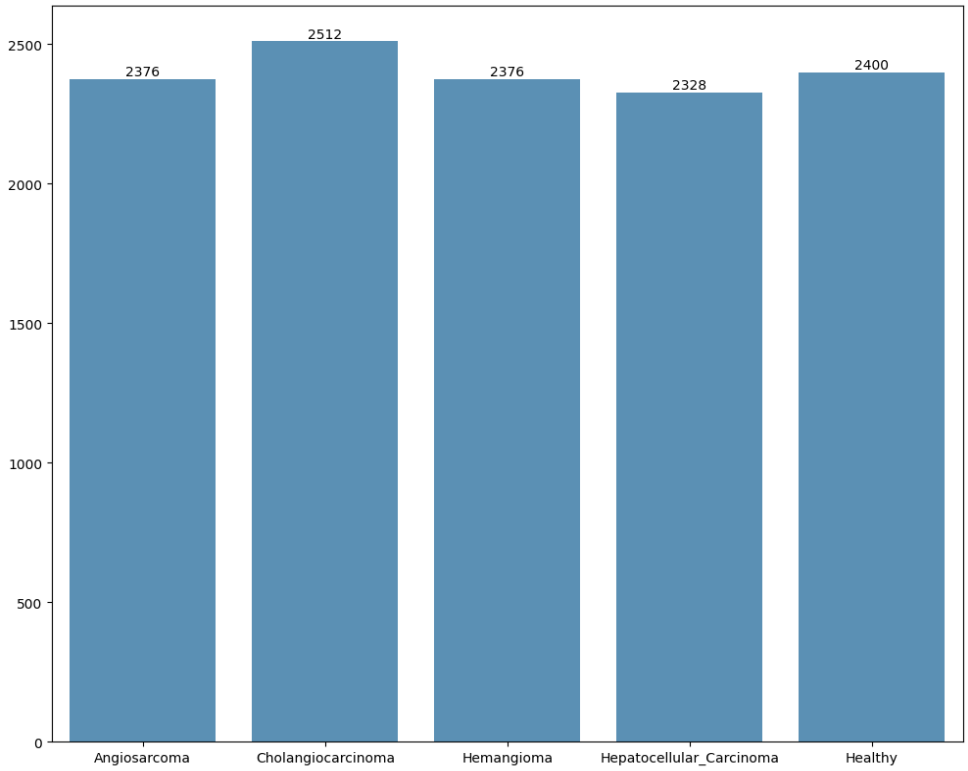

Figure. 3 Class Distribution

The augmentation strategy includes random rotations up to 30°, width and height shifts of 0.2, shear intensity of 0.15, zoom range of 0.15, and horizontal flipping, with the fill mode set to nearest. These preprocessing and augmentation operations enhance the robustness and discriminative quality of the input samples, enabling more effective learning of liver lesion characteristics by the proposed hybrid model.

3.3 Model development

The proposed model leverages NASNet to extract robust convolutional features and ViT to capture long-range global dependencies across image regions. A cross-gating mechanism is integrated to adaptively fuse complementary features, emphasizing clinically relevant regions while suppressing redundant information. Together, these components enhance representation learning, leading to improved accuracy and generalization in medical image analysis.

3.3.1. NasNet-Large network (neural architecture search network)

NASNet-Large is a convolutional neural network architecture developed using a scalable Neural Architecture Search (NAS) framework. In this approach, the Google ML research group employed a reinforcement learning–based strategy to automatically discover optimal network designs. Rather than manually designing deep architectures, NAS learns architectural patterns directly from data, enabling the construction of highly efficient and high-performing networks. The core idea behind NASNet-Large is to identify an optimal cell structure within a very limited dataset and then transfer this learned cell to build a deeper and wider network suitable for large-scale image recognition tasks such as ImageNet. This transferability significantly reduces the computational burden of architecture search while achieving state-of-the-art classification accuracy and strong generalization capability [23].

NASNet-Large is composed of two reusable building blocks known as the Normal Cell and the Reduction Cell [24]. These cells act as modular units that are stacked repeatedly to form the full network. The normal cell preserves the spatial resolution of feature maps and is primarily responsible for extracting rich features through combinations of standard convolution, depthwise separable convolution, pooling operations, and skip connections. The reduction cell, on the other hand, increases the number of channels while decreasing the feature maps' spatial dimensions, allowing for hierarchical abstraction and a larger receptive field. Fig. 4 shows the structural arrangements of the Normal Cell and Reduction Cell used in NASNet.

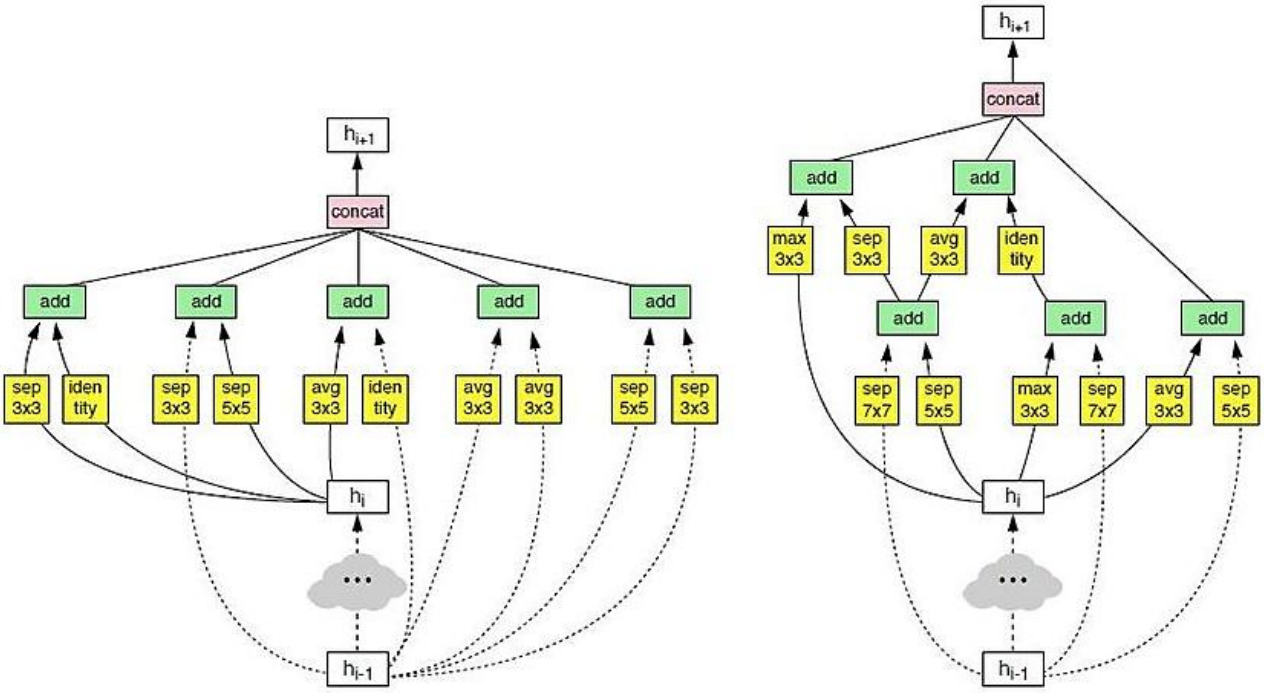

Figure. 4 Normal Cell and Reduction cell in NASNet

In NASNet-based image classification, the input image is mathematically represented to define its spatial and channel dimensions before being processed by the network. Eq. (3) defines the structure of the image before it is passed into the network.

$$X \in R^{H \times W \times 3} \tag{3}$$

where H and W indicate the illustration's height and width, respectively and 3 corresponds to RGB color channels. This representation serves as the starting point for feature extraction.

The network initiates with a convolutional stem that extracts low-level visual properties including edges and textures before transmitting the data into the NASNet cells. In convolutional neural networks, the convolution operation applies learnable filters across spatial locations to capture local patterns while sharing weights, thereby reducing the number of trainable parameters. The convolution operation's output is calculated as shown in Eq. (4).

$$z_{i,j,k} = \sum_{m=1}^{K} \sum_{n=1}^{K} \sum_{c=1}^{C} w_{m,n,c,k} \; x_{i+m-1,j+n-1,c} + b_k \tag{4}$$

where $w_{m,n,c,k}$ symbolizes the convolution kernel's weight, $x_{i,j,c}$ represents the input value at spatial location (i, j) in channel c and b_k denotes the bias term of the k th filter. This operation enables the network to learn discriminative spatial features that are propagated through successive NASNet cells.

To stabilize training and ensure consistent feature scaling, Layer Normalization (LN) is applied after convolution. LN normalizes activations within each sample by computing the mean and variance across feature dimensions. The mean μ and variance σ^2 are computed as described in Eqs. (5) and (6).

$$\mu = \frac{1}{d} \sum_{i=1}^{d} x_i \tag{5}$$

$$\sigma^2 = \frac{1}{d} \sum_{i=1}^{d} (x_i - \mu)^2 \tag{6}$$

where d represents the number of features.

The normalized activation is then obtained as given by Eq. (7).

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{7}$$

where ϵ is a minor constant incorporated for numerical stability, is used. This normalization is followed by a scale-and-shift transformation as shown by Eq. (8).

$$y = \gamma \hat{x} + \beta \tag{8}$$

where γ and β are learnable parameters that restore the network's representational capacity. The

Rectified Linear Unit (ReLU) activation function adds non-linearity, which enables the model to learn complex decision bounds.

At the end of the network, Global Average Pooling (GAP) is utilized to aggregate spatial features into compact global representations. GAP calculates each feature map's average value rather than depending on completely connected layers with a lot of parameters, thereby reducing model complexity and mitigating overfitting. For a feature map F of size $H \times W$, the GAP operation is defined as shown in Eq. (9).

$$GAP_c = \frac{1}{H*W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{i,j,c} \quad (9)$$

where GAP_c denotes the pooled output for channel c and $F_{i,j,c}$ represents the activation at spatial location (i, j). In order to produce class probabilities, the pooled feature vector is finally processed through a fully connected layer and then the SoftMax activation function. The SoftMax output for class k is computed as shown in Eq. (10).

$$y_k = \frac{e^{z_k}}{\sum_{i=1}^{K} e^{z_i}} \quad (10)$$

where z_k denotes the logit corresponding to class k and K denotes the aggregate number of classes.

3.3.2. Vision transformer

A deep learning model called Vision Transformer (ViT) makes use of the Transformer architecture for image analysis [25]. Similar to how words are processed in a sentence, ViT interprets an image as a series of patches instead of relying on convolutional layers. Unlike conventional Convolutional Neural Networks (CNNs), which emphasize local feature extraction, ViT predominantly utilizes the self-attention mechanism to detect connections throughout the whole image [26]. First, a fixed-size patch segmentation of an input image is performed. Patch embedding is the process of flattening and converting each patch into a numerical vector. Positional embeddings are incorporated to maintain spatial information. These embeddings are processed using feed-forward networks and a multihead self-attention module, which allows the model to learn correlations between various visual areas as well as global relationships throughout the entire image. Residual connections and layer normalization are applied to stabilize training, followed by dense layers for feature transformation. Finally, global average pooling summarizes the learned features into a compact representation for classification. The final output representation is used for image classification, detection, or segmentation. ViT is appropriate for complicated visual tasks because it excels at collecting global context and long-range interdependence. However, it typically requires large training datasets or pretraining to achieve significant performance compared to convolutional neural networks [27]. Illustration of Vision Transformer is shown in Fig. 5.

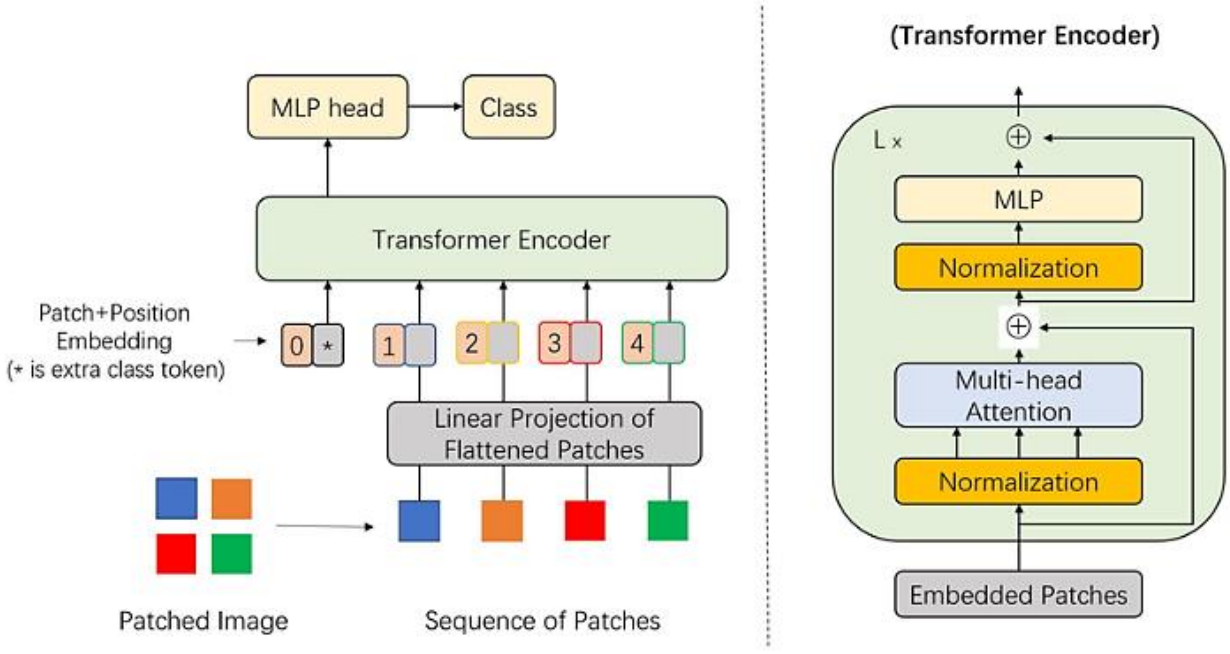

Figure. 5 Illustration of Vision Transformer

The ViT processes images by dividing them into non-overlapping, fixed-size patches that encompass the whole image [28]. For an input image of spatial resolution $H \times W$, the image is divided into $P \times P$ patches. Eq. (11) indicates how many patches are present overall.

$$N = \left(\frac{H}{P}\right) * \left(\frac{W}{P}\right) \tag{11}$$

where N is the total number of patches. The image is enlarged or padded so that both H and W are multiples of P, ensuring N is a positive integer. After being transformed into a one-dimensional vector, each patch is then transferred using a learnable linear transformation to a D-dimensional embedding space, as illustrated in Eq. (12).

$$z_i = W_p\, x_i + \, b_p, i = 1,2,3 \ldots . \mathrm{N} \tag{12}$$

where $x_i \in R^{P^2.C}$ is the flattened vector of the i-th image patch (with C color channels), b_p is a learnable bias and $W_p \in R^{D*(P^2.C)}$ is a trainable weight matrix [29]. Each patch embedding comprises of positional embeddings to maintain spatial information. The resulting sequence is expressed as shown in Eq. (13).

$$\hat{z}_i = z_i + \, E_i \; i = 1,2 \ldots, N \tag{13}$$

The multi-head self-attention (MSA) method enables global interactions among all patch embeddings, enabling contextual feature learning throughout the image. For a given input sequence $Z \in R^{(N+1)*D}$, the Query (Q), Value (V) and Key (K) matrices are obtained using different linear projections as seen in Eqs. (14)-(16).

$$Q = ZW_Q \tag{14}$$

$$K = ZW_K \tag{15}$$

$$V = ZW_V \tag{16}$$

where W_Q, W_K, $W_V \in R^{D*D_k}$ are projection matrices that can be trained and D_k is the key vectors' dimension.

Each transformer block uses a feed-forward network (FFN) with two linear layers separated by a Gaussian Error Linear Unit (GELU) activation to process its output following the attention operation, as indicated in Eq. (17).

$$FFN(x) = GELU(xW_1 + \, b_1)W_2 + b_2 \tag{17}$$

where W_1 and W_2 denote the learnable weight matrices, with b_1 and b_2 representing their corresponding bias terms [30].

To preserve steady gradient flow and lessen deterioration effects, each transformer block uses residual connections and layer normalization (LN) as shown in Eqs. (18) and (19).

$$\hat{z} = LN(z_{l-1}) + MSA(z_{l-1}) \tag{18}$$

$$\hat{z} = LN(z_l) + MSA(z_l) \tag{19}$$

where z_l is the final output of the transformer block at layer l.

After the class token has traversed several transformer layers, its ultimate representation is extracted and fed into a fully linked layer activated by softmax for classification which is represented by Eq. (20).

$$y = softmax(W_c\, z_L^0 + b_c) \tag{20}$$

where W_c and b_c are variable parameters, and z_L^0 is the final class token embedding. The output y represents the probability distribution over the target classes [31, 32].

3.3.3. Cross gating mechanism

A cross-gating mechanism is used to control and combine information between different branches of a neural network. It enables the model to exploit the complementary strengths of multiple feature representations. Through learned gating functions, one feature stream can either enhance or suppress the other, allowing selective information flow. This approach helps the network to concentrate on relevant features while reducing noise and is widely used in multi-modal, multi-scale, and encoder–decoder architectures to improve performance. In cross-gating, each branch generates a gate for the other branch as shown in Eqs. (21) and (22).

$$G_{CNN} = \sigma(W_1 F_{ViT} + b_1) \tag{21}$$

$$G_{ViT} = \sigma(W_2 F_{CNN} + b_2) \tag{22}$$

where W_1, W_2 denotes learnable weight matrices, b_1, b_2 denotes biases, $\sigma(\cdot)$ is sigmoid function and Gate values lie in [0, 1].

3.3.4. Proposed hybrid NASNet -ViT with cross gating fusion model

The developed model combines NASNet with ViT for effective liver MRI image classification. Each input MRI image of size 224×224×3 is simultaneously fed into two parallel feature extraction branches. The first branch employs a pretrained NASNet-Large convolutional neural network as a frozen backbone, consisting of a deep stack of normal and reduction cells with depthwise separable convolutions, enabling the extraction of high-level hierarchical features related to lesion texture, shape, and morphological patterns. The output feature maps from NASNet-Large are compressed by a global average pooling layer, yielding a 4032-dimensional compact representation. In parallel, the Vision Transformer branch converts the incoming image into discrete, non-overlapping patches measuring 16×16 using a convolutional embedding layer, producing 196 patch tokens with an embedding dimension of 64. To help the model identify long-term dependencies and global spatial context within the hepatic area, these tokens are processed through two stacked transformer encoder layers, each of which has four attention heads and multi-head self-attention. Subsequently, feed-forward networks and layer normalization are highlighted. The transformer output is aggregated using global average pooling to form a compact global feature vector. Using fully connected layers, both feature vectors are projected onto a common 256-dimensional embedding space to allow for efficient interaction between the heterogeneous CNN and transformer features. The core novelty of the architecture lies in the Cross-Gating Fusion Block, where a bidirectional gating mechanism is applied instead of direct feature concatenation. In this block, the transformer features generate a gating signal that selectively modulates the CNN features, while simultaneously, the CNN features generate a complementary gate to modulate the transformer features. This mutual gating mechanism allows each branch to emphasize task-relevant information while suppressing redundant or noisy activations based on cross-branch context. The gated feature vectors are then concatenated and refined using an additional dense layer to form a unified representation. Finally, the integrated characteristics are subsequently processed through a fully connected classification head with dropout regularization to reduce overfitting, followed by a softmax layer that outputs probability scores for five liver cancer classes. This architecture effectively combines strong local feature learning from NASNet-Large with global dependency modeling from ViT, while the cross-gating mechanism ensures adaptive and context-aware feature fusion for improved liver cancer classification performance. The algorithm for the proposed model is given below.

Algorithm 1: *Hybrid NASNet - ViT with Cross Gating Fusion Model for Liver MRI Classification*

Input: *Liver MRI images* $I \in \mathbb{R}^{224\times224\times3}$
Output: *Predicted class label* $\hat{y} \in \{1,2,3,4,5\}$

Start:
Step 1: *Load and Preprocess Data*
- *Load liver MRI images and corresponding labels for 5 classes.*
- *Resize each image to* $224 \times 224 \times 3$.
- *Normalize pixel intensities (min-max to [0,1]).*
- *Apply on-the-fly augmentation only on the training set (rotation, horizontal/vertical flip, mild zoom, and slight scaling).*
- *Divide the dataset into subsets for training and testing utilizing stratified sampling to maintain class distribution.*

Step 2: *Initialize NASNet–ViT Hybrid Architecture*
- *CNN Feature Extraction (NASNet branch):*
  - *Load pretrained NASNet backbone with classification head removed (include _top=False).*
  - *Feed each image through NASNet and extract the CNN feature representation.*
- *Transformer Feature Extraction (ViT branch):*
  - *Tokenize and embed the provided image into non-overlapping patches.*
  - *Transmit tokens through transformer encoders and extract the ViT feature representation.*

Step 3: *Project Features into a Common Embedding Space*
- *Project CNN features into embedding dimension D.*
- *Project ViT features into embedding dimension D.*

Step 4: *Cross-Gating Fusion Block*
- *Generate gate from ViT features to modulate CNN features.*
- *Generate gate from CNN features to modulate ViT features.*

- *Apply element-wise modulation to obtain gated feature outputs.*

Step 5*: Feature Fusion and Classification Head*

- *Concatenate gated CNN and ViT features.*
- *Apply dense layers with non-linear activation.*
- *Apply Dropout to reduce overfitting.*
- *Compute class probabilities using Softmax.*

Step 6: *Model Training*

- *Loss function: Categorical Cross-Entropy*
- *Optimizer: Adam with tuned learning rate and batch size.*
- *Train the network in mini-batches across epochs.*
- *Update parameters using backpropagation.*
- *Implement early termination depending on validation performance.*

Step 7: *Model Evaluation and Inference*

- *Preprocess test MRI image.*
- *Extract CNN and ViT features, perform cross-gating fusion, and compute Softmax outputs.*
- *Assign the final predicted class label*

Save Model

End

3.4 Simulation setup

All experiments were conducted using the Google Collaboratory platform, which provides cloud-based computational resources with GPU acceleration suitable for deep learning–based medical image analysis. The implementation was carried out using Python (version 3.9) with the TensorFlow deep learning framework (version 2.12) and the Keras API, enabling modular construction, training, and evaluation of the proposed hybrid architecture. To support reproducibility, random seeds were fixed consistently across Python, NumPy, and TensorFlow throughout all experiments. Model training was performed using the Adam optimizer with a learning rate of 1×10^{-5} and a batch size of 32. Training was conducted for a maximum of 100 epochs, with early stopping applied based on validation loss using a patience of 10 epochs to prevent overfitting. Dropout regularization was employed within the classification head to further enhance generalization. The hyperparameters used for training and evaluation of the proposed hybrid model are summarized in Table 1.

Table 1. Hyperparameters employed in the proposed model

Hyperparameter	Value
Epochs	100
Train: Test Ratio	80: 20
Patch Size	16 × 16
Learning Rate	0.00001
Optimizer	Adam
Activation Functions	ReLU, Softmax
Loss Function	Categorical Cross-Entropy
Dropout Rate	0.5

4. Results and discussions

4.1 Performance evaluation

The accuracy plot illustrates a consistent improvement in both training and validation accuracy across epochs, highlighting the model's excellent learning capacity as shown in Fig. 6. At the initial epoch, the training accuracy is approximately 0.39, while the validation accuracy begins at around 0.57, reflecting the model's initial learning stage with randomly initialized trainable parameters. By the second epoch, training accuracy improves to nearly 0.67 and validation accuracy rises to about 0.77, indicating rapid feature learning. As training progresses from epochs 3 to 5, training accuracy increases steadily from approximately 0.76 to 0.86, while validation accuracy improves from about 0.82 to 0.91, demonstrating effective convergence and balanced learning. Between epochs 6 and 10, both curves continue to rise gradually, reaching around 0.95 for training accuracy and 0.97 for validation accuracy, with minimal divergence. Toward the final epochs, the training accuracy converges close to 0.98, while validation accuracy approaches 0.99, indicating strong generalization capability. The model gains strong discriminative features without overfitting, as evidenced by the training and validation accuracy curves constant alignment.

The loss plot signifies a smooth and monotonic decrease in both training and validation loss values, reflecting stable optimization of the proposed hybrid model as shown in Fig. 7. Although the total number of epochs was initially established at 100, the model was automatically terminated at the 16th epoch through early stopping, indicating that further training would not yield meaningful improvement. At the initial epoch, the training loss is approximately 1.48, while the validation loss starts at around 1.30, indicating a high error rate due to untrained network parameters.

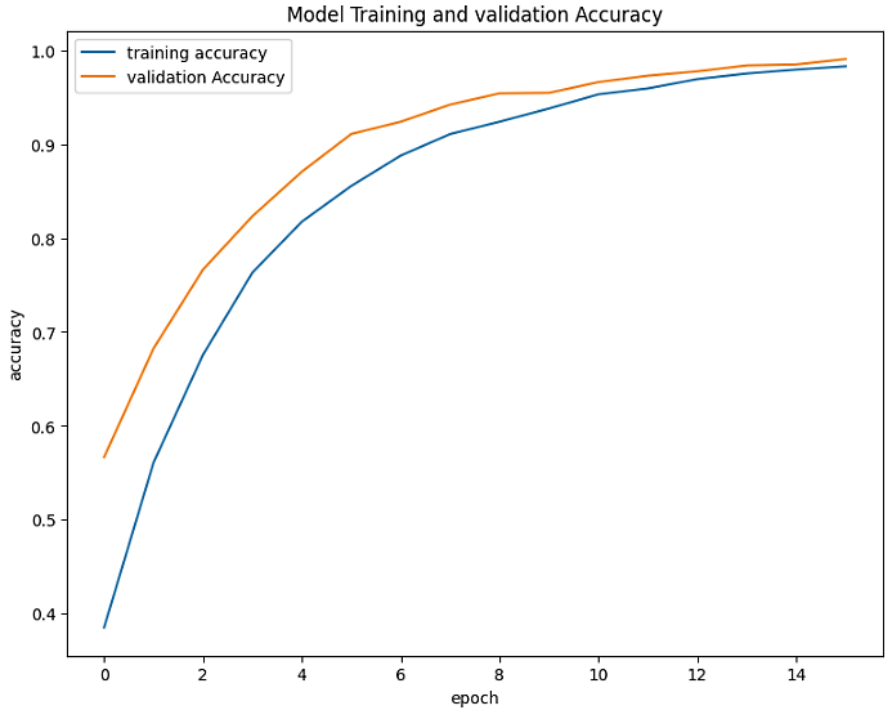

Figure. 6 Accuracy plot of the proposed model

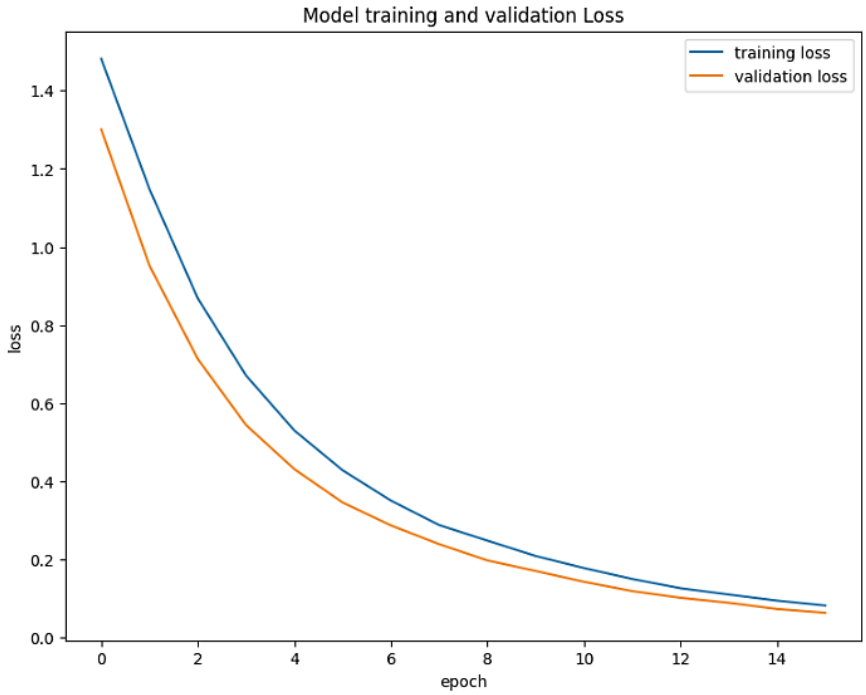

Figure. 7 Loss plot of the proposed model

By the second epoch, the training loss decreases sharply to nearly 0.87 and the validation loss reduces to about 0.72, demonstrating rapid convergence in the early training phase. From epochs 3 to 6, the loss curves decline steadily in a closely aligned manner, with training loss reducing from approximately 0.67 to 0.38 and validation loss from about 0.54 to 0.29, reflecting stable learning behavior. During the later epochs, the loss continues to decrease gradually, and by the final epoch, the training loss converges to nearly 0.09 while the validation loss stabilizes around 0.06. The consistently small gap between training and validation loss throughout the training process indicates effective generalization with minimal

overfitting. Overall, the smooth convergence of the loss curves validates the efficacy of the selected hyperparameters, optimization method and regularization mechanisms employed in the NASNet–ViT cross-gated hybrid architecture.

The efficacy of the suggested model was evaluated utilizing conventional classification metrics obtained from the confusion matrix, namely true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These measures offer a thorough and dependable evaluation of the model's predictive capability by quantifying its ability to correctly identify liver cancer cases as well as normal liver conditions from MRI images. As defined in Eq. (23)– (26), accuracy assesses the comprehensive correctness of classification, recall evaluates the model's efficacy in detecting true liver cancer cases, precision reflects the dependability of favorable predictions, and the F1-score offers a comprehensive evaluation by integrating precision and recall. Together, these evaluation metrics offer a rigorous and widely accepted framework for analysing classification performance, error characteristics, and the discriminative power of the proposed model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{23}$$

$$Recall = \frac{TP}{TP+FN} \tag{24}$$

$$F1-score = 2 * \frac{Precision \times Recall}{Precision+Recall} \tag{25}$$

$$Precision = \frac{TP}{TP+FP} \tag{26}$$

The achieved accuracy of 98.54% demonstrates the high reliability and robustness of the suggested model in accurately classifying liver carcinoma from MRI images. The detailed class-wise evaluation results demonstrate the resilience and dependability of the suggested model which is depicted in Fig. 8. Notably, the Healthy class attains perfect precision (1.00) and an F1-score of 1.00, indicating flawless identification of normal liver MRI scans. Hepatocellular Carcinoma also exhibits excellent performance achieving a recall of 1.00 and an F1-score of 0.99, underscoring the model's strong sensitivity toward malignant cases. The remaining tumor classes consistently achieve precision and recall values above 0.96, reflecting balanced and reliable classification behavior. The overall classification accuracy of 0.99 validates the efficacy of the suggested model on unseen test data.

	precision	recall	f1-score
Angiosarcoma	0.97	0.99	0.98
Cholangiocarcinoma	0.99	0.99	0.99
Healthy	1.00	0.99	1.00
Hemangioma	0.98	0.96	0.97
Hepatocellular_Carcinoma	0.99	1.00	0.99
accuracy			0.99
macro avg	0.99	0.99	0.99
weighted avg	0.99	0.99	0.99

Fig. 8 Classification Report

Furthermore, the macro-average and weighted-average precision, recall, and F1-score values of 0.99 indicate uniform performance across all classes, irrespective of class imbalance. These results confirm the suggested hybrid model's capacity to accurately discriminate between healthy liver tissue and multiple liver cancer types in MRI images.

Fig. 9 depicts the confusion matrix which demonstrates the grouping efficacy of the constructed model. Among 474 Angiosarcoma cases, 467 images were correctly classified, with 7 samples misidentified as Hemangioma. For Cholangiocarcinoma, 518 out of 522 images were accurately recognized, while 4 cases were misclassified as Healthy or Hepatocellular Carcinoma. In the Healthy category, 502 of 505 images were correctly labeled, with only 3 instances confused with Cholangiocarcinoma. The Hemangioma class achieved 438 correct predictions out of 457 cases, where 19 images were mainly misclassified as Angiosarcoma or Hepatocellular Carcinoma. Similarly, 439 out of 441 Hepatocellular Carcinoma images were accurately detected, with 2 misclassifications into the Cholangiocarcinoma class. Overall, the model produced 2364 correct predictions and 35 misclassifications. The low confusion among clinically similar liver tumor types highlights the efficacy of the proposed model in incorporating both local structural and global contextual features from MRI scans.

A threshold-independent assessment of classification performance is provided by the Receiver Operating Characteristic (ROC) curve, which shows the relationship between the False Positive Rate (FPR) and the True Positive Rate (TPR) at several decision levels. Fig. 10 illustrates that the multi-class ROC curves of the developed model are consistently positioned near the upper-left region of the ROC space, indicating excellent discriminative capability. The micro-averaged ROC curve achieves an AUC of 0.9997, reflecting robust overall classification performance across all liver classes.

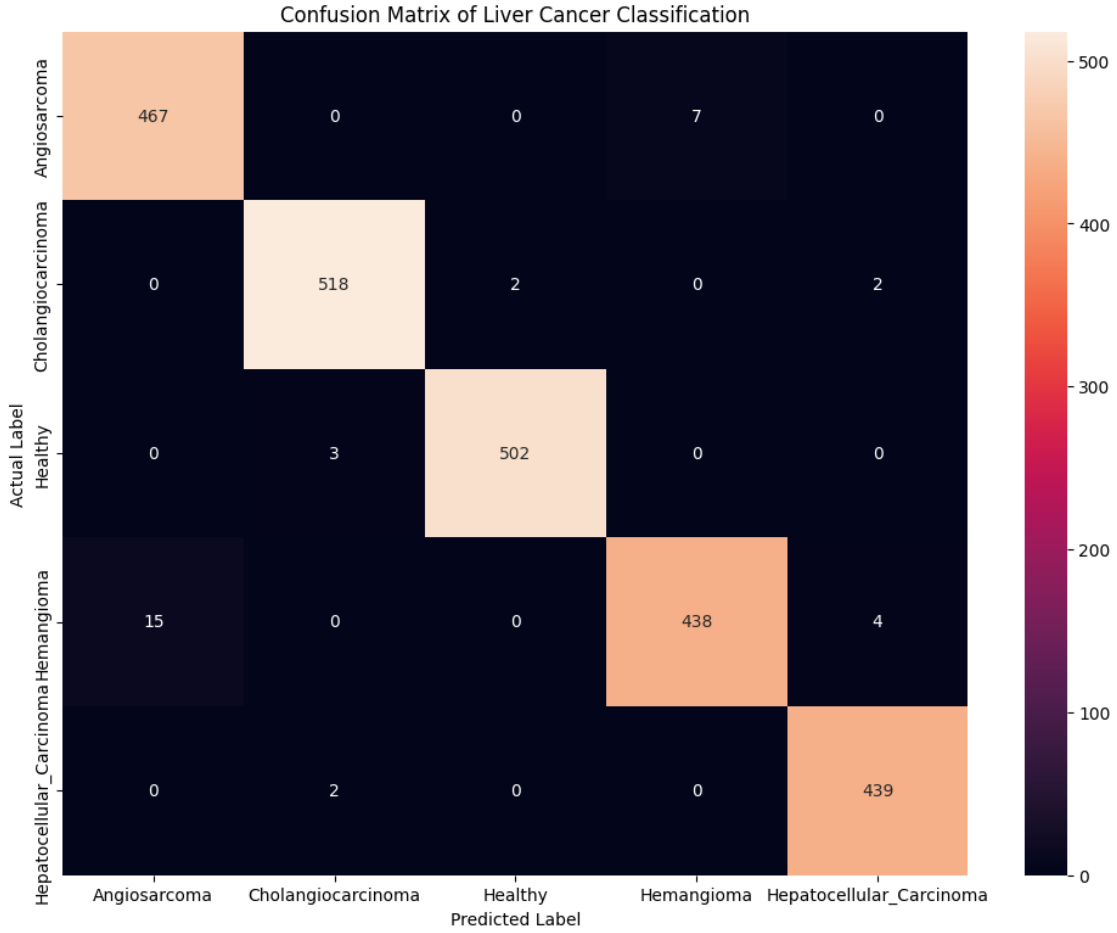

Figure. 9 Confusion Matrix

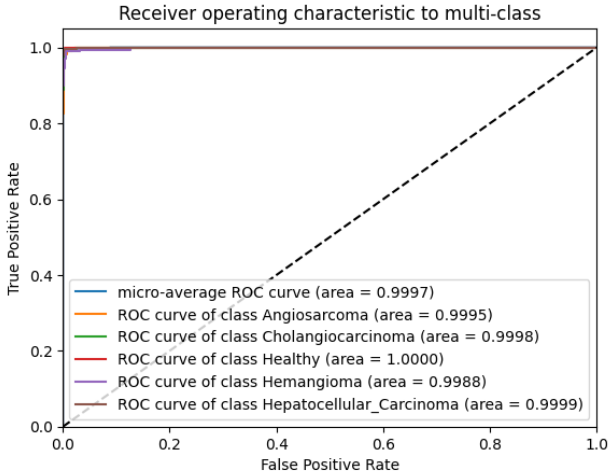

Figure. 10 ROC curve

In addition, class-wise ROC curves exhibit near-perfect separability, with AUC values of 0.9995 for Angiosarcoma, 0.9998 for Cholangiocarcinoma, 1.0000 for Healthy, 0.9988 for Hemangioma, and 0.9999 for Hepatocellular Carcinoma.

The clear separation from the random baseline and the rapid convergence toward a TPR of 1.0 at very low FPR values highlight the trustworthiness and sustainability of the suggested model. This result validates the model's efficacy for precise and reliable liver cancer diagnosis in MRI-based clinical decision support systems.

Fig. 11 illustrates representative prediction outcomes of the developed model across multiple clinical categories. The model accurately identifies Angiosarcoma, Cholangiocarcinoma, Hemangioma, Hepatocellular Carcinoma, and Healthy cases under diverse anatomical appearances and contrast variations. The close correspondence between true and predicted labels in the majority of samples visually corroborates the strong quantitative performance achieved by the model. These visual results demonstrate the robustness of the cross-gated NASNet–ViT architecture in capturing discriminative tumor-specific patterns and reliably differentiating between malignant, benign, and healthy liver tissues, highlighting its potential effectiveness for MRI-based liver cancer diagnosis.

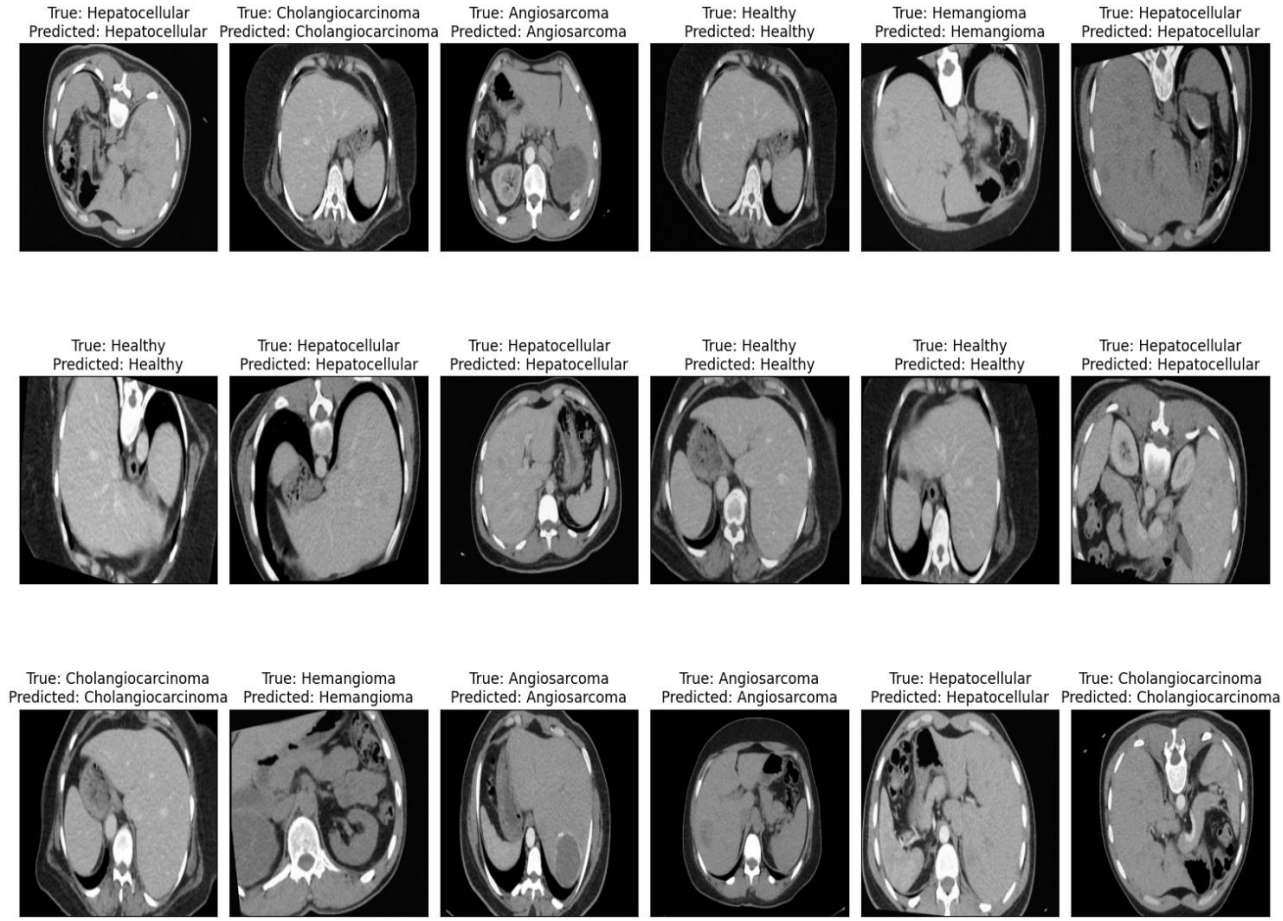

Figure. 11 Predicted outputs of the proposed model

4.2 Cross-validation analysis

To address concerns related to single-split evaluation and potential data leakage in image-level datasets, five-fold stratified cross-validation was conducted. As shown in Table 2, the proposed hybrid NASNet–ViT model demonstrates consistently strong performance across all folds, with classification accuracy ranging from 97.85% to 98.76%. The relatively small standard deviation (±0.39%) indicates stable image-level generalization under repeated data partitioning. Macro-averaged precision, recall, and F1-score values remain uniformly high across folds, suggesting balanced discrimination among all five liver classes. Furthermore, the macro-averaged AUC values consistently approach unity, confirming the robustness of the learned feature representations. These results validate that the reported performance is not dependent on a single random split and remains stable under repeated stratified evaluation, within the constraints of image-level dataset availability.

Table 2. Five-fold stratified cross-validation results of the proposed hybrid NASNet–ViT model

Fold	Accuracy (%)	Precision (macro)	Recall (macro)	F1-score (macro)	AUC (macro)
Fold 1	97.98	0.979	0.976	0.978	0.997
Fold 2	98.34	0.984	0.982	0.983	0.999
Fold 3	98.67	0.987	0.985	0.986	0.999
Fold 4	97.85	0.976	0.974	0.975	0.996
Fold 5	98.76	0.988	0.986	0.987	0.999
Mean ± Std	98.32 ± 0.39	0.983 ± 0.005	0.981 ± 0.005	0.982 ± 0.005	0.998 ± 0.001

Table 3. Ablation study of model components

Model Variant	Accuracy (%)	Precision (macro)	Recall (macro)	F1-score (macro)
NASNet only	95.84	0.956	0.954	0.955
ViT only	94.62	0.945	0.942	0.943
NASNet + ViT (Simple Concatenation)	96.91	0.969	0.967	0.968
NASNet + ViT (Late Fusion)	97.28	0.972	0.970	0.971
NASNet + ViT + Cross-Gating (Proposed)	98.54	0.985	0.984	0.984

4.3 Ablation study on model components

A controlled ablation study was performed to isolate the contribution of individual architectural components under identical experimental conditions. As shown in Table 3, the NASNet-only model achieves an accuracy of 95.84% with a macro F1-score of 0.955, reflecting its strong ability to capture fine-grained local spatial and textural features of liver lesions, but also indicating limitations in modeling long-range contextual relationships across heterogeneous tumor regions. The Vision Transformer–only configuration attains an accuracy of 94.62% and a macro F1-score of 0.943, demonstrating effective global context modeling and inter-channel dependency learning, while exhibiting reduced sensitivity to localized anatomical patterns when used without a convolutional backbone.

Combining NASNet and Vision Transformer through simple feature concatenation improves accuracy to 96.91% and macro F1-score to 0.968, indicating that joint access to local and global representations enhances discriminative capability. The late-fusion strategy further increases accuracy to 97.28% with a macro F1-score of 0.971, suggesting that decision-level aggregation of complementary predictions provides additional robustness. However, these fusion approaches yield only incremental gains, implying that performance improvements cannot be attributed solely to increased model capacity or the presence of dual-branch architectures. In contrast, the proposed cross-gated NASNet–ViT model achieves the highest performance across all metrics, with an accuracy of 98.54%, macro precision of 0.985, macro recall of 0.984, and macro F1-score of 0.984. This consistent improvement demonstrates that the cross-gating mechanism enables adaptive bidirectional modulation between convolutional and transformer feature streams, selectively emphasizing complementary information while suppressing redundant activations, thereby delivering superior discriminative performance beyond architectural scale or simple fusion alone.

4.4 Performance comparison

Table 4 and Fig. 12 present a comparative overview of representative deep learning–based liver cancer classification approaches reported in the literature. The comparison is intended to highlight general performance trends and architectural evolution rather than to establish direct numerical superiority, as the referenced studies differ in datasets, imaging modalities, task formulations, and evaluation protocols. Earlier transformer-based radiomics approaches, such as the MRI radiomics with biomarker fusion model reported by [15], achieved relatively modest accuracy (75.80%), which can be attributed to reliance on handcrafted radiomic features, limited modeling of local spatial patterns, and smaller or more heterogeneous datasets. Similarly, approaches integrating conventional CNN backbones such as VGG19 [16] and ResNet50 [13] demonstrate improved performance (85.00% and 89.50%, respectively), reflecting the benefits of deeper feature hierarchies and residual learning; however, these models primarily focus on convolutional feature extraction and may be constrained in capturing long-range contextual dependencies and subtle inter-class variations present in liver MRI data.

More recent hybrid architectures incorporating segmentation or multi-stage learning strategies, such as UNet70 [7] and V-Net combined with VGG16 [12], further enhance classification accuracy by explicitly modeling lesion regions and reducing background interference. Despite these improvements, such approaches often depend on manual or semi-automatic segmentation and are typically evaluated on different imaging modalities or task settings, limiting direct comparability.

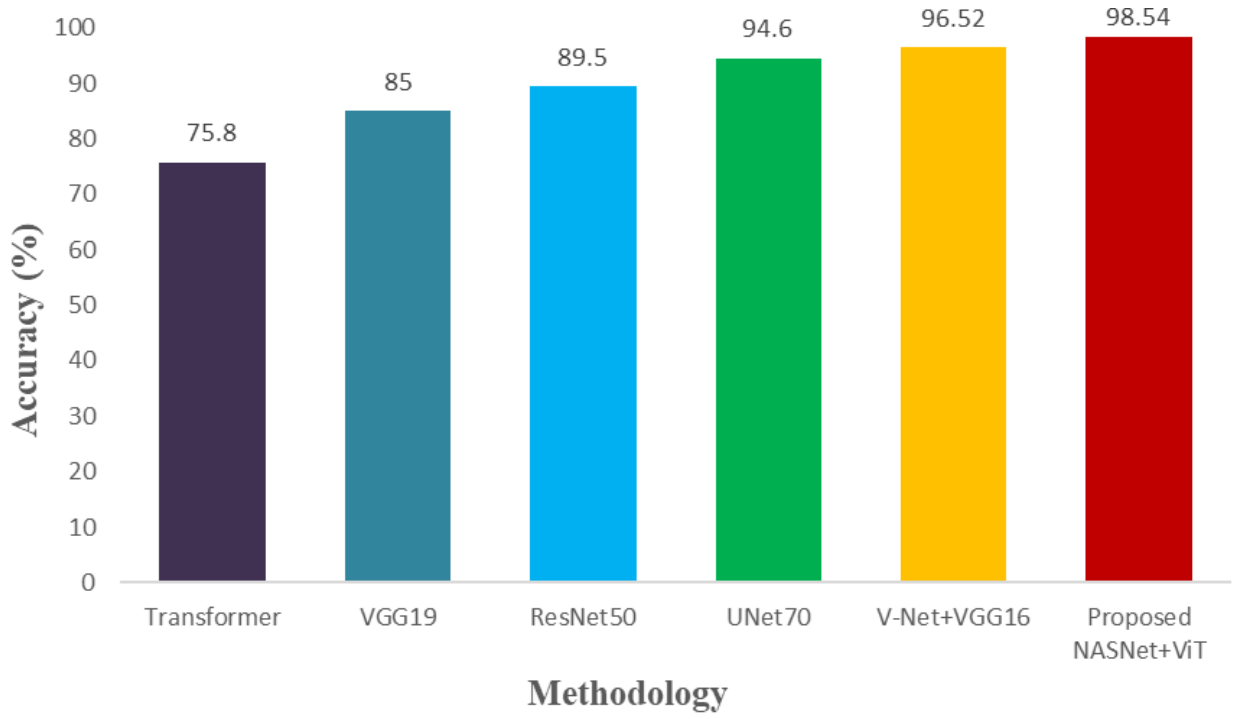

Figure. 12 Graphical illustration of accuracy comparison of the proposed model with existing methods

Table 4. Comparison of accuracy of proposed model with existing methods

Author	Methodology	Accuracy (%)
He et al. [15]	Transformer	75.80
Wu et al. [16]	VGG19	85.00
Yin et al. [13]	ResNet50	89.50
Gowda N et al. [7]	UNet70	94.60
Slama et al. [12]	V-Net + VGG16	96.52
Proposed Hybrid NASNet - ViT with Cross Gating Fusion Model		**98.54**

Note: Results are taken from the original studies and shown only to indicate general performance trends; direct numerical comparison is not intended due to differences in datasets, modalities, and evaluation protocols.

In contrast, the proposed model integrates a neural architecture search–optimized convolutional backbone for fine-grained local feature extraction with Vision Transformer modules capable of modeling global contextual relationships. The cross-gating fusion mechanism further enables adaptive interaction between convolutional and transformer features, selectively emphasizing complementary information while suppressing redundant responses. This integrated design contributes to the strong performance observed on the employed five-class MRI dataset under the defined experimental protocol. Overall, the comparison illustrates a progressive trend toward more expressive hybrid architectures for liver cancer classification, while underscoring that reported performance values should be interpreted within their respective experimental contexts.

5. Conclusion

Accurate liver cancer detection from MRI images remains a challenging task due to subtle inter-class variations, heterogeneous tumor appearances, and low contrast between malignant, benign, and healthy liver tissues. These challenges are addressed by the proposed hybrid NASNet–ViT architecture with cross-gating fusion, which combines transformer-based global contextual modeling with convolutional neural networks for local spatial feature extraction. The NASNet backbone captures fine-grained anatomical and textural characteristics of liver lesions, while the Vision Transformer modules model long-range dependencies and inter-channel relationships. The cross-gating fusion mechanism selectively enhances complementary feature responses, resulting in improved discriminative capability. The proposed model achieves an overall classification accuracy of 98.54%, with consistently high precision, recall, and F1-score values across all five liver classes, indicating strong image-level generalization and minimal misclassification. Stable convergence behavior, near-perfect ROC characteristics, and low confusion among clinically similar tumor types

further support the reliability of the proposed approach. Future work will focus on subject-wise and multi-center validation using datasets with patient-level metadata, as well as extension to three-dimensional, multi-phase, and multi-modal MRI data, to further assess clinical generalizability and real-world applicability.

Conflicts of interest

The authors declare that there are no conflicts of interest related to this manuscript.

Author contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision and project administration, have been done by 2nd author.

References

[1] J. H. Oh, D. W. Jun, "The latest global burden of liver cancer: A past and present threat", *Clinical and Molecular Hepatology*, Vol. 29, No. 2, 355, 2023.

[2] B. McMahon *et al.*, "Opportunities to address gaps in early detection and improve outcomes of liver cancer", *JNCI Cancer Spectrum*, Vol. 7, No. 3, pkad034, 2023.

[3] K. Kobayashi *et al.*, "Screening methods for early detection of hepatocellular carcinoma", *Hepatology*, Vol. 5, No. 6, pp. 1100–1105, 1985.

[4] S. Sontakke, J. Lohokare, R. Dani, "Diagnosis of liver diseases using machine learning", In: *Proc. of Int. Conf. Emerging Trends & Innovation in ICT (ICEI)*, pp. 129–133, 2017.

[5] L. Huang *et al.*, "Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning", *Nature Communications*, Vol. 14, No. 1, 48, 2023.

[6] A. Bakrania *et al.*, "Artificial intelligence in liver cancers: Decoding the impact of machine learning models in clinical diagnosis of primary liver cancers and liver cancer metastases", *Pharmacological Research*, Vol. 189, 106706, 2023.

[7] Y. Gowda, R. V. Manjunath, "Automatic liver tumor classification using UNet70: A deep learning model", *Journal of Liver Transplantation*, Vol. 18, 100260, 2025.

[8] S. Song *et al.*, "Deep learning based on intratumoral heterogeneity predicts histopathologic grade of hepatocellular carcinoma", *BMC Cancer*, Vol. 25, No. 1, 497, 2025.

[9] R. Stollmayer *et al.*, "LI-RADS-based hepatocellular carcinoma risk mapping using contrast-enhanced MRI and self-configuring deep learning", *Cancer Imaging*, Vol. 25, No. 1, 36, 2025.

[10] L. Jia *et al.*, "Deep learning-based MRI model for predicting P53-mutated hepatocellular carcinoma", *BMC Medical Imaging*, Vol. 25, No. 1, 506, 2025.

[11] H. Wei *et al.*, "Deep learning-based 3D quantitative total tumor burden predicts early recurrence of BCLC A and B HCC after resection", *European Radiology*, Vol. 35, No. 1, pp. 127–139, 2025.

[12] A. B. Slama, H. Sahli, Y. Amri, S. Labidi, "V-NET-VGG16: Hybrid deep learning architecture for optimal segmentation and classification of multi-differentiated liver tumors", *Intelligence-Based Medicine*, Vol. 11, 100210, 2025.

[13] L. Yin *et al.*, "Deep learning-based CT radiomics predicts prognosis of unresectable hepatocellular carcinoma treated with TACE-HAIC combined with PD-1 inhibitors and tyrosine kinase inhibitors", *BMC Gastroenterology*, Vol. 25, No. 1, 24, 2025.

[14] M. Li *et al.*, "Interactive explainable deep learning model for hepatocellular carcinoma diagnosis at gadoxetic acid–enhanced MRI: A retrospective, multicenter diagnostic study", *Radiology: Imaging Cancer*, Vol. 7, No. 3, e240332, 2025.

[15] R. He *et al.*, "Transformer-based deep learning for preoperative prediction of microvascular invasion in hepatocellular carcinoma", *Cancers*, Vol. 17, No. 20, 3314, 2025.

[16] Q. Wu *et al.*, "MRI-based deep learning radiomics to differentiate dual-phenotype hepatocellular carcinoma from HCC and intrahepatic cholangiocarcinoma: A multicenter study", *Insights into Imaging*, Vol. 16, No. 1, 27, 2025.

[17] B. Jeong *et al.*, "Predicting post-hepatectomy liver failure in patients with hepatocellular carcinoma: Nomograms based on deep learning analysis of gadoxetic acid-enhanced MRI", *European Radiology*, Vol. 35, No. 5, pp. 2769–2782, 2025.

[18] Y. Zhang *et al.*, "Deep learning empowered gadolinium-free contrast-enhanced abbreviated

MRI for diagnosing hepatocellular carcinoma", *JHEP Reports*, Vol. 7, No. 5, 101392, 2025.
[19] M. Gu *et al.*, "Multilayer perceptron deep learning radiomics model based on Gd-BOPTA MRI to identify vessels encapsulating tumor clusters in hepatocellular carcinoma: A multicenter study", *Cancer Imaging*, Vol. 25, No. 1, 87, 2025.
[20] Y. Dai *et al.*, "Automated predictive framework using AI and deep learning approaches for early detection and classification of liver cancer", *Frontiers in Oncology*, Vol. 15, 1650800, 2025.
[21] G. N. Yashaswini *et al.*, "Deep learning technique for automatic liver and liver tumor segmentation in CT images", *Journal of Liver Transplantation*, Vol. 17, 100251, 2025.
[22] Kaggle, "Liver cancer multiclass dataset", [Online]. Available: https://www.kaggle.com/datasets/ucimachinelearning/liver-cancer-multiclass-dataset
[23] S. Vallukappully, I. van der Linde, A. Chakraborty, "Early detection and classification of diabetic retinopathy by transfer learning of NASNet-Large and ResNet-50 convolutional neural networks", *Informatics in Medicine Unlocked*, 101688, 2025.
[24] B. Zoph, V. Vasudevan, J. Shlens, Q. V. Le, "Learning transferable architectures for scalable image recognition", In: *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 8697–8710, 2018.
[25] H. Yin *et al.*, "A-ViT: Adaptive tokens for efficient vision transformer", In: *Proc. of IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 10809–10818, 2022.
[26] K. Al-Hammuri *et al.*, "Vision transformer architecture and applications in digital health: A tutorial and survey", *Visual Computing for Industry, Biomedicine, and Art*, Vol. 6, No. 1, 14, 2023.
[27] A. O. Adedoja, P. A. Owolawi, T. Mapayi, "Intelligent mobile plant disease diagnostic system using NASNet-Mobile deep learning", *IAENG International Journal of Computer Science*, Vol. 49, No. 1, 2022.
[28] S. Saleem, M. I. Sharif, "An integrated deep learning framework leveraging NASNet and vision transformer with mixprocessing for accurate and precise diagnosis of lung diseases", *arXiv Preprint*, arXiv:2502.20570, 2025.
[29] G. Chen *et al.*, "Predicting bone metastasis risk of colorectal tumors using radiomics and deep learning ViT model", *Journal of Bone Oncology*, Vol. 51, 100659, 2025.
[30] S. Amir, Y. Gandelsman, S. Bagon, T. Dekel, "Deep ViT features as dense visual descriptors", *arXiv preprint* arXiv:2112.05814, 2021.
[31] W. M. S. P. B. Wickramasinghe, M. B. Dissanayake, "Vision transformers for glioma classification using T1 magnetic resonance imaging", *Artificial Intelligence in Health*, Vol. 2, pp. 68–80, 2024.
[32] A. Ullah *et al.*, "A vision transformer model-integrated mobile application for early and accurate detection of lumpy skin disease in cattle", *Scientific Reports*, 2025.