

RESEARCH

Advanced Liver Fibrosis Detection and Classification Through Deep Learning-Driven Image Analysis

Umesh Kumar Lilhore · Sarita Simaiya · Roobaea Alroobaea · Abdullah M. Baqasah · Majed Alsafyani · Afnan Alhazmi · MD Monish Khan

Received: 28 March 2025 / Revised: 10 October 2025 / Accepted: 2 November 2025

© The Author(s) 2025

Abstract

Liver fibrosis is a progressive liver disease that impairs liver function, necessitating a precise and timely diagnosis to ensure effective treatment and management. This paper describes a hybrid deep learning method for classifying liver fibrosis that utilises Convolutional Neural Networks (CNNs) in conjunction with an advanced Attention Mechanism to enhance performance in medical image analysis. Traditional methods for detecting liver fibrosis often suffer from issues such as overfitting, poor generalisation across different datasets, and difficulties in highlighting critical areas of interest in medical imaging. We propose a novel hybrid model that combines the EfficientNet-B7 architecture and a dual-attention mechanism to address these limitations. Incorporating spatial and channel attention mechanisms enhances EfficientNet-B7, a model renowned for its superior performance and computational efficiency. This integration enables the model to focus on the most crucial features of liver scans, thereby enhancing the accuracy of fibrosis detection and severity classification. We utilise multi-scale feature extraction with dilated convolutions to capture finer details, thereby enhancing model sensitivity. This approach introduces a significant innovation by combining EfficientNet-B7 with dual-attention mechanisms, which improves classification accuracy and computational efficiency while addressing common limitations in current solutions. The model is evaluated against the publicly available Liver Fibrosis Dataset (LiverFib), which comprises liver imaging scans annotated with corresponding fibrosis stages. Our hybrid model outperforms traditional CNN architectures, such as VGGNet, DenseNet, ResNet, and U-Net, as well as attention-based models, achieving 98.5% accuracy, 97.2% precision, 96.5% recall, 96.8% F1-score, and an AUC of 0.98. The attention mechanism enhances classification performance by focusing on critical areas of images, thereby increasing model interpretability and facilitating clinicians' ability to identify fibrosis-related regions. The findings demonstrate that the proposed method is a highly effective tool for the automated detection and classification of liver fibrosis, providing high accuracy and valuable insights for informed clinical decision-making.

Keywords Liver fibrosis · Deep learning · EfficientNet-B7 · Dual-attention mechanism · Medical image classification



Abbreviations

AI	Artificial intelligence
AUC	Area under the curve
CNN	Convolutional neural networks
CT	Computed tomography
DL	Deep learning
F1-score	F-measure
MRI	Magnetic resonance imaging
ResNet	Residual networks
U-Net	U-Net architecture
VGGNet	Visual geometry group network
B7	EfficientNet-B7
XAI	Explainable AI
SVM	Support vector machine
CT	Computed tomography
LSTM	Long short-term memory
ML	Machine learning
VGGNet	Visual geometry group network
DNN	Deep neural network
ResNet	Residual networks
U-Net	U-Net architecture
ROI	Region of interest
LiverFib	Liver fibrosis dataset
LiverFib	Liver fibrosis dataset
AUC-ROC	Area under the curve, receiver operating characteristic

1 Introduction

Liver fibrosis is a progressive disorder characterised by the accumulation of extracellular matrix proteins in the liver, leading to the formation of scar tissue. If untreated, this scarring may advance to cirrhosis, hepatic failure, and potentially liver carcinoma [1]. The prompt and precise identification of liver fibrosis facilitates effective management and treatment, thereby averting additional complications. Liver biopsy has conventionally been the standard approach for diagnosing and staging fibrosis; nonetheless, this method is invasive, costly, and susceptible to sampling errors. Consequently, non-invasive imaging modalities, such as ultrasound, MRI, and CT scans, are increasingly utilised as alternatives. Due to the complexity of liver tissue and the subtle changes associated with fibrosis, accurately identifying and staging liver fibrosis remains challenging using these imaging modalities despite their potential [2, 3].

The application of AI and deep learning has shown considerable promise in improving the accuracy and effectiveness of medical image analysis. CNNs have demonstrated exceptional efficacy across various tasks, including organ segmentation, tumour identification, and disease classification. Notwithstanding these achievements, conventional CNNs face considerable constraints in the context of liver fibrosis detection [4, 5]. These include overfitting, inadequate generalisation across diverse datasets, and the inability to focus on relevant regions of the image that signify fibrosis. Attention mechanisms have arisen as a remedy to these issues. Attention mechanisms enhance model performance in detecting subtle features by directing focus to the most informative regions of an image, particularly in medical imaging tasks where fine details are essential. Nonetheless, the utilisation of

attention mechanisms in detecting liver fibrosis remains insufficiently investigated. This gap presents an opportunity to improve the precision and comprehensibility of deep learning models for this task [6, 7].

Liver fibrosis is a severe and progressive liver disease that can lead to cirrhosis, liver failure, and even liver cancer if not detected and treated early. A prompt and precise diagnosis is essential to avert the progression of the disease to these perilous stages [8]. Liver biopsy has conventionally been regarded as the definitive method for diagnosing and staging fibrosis. This procedure is invasive, costly, and entails risks, including haemorrhage and infection. Liver biopsies are susceptible to sampling errors and are unsuitable for routine screening. Consequently, non-invasive imaging modalities such as ultrasound, MRI, and CT scans have gained prevalence in detecting liver fibrosis. Although these methods are safer and more accessible, they nonetheless possess limitations [9].

Liver fibrosis progresses incrementally, and alterations in liver tissue are frequently nuanced, particularly during the initial phases. These imaging modalities may struggle to accurately identify subtle alterations and often rely on manual interpretation, which can be labour-intensive and subjective, depending on the clinician's expertise. This presents a considerable challenge: identifying liver fibrosis early, accurately, and reliably without the need for invasive procedures. The demand for improved diagnostic tools is increasingly critical as the global population at risk of liver fibrosis continues to escalate due to factors such as viral hepatitis, alcohol intake, and non-alcoholic fatty liver disease [2, 3].

Existing imaging techniques are insufficient to meet the growing demand for precise, non-invasive, and effective fibrosis detection. This is the domain where AI and DL technologies can effectuate change. Artificial intelligence has demonstrated significant potential in enhancing the precision of medical image analysis by automating disease detection and classification. Nonetheless, despite the use of sophisticated deep learning models, issues such as overfitting, generalisation across heterogeneous datasets, and the emphasis on the most pertinent aspects of the images persist. These challenges constrain the efficacy of AI models in detecting liver fibrosis [9]. In light of these challenges, there is an urgent need for more sophisticated, reliable, and efficient solutions capable of overcoming the limitations of conventional imaging techniques. The primary objective is to develop a system capable of accurately and consistently identifying early-stage liver fibrosis, thereby enabling clinicians to make more informed and expedient decisions, ultimately enhancing patient outcomes without the need for invasive methods.

The presented research presents a hybrid deep learning model that combines the computational efficiency of EfficientNet-B7 with the functionalities of dual-attention mechanisms to address the issues outlined in the previous section [10]. Integrating attention mechanisms allows the model to independently focus on the most informative areas of liver images, improving accuracy and interpretability. The key contributions of this research are as follows.

- *Hybrid Deep Learning Model:* This research introduces an innovative hybrid deep learning model that integrates the EfficientNet-B7 architecture with dual-attention mechanisms. This combination enables the model to concentrate more effectively on the critical features in liver images, thereby enhancing accuracy in fibrosis detection.
- *Attention Mechanisms for Enhanced Focus:* The model employs spatial and channel attention mechanisms to highlight the key regions of the liver that exhibit fibrosis. This enhances the detection of nuanced alterations and improves the model's interpretability, facilitating a deeper understanding of which areas of the image are most crucial for diagnosis.
- *Outstanding Performance:* The proposed model surpasses the conventional CNN architectures, including VGGNet, DenseNet, ResNet, and U-Net, yielding exceptional results. With an accuracy of 98.5%, a precision of 97.2%, a recall of 96.5%, an F1-score of 96.8%, and an AUC of 0.98, it exhibits high efficacy in detecting liver fibrosis.

- **Strong Generalisation Across Datasets:** The model exhibits exceptional generalisation capabilities, consistently performing well across diverse liver imaging datasets. This robustness ensures its utility in various clinical environments, making it a versatile tool for practical medical applications.

The remainder of the paper is structured as follows: Sect. 2 reviews the literature on liver fibrosis detection using deep learning, highlighting the key challenges in the field. In Sect. 3, we describe the materials, methods, and architecture of the proposed hybrid model. Section 4 presents the experimental results, including comparisons with traditional CNN-based approaches. Finally, Sect. 5 concludes the paper and discusses potential future directions for improving liver fibrosis detection using AI.

2 Literature Review

Liver fibrosis is a progressive ailment that may result in severe conditions, including cirrhosis and liver cancer, if not accurately diagnosed and managed. A prompt and precise diagnosis is crucial to avert additional harm to the liver. Although conventional techniques such as liver biopsy have been the benchmark for evaluating liver fibrosis, they are invasive and susceptible to inaccuracies. Consequently, there is an increasing interest in non-invasive imaging modalities, including ultrasound, MRI, and CT scans, integrated with machine learning algorithms to enhance the detection and staging of liver fibrosis. This literature review examines prior studies on Deep Learning for Liver Fibrosis Detection, Attention Mechanisms in Medical Imaging, and Hybrid Approaches in Liver Disease Diagnosis, all of Which Are pertinent to improving the precision and dependability of fibrosis detection.

2.1 Deep Learning for Liver Fibrosis Detection

Deep learning techniques, particularly CNNs, have demonstrated considerable potential in enhancing the automated detection of liver fibrosis. Numerous studies have shown the efficacy of these methods in accurately classifying liver conditions and identifying fibrosis. Lee et al. (2020) investigated the application of deep convolutional neural networks to classify liver fibrosis using ultrasound images, achieving robust results that demonstrate the efficacy of deep learning. Wong et al. (2021) employed deep learning models to predict the progression of non-alcoholic fatty liver disease (NAFLD) and liver fibrosis, highlighting the growing importance of AI in liver disease diagnostics.

Rahman et al. (2022) employed deep-learning liver and tumour segmentation methodologies in CT images, successfully identifying liver fibrosis and associated conditions. Dai et al. (2024) introduced a deep learning model to predict the risk of liver fibrosis progression in patients with metabolic dysfunction-associated steatotic liver disease (MASLD), demonstrating the applicability of deep learning across diverse patient demographics. Furthermore, Shaheen et al. (2023) investigated a hybrid model that integrates CNN and Capsule Networks to classify cirrhosis, attaining enhanced outcomes in managing intricate liver disease characteristics. Notwithstanding these achievements, challenges, including overfitting, inadequate generalisation, and difficulty detecting subtle changes in fibrosis, persist as substantial obstacles for existing models.

2.2 Attention Mechanisms in Medical Imaging

Attention mechanisms have become increasingly important in medical image analysis, enabling models to focus on the most relevant areas of an image, which is particularly useful for detecting the subtle features of liver fibrosis. Zha et al. (2024) incorporated attention mechanisms into MRI scans for liver fibrosis triage, showing that the ability to highlight specific regions of interest significantly improved the model's performance in detecting clinically significant fibrosis. Alshagathrh et al. (2024) applied attention-based deep-learning models to ultrasound images for detecting hepatocyte ballooning, which is associated with liver fibrosis. Their research demonstrated

that attention mechanisms enabled the model to focus on the liver regions most severely affected by fibrosis, thereby improving detection accuracy.

Similarly, Abinaya et al. (2024) integrated attention mechanisms into a hybrid MRMR-BiLSTM-CNN architecture to improve liver fibrosis detection. The attention mechanism enabled the model to focus on key regions that showed early signs of fibrosis, enhancing its ability to detect subtle changes. Yu et al. (2018) also leveraged attention mechanisms to automatically score liver fibrosis stages, a critical step toward making AI-based diagnostic tools more interpretable and clinically useful. Shaban et al. (2024) demonstrated that incorporating attention mechanisms into machine learning models for liver disease classification enhanced their ability to detect key features and improve overall performance.

2.3 Hybrid Approaches in Liver Disease Diagnosis

Hybrid models that combine various machine-learning techniques have gained popularity for addressing the complex nature of liver disease detection and classification. Deshmukh et al. (2023) proposed a hybrid model that combines CNN with Extreme Learning Machines (ELM) for liver disease forecasting. Their approach improved liver disease detection accuracy, demonstrating the potential of hybrid models to capture different aspects of the disease. Jabbar et al. (2023) also worked with hybrid machine learning models to classify liver fibrosis stages, finding that integrating multiple approaches improved the performance across various fibrosis stages.

Additionally, Prabagaran et al. (2024) introduced a hybrid approach that combined different models to classify non-alcoholic fatty liver disease (NAFLD) stages, achieving enhanced classification performance. Tanfoni et al. (2024) used a hybrid deep learning model for liver tumour segmentation, integrating DeepLabV3+ and Hidden Markov Models, which led to improved diagnostic precision and segmentation quality. Furthermore, Zha et al. (2024) developed a fully automated hybrid approach using conventional MRI to triage clinically significant liver fibrosis, making the process faster and more reliable. These hybrid models can handle complex data more effectively and are particularly valuable in real-world clinical settings where diverse imaging modalities are often used.

2.4 Comparative Analysis

Recent advancements in deep learning have significantly improved the detection of liver fibrosis, with CNN models demonstrating strong capabilities in classifying liver conditions. However, challenges such as overfitting and generalisation across datasets remain. Attention mechanisms have emerged as an effective solution to these challenges by allowing models to focus on the most relevant features of liver images, improving detection accuracy and interpretability. Additionally, hybrid approaches, which combine various machine learning techniques, have proven successful in enhancing diagnostic performance and generalisation across different clinical settings. The continued integration of attention mechanisms and hybrid models is expected to advance liver fibrosis detection further, making it a more reliable and efficient tool for clinicians. Table 1 presents a comparative analysis of various existing research on liver disease.

3 Materials and Methods

This section outlines the architecture and methodology of the proposed hybrid deep learning model, which integrates EfficientNet-B7 with dual-attention mechanisms. The document encompasses the dataset utilised, pre-processing procedures, model architecture, training methodology, and performance evaluation metrics.

Table 1 Comparative analysis of various existing research on liver disease

Study	Model/approach	Focus area	Key findings	Performance	Challenges
Lee et al. (2020)	CNN	Ultrasound-based fibrosis classification	High accuracy in detecting fibrosis from ultrasound images	High accuracy and specificity	Limited dataset, image quality
Wong et al. (2021)	Deep learning (multiple models)	NAFLD and fibrosis progression	AI models effectively predict and monitor fibrosis in NAFLD	Strong predictive ability	Oversetting, dataset variability
Rahman et al. (2022)	ResUNet	Liver and tumour segmentation from CT scans	ResUNet performs well in liver and tumour segmentation	High segmentation quality	Sensitivity to subtle fibrosis signs
Dai et al. (2024)	Deep learning for fibrosis prediction	Fibrosis risk in MASLD	Predicts fibrosis progression in MASLD patients accurately	High prediction accuracy	Needs diverse datasets
Shaheen et al. (2023)	Hybrid CNN+capsule network	Cirrhosis and fibrosis classification	The hybrid model improves cirrhosis detection accuracy	High F1-score	Model complexity, interpretability issues
Zha et al. (2024)	Hybrid+ attention mechanisms	MRI-based fibrosis triage	Attention mechanisms improve accuracy by focusing on key areas	High AUC, sensitivity, and specificity	MRI quality and dataset variation
Alshagathr et al. (2024)	Hybrid deep learning + machine learning	Hepatocyte ballooning detection in ultrasound	Improves detection of ballooning hepatocytes, indicating fibrosis	High accuracy, precision, and recall	Ultrasound image quality, low resolution
Abinaya et al. (2024)	Hybrid MRMR-BiLSTM-CNN with Attention	Liver fibrosis detection	The attention-based hybrid model enhances fibrosis detection	High accuracy, precision, and recall	Requires large datasets
Prabagaran et al. (2024)	Hybrid CNN+Other models	NAFLD fibrosis classification	Improves classification accuracy for NAFLD stages	High accuracy	Complexity, oversetting risks
Tanfoni et al. (2024)	Hybrid deep learning (DeepLabV3+ + HMM)	Liver tumour segmentation	The hybrid model aids in liver tumour segmentation and fibrosis detection	High segmentation quality	High computational cost, complex integration

Table 2 Distribution of images across fibrosis stages

Fibrosis stage	Number of images
Stage 0 (healthy liver)	500 images
Stage 1 (mild fibrosis)	700 images
Stage 2 (moderate fibrosis)	800 images
Stage 3 (severe fibrosis)	600 images
Stage 4 (cirrhosis)	500 images
Total	3100 images

3.1 Dataset Details

This research used the Liver Fibrosis Dataset (LiverFib), a publicly accessible database that contains liver images labelled with different stages of liver fibrosis [11, 12]. This dataset is essential for creating and evaluating the proposed hybrid deep learning model. It provides the necessary data for the model to detect and classify various stages of liver fibrosis using medical imaging. Table 2 presents the dataset count and categories of the images.

Table 2 illustrates the distribution of images in the LiverFib dataset across the five stages of liver fibrosis. In total, the dataset contains 3100 images. Stage 0 (Healthy Liver) consists of 500 images, which is comparatively fewer given the rarity of healthy liver scans in clinical settings. Stage 1 (Mild Fibrosis) contains 700 images,

followed by Stage 2 (Moderate Fibrosis) with 800 images, representing the highest number due to its higher occurrence and easier identification in medical imaging. Stage 3 (Severe Fibrosis) contains 600 images, and Stage 4 (Cirrhosis) has 500 images, reflecting more advanced stages of liver damage. This distribution is typical of real-world datasets, where moderate stages are more commonly represented and identifiable, while advanced stages are less frequently available.

3.1.1 Key Features of the Dataset

This dataset primarily consists of MRI and CT scans, two common imaging modalities for assessing liver conditions. The scans are of high quality, with a resolution of (512×512) pixels, ensuring the visibility and detectability of essential characteristics related to liver fibrosis by the deep learning model [2, 4, 9]. Each image is presented in grayscale (single channel), highlighting intensity values and allowing the model to identify patterns based on varying tissue density levels indicative of fibrosis stages. Additionally, medical experts annotate each image with the fibrosis stage, ensuring precise labelling for training objectives.

3.1.2 Challenges in the Dataset

The liver images in this dataset display variability in quality due to factors such as imaging equipment, patient health, and scan resolution. Some scans may exhibit enhanced contrast, while others may contain noise or artefacts that hinder the identification of fibrosis. The dataset includes diverse patient demographics, such as age, gender, and ethnicity, enhancing the model's capacity to generalise to a broad population [7, 9, 13]. These variations present challenges but provide substantial opportunities for our model to develop the capability to handle diverse conditions, thus ensuring optimal performance in real-world clinical situations.

3.2 Data Pre-processing

Data pre-processing is crucial in preparing the dataset for training a deep learning model, as it directly influences the model's performance and generalisation ability. The preliminary stage of this procedure entails resizing images to a consistent dimension of (224×224) pixels. Deep learning models, including EfficientNet, require input images to have consistent dimensions. Following resizing, the images undergo normalisation, calibrating the pixel values to a range of $(0-1)$ or $(-1-1)$. Normalising the images prevents any singular feature from dominating others during training, thereby expediting model convergence and enhancing learning efficiency [2, 4, 7, 14].

Data augmentation is utilised to artificially expand the dataset and improve the diversity of the training data. Augmentation techniques applied randomly encompass rotation, zooming, flipping, translation, and shearing of images. This enhances the model's generalisation and reduces overfitting by providing a wider range of training examples. Additionally, noise reduction techniques like Gaussian smoothing or median filtering are utilised to improve the images by removing superfluous details that could obscure the model [14, 15].

Table 3 Distribution of images across fibrosis stages (after pre-processing and augmentation)

Fibrosis stage	Number of images (before augmentation)	Number of images (after augmentation)
Stage 0 (healthy liver)	500 images	1500 images
Stage 1 (mild fibrosis)	700 images	2100 images
Stage 2 (moderate fibrosis)	800 images	2400 images
Stage 3 (severe fibrosis)	600 images	1800 images
Stage 4 (cirrhosis)	500 images	1500 images
Total	3100 images	9800 images

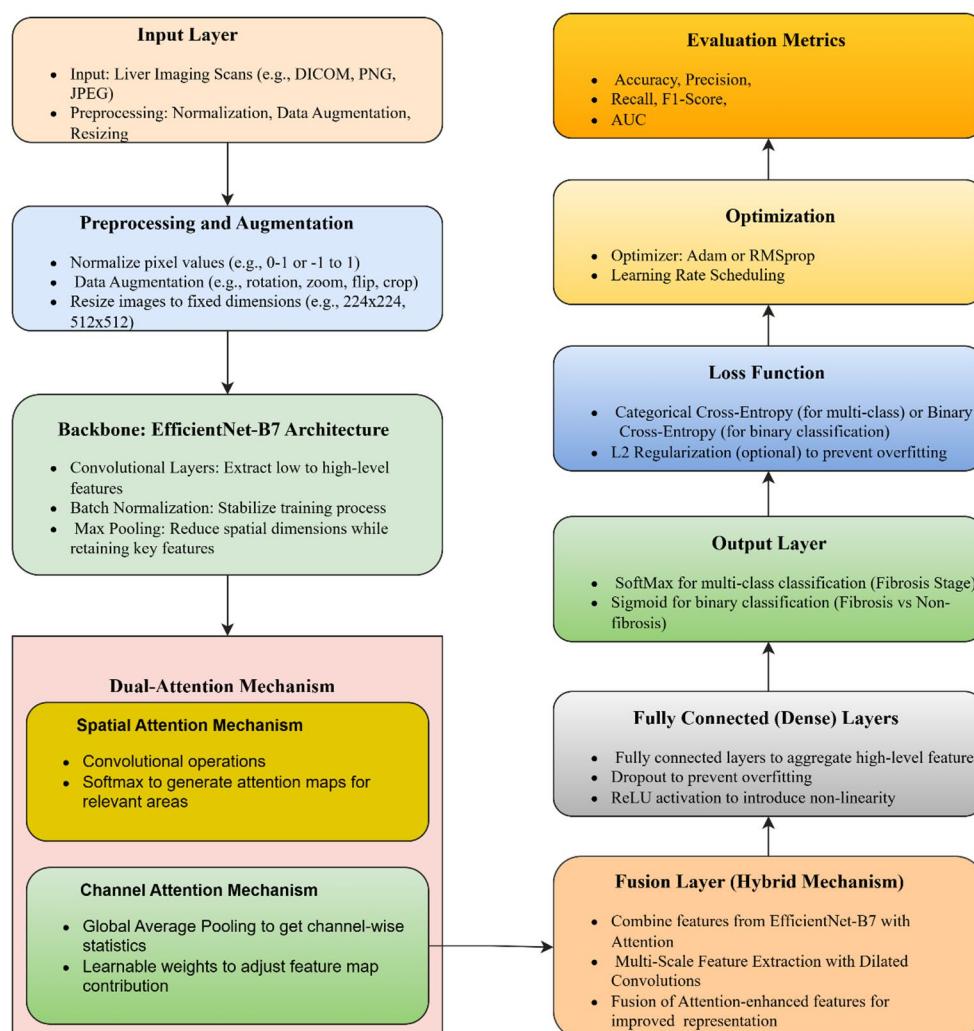
Cropping and masking are utilised to direct the model's focus on critical areas of the image, such as the liver, thereby enhancing its capacity to identify fibrosis. Concerning the dataset, after the pre-processing procedures, the images are divided into training, validation, and test sets, typically allocated as 70% for training, 15% for validation, and 15% for testing. This guarantees that the model is assessed on new data, thus enabling the evaluation of its generalisation abilities.

Table 3 displays the dataset count before and after data pre-processing. Following data augmentation, the dataset's image count escalates from 3100 to 9800, enhancing the dataset for model training. The pre-processing steps of resizing, normalisation, augmentation, and splitting ensure the dataset is adequately prepared for deep learning, enabling the model to effectively learn essential features for precise liver fibrosis classification.

3.3 Proposed Model Architecture

The proposed hybrid deep learning model for liver fibrosis detection integrates EfficientNet-B7 with a dual-attention mechanism to improve the classification of liver fibrosis stages (as presented in Fig. 1). EfficientNet is a convolutional neural network architecture recognised for its high performance, achieved with fewer parameters and lower computational costs. EfficientNet-B7, the most advanced iteration of EfficientNet, functions as the backbone for feature extraction owing to its superior accuracy and computational efficiency [9, 15–17]. Incorporating the dual-attention mechanism enhances the model's capability to concentrate on the most informative

Fig. 1 Architecture of proposed hybrid model

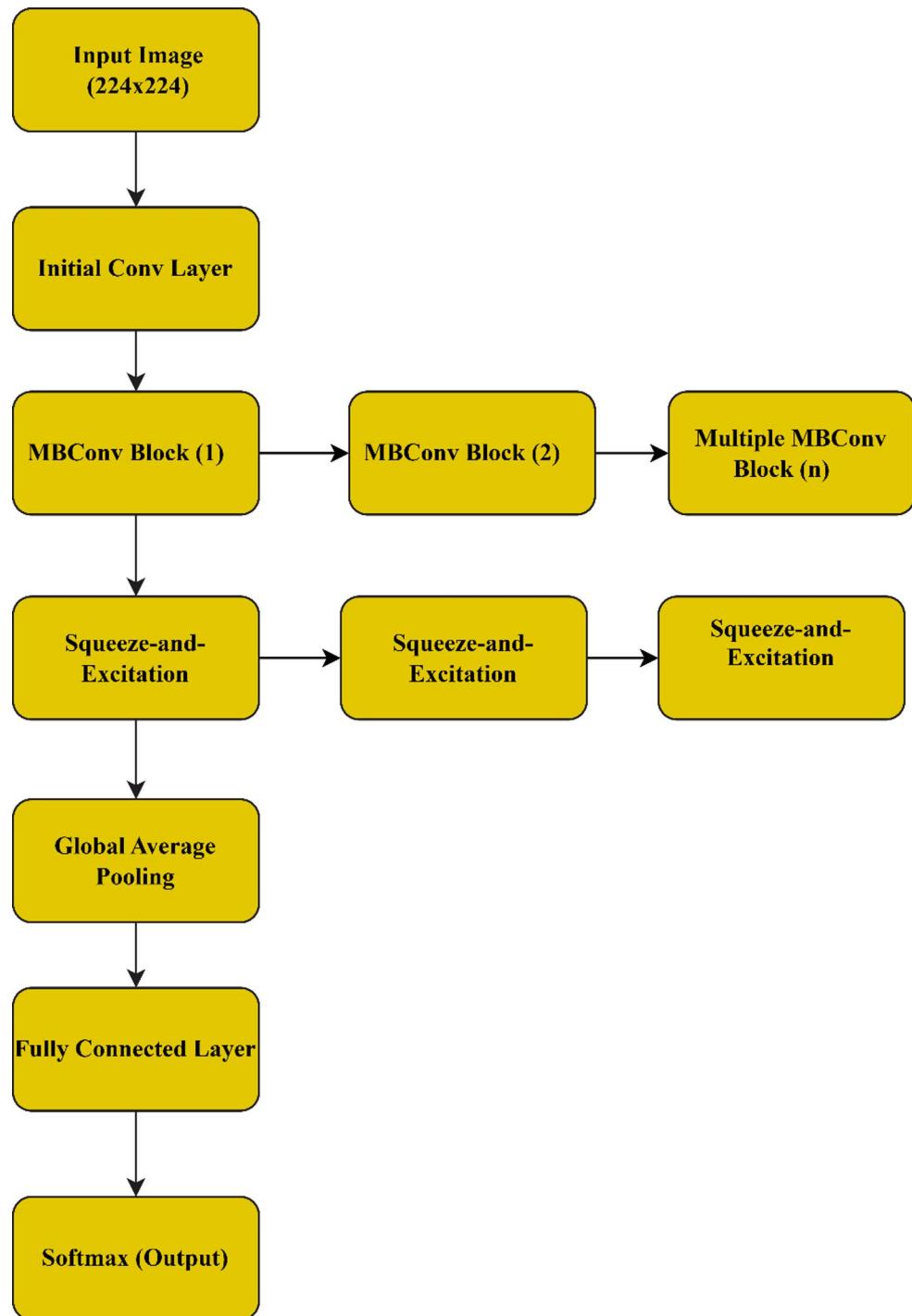


aspects of the input image, thereby facilitating the detection of subtle variations in liver scans indicative of fibrosis. This attention mechanism comprises two types: spatial attention and channel attention, as presented in Fig. 1.

3.4 Working of the Proposed Model

The complete working of the proposed hybrid model is as follows.

Fig. 2 Layered EfficientNet-B7 architecture



3.4.1 EfficientNet-B7 (Feature Extraction)

The EfficientNet-B7 model is the feature extractor in the proposed deep-learning architecture for liver fibrosis detection. The model's architecture was selected for its capacity to attain high accuracy while utilizing fewer parameters and operations, rendering it suitable for medical image analysis, which often involves large-scale datasets and high-resolution images [18]. Figure 2 presents the layered architecture of the EfficientNet-B7 model; it contains the following layers.

- *Input Layer*: The model initiates the process by resizing the input image to dimensions of (224×224) pixels. This achieves a balance between maintaining sufficient detail and ensuring computational efficiency. The pixel values are normalized to facilitate smoother model training [18–20].
- *Initial Convolution Layer*: The model extracts fundamental patterns like edges and textures. The implementation employs a (3×3) filter size, characterized by its compactness and efficiency, alongside a stride of 2 to expedite the reduction of the image's dimensions. This conserves computational resources while retaining essential information.
- *MBConv Blocks (Multiple Blocks)*: The blocks serve as the foundational components of EfficientNet-B7. MBConv, or Mobile Inverted Bottleneck Convolution, is a specialised form of convolution employed in this context. This operation efficiently decomposes image features layer-wise while maintaining low computational costs. Following each block, a mechanism known as Squeeze-and-Excitation (SE) is employed, which enhances the model's ability to concentrate on the most significant aspects of the image through the recalibration of feature maps [21].
- *Squeeze-and-Excitation*: Consider this the model taking a step back to determine which features are most important. SE layers emphasise important features while minimising less relevant ones. It's similar to focusing on the most critical details. Using Squeeze, each feature map receives global average pooling, resulting in a scalar value representing each channel as presented by Eq. 1. Where Z_c : squeezed value for channel c, H, W: Weight, Width.

$$Z_c = \left[\frac{1}{H \times W} \right] \left(\sum_{i=1}^H \sum_{j=1}^W X_{i,j,c} \right) \quad (1)$$

In the extraction operation, a fully connected layer is used to learn channel-wise dependencies, followed by sigmoid activation to normalise the features in Eq. 2. Where S_c : Scale to feature, W_1 and W_2 : Learnable weights, σ : Sigmoid Function

$$S_c = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot Z_c)) \quad (2)$$

- *Global Average Pooling*: GAP simplifies things further by averaging the feature map's spatial dimensions after extracting all of the key features. This converts the feature map into a single 1D vector highlighting the most critical information.
- *Fully Connected Layer*: At this point, the extracted features are fed into a dense layer that holds everything together. This is where the model begins making predictions by mapping the features to the various classes in your task.
- *SoftMax Layer*: The output goes through the SoftMax function, which turns the raw numbers into probabilities. These probabilities show how confident the model is about each class, for example, the fibrosis stages.
- *Activation Functions*: EfficientNet uses the ‘Swish activation function’ to improve training performance over traditional ReLU, as presented in Eq. 3, where $\sigma(x)$: Sigmoid Function.

$$\text{Swish } (x) = [x \cdot \sigma(x)] \quad (3)$$

EfficientNet-B7 uses compound scaling, which increases the model's capacity by proportionally modifying its depth, width, and resolution [2]. The compound scaling method adjusts the network's depth, width, and resolution according to the subsequent Eq. (4). Where NM_{Scale} : New Model Scale, α , β and γ : Scaling coefficients

$$NM_{Scale} = \alpha^{\text{depth}}, \beta^{\text{width}}, \gamma^{\text{resolution}} \quad (4)$$

- *Depth*: The number of layers in the network is referred to as its depth, and as the network's depth increases, the model can understand more complex and hierarchical patterns.
- *Width*: The width indicates the number of channels in each layer, which improves the model's ability to extract diverse and complex features. The channels act as additional "eyes" to examine various aspects of the input.
- *Resolution*: In contrast, resolution determines the size of the input image; higher resolutions produce more detail, allowing the model to handle more complex data.

Unlike traditional methods, which scale a single dimension independently, EfficientNet-B7 scales all three dimensions simultaneously in a balanced manner using a compound coefficient. This approach improves model performance while lowering computational requirements, making it practical and efficient.

3.4.2 Dual-Attention Mechanism (Spatial + Channel)

The proposed hybrid deep learning model for liver fibrosis detection uses a Dual-Attention Mechanism (Spatial+Channel), which significantly improves network performance by focusing on the most relevant regions of liver images and critical feature channels. The mechanism aims to improve fibrosis detection precision by highlighting essential areas of the image and focusing on the most informative features in each channel. The following section describes how the proposed model employs the dual-attention mechanism, including equations and architectural details [2, 4, 5].

The Dual-Attention Mechanism consists of two primary components. Both attention mechanisms are employed sequentially to enhance the feature maps produced by the backbone (EfficientNet-B7) before their transmission to subsequent layers for classification or other functions [20, 22].

3.4.2.1 Spatial Attention

It produces a spatial attention map that emphasises specific spatial regions of the image that require focus for a more accurate diagnosis of liver fibrosis.

- *Global Average Pooling and Max Pooling*: The input feature map F , derived from the EfficientNet-B7 output with dimensions $(H \times W \times C)$, where H denotes height, W indicates width, and C shows the number of channels, executes Global Average Pooling (GAP) along with Global Max Pooling (GMP) operations. It can be measured by using Eq. 5.

$$F_{AVG} = GAP(F), F_{MAX} = GMP(F) \quad (5)$$

- *Concatenation*: The pooled outcomes from GAP and GMP are combined along the dimension of channels to generate a new feature map, utilising Eq. 6.

$$F_{CONCAT} = CONCAT(F_{AVG}, F_{MAX}) \quad (6)$$

- *Convolution to Generate Spatial Attention Map:* The concatenated feature map F_{concat} is subjected to a convolutional layer (typically a (1×1) convolution) to produce the spatial attention map $A_{SPATIAL}$. As presented by Eq. 7. Where σ : Sigmoid function, $W_{SPATIAL}$: convolutional weight matrix, \times : Convolution operation.

$$A_{SPATIAL} = \sigma(F_{CONCAT} \times W_{SPATIAL}) \quad (7)$$

- *Feature Map Refinement:* Spatial attention map to refine the spatial information, $A_{SPATIAL}$ is multiplied element-wise with the original feature map F . This focuses on the most relevant regions of the liver image, as presented by Eq. 8.

$$F_{REFINED} = A_{SPATIAL} \times F \quad (8)$$

3.4.2.2 Channel Attention It readjusts the significance of each feature channel to amplify the representation of pertinent features and diminish irrelevant ones. It facilitates the adjustment of the importance of each channel, enabling the model to concentrate on the more informative feature channels [9, 16, 17].

- *Global Pooling for Channel Attention:* Similar to spatial attention, Global GAP and GMP are utilised on the feature map F across the channel dimension as presented in Eq. 9.

$$F_{AVG} = GAP(F) \text{ and } F_{MAX} = GMP(F) \quad (9)$$

- *Channel Attention Map Generation:* The pooled feature vectors F_{AVG} and F_{MAX} , are processed through fully connected (FC) layers, which implement a squeeze-and-excitation operation to produce the channel attention map $A_{CHANNEL}$ as presented in Eq. 10. Where $W_{CHANNEL}$: weight matrix of the fully connected layers, \times : Matrix multiplication.

$$A_{CHANNEL} = \sigma(W_{CHANNEL}) \times (F_{AVG} + F_{MAX}) \quad (10)$$

- *Feature Map Refinement:* The channel attention map is a representation that highlights the significance of different channels in a given dataset, allowing for a focused analysis of the most relevant features. $A_{CHANNEL}$ is multiplied element-wise with the feature map F to adjust the channel-wise feature representations as presented by Eq. 11.

$$F_{FINAL} = (A_{CHANNEL}) \times (F) \quad (11)$$

3.4.3 Multi-scale Feature Extraction (Dilated Convolutions)

Multi-scale feature extraction is a method employed to obtain information across various spatial scales or resolutions. The proposed hybrid model for liver fibrosis detection utilises dilated convolutions to effectively capture intricate details and contextual information from liver images across multiple scales. This method is especially advantageous in medical imaging, where it is crucial to detect patterns across different levels of detail while maintaining spatial resolution [3–6, 8].

The dilated convolutional layer modifies the standard convolution by incorporating a dilation factor d , which regulates the spacing between the kernel elements. This facilitates the expansion of the receptive field while maintaining the same number of parameters and computational cost, thereby allowing the network to capture features from a broader context.

- *Dilated Convolution Operation:* The dilated convolution operation can be defined by Eq. 12.

$$Y(i, j) = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} (x((i + m \times d), (j + n \times d)) \times (W(m, n))) \quad (12)$$

- *Advantages in the Context of Liver Fibrosis Detection:* Dilated convolutions are essential for capturing larger spatial patterns in images, including the unique shapes and structures associated with liver fibrosis, while avoiding the need for larger filters or extra pooling layers. Increasing the receptive field through dilated convolutions enables the model to concentrate on broader contextual information, which is crucial for interpreting complex medical images [6]. This technique concurrently maintains the spatial resolution of the image, which is essential in medical image analysis, as it ensures the retention of fine details necessary for accurate diagnosis. The ability to capture large-scale features while maintaining high-resolution information renders dilated convolutions highly effective for tasks such as liver fibrosis detection, where it is essential to identify subtle patterns in the images for accurate classification.

3.4.4 Fully Connected (Dense) Layers

The fully connected, dense layers serve as essential elements within the proposed hybrid deep learning model. The layers in question integrate features extracted by the convolutional layers, encompassing dilated convolutions and attention mechanisms, to produce final predictions regarding the severity of liver fibrosis. The layers are essential for feature aggregation and classification [16]. The output z of a fully connected layer is computed according to Eq. 13. Where W : Weight Matrix, x : Input vector, b : bias, Z : Outcome using an activation function.

$$Z = (W \times x) + b \quad (13)$$

The fully connected layers within the proposed model integrate the features obtained from preceding layers and execute the final decision-making process, which determines the classification of the liver fibrosis stage. Implementing these layers, typically utilising a Multi-layer Perceptron (MLP), allows the model to perform non-linear transformations on the extracted features [18]. The MLP improves the model's capacity to identify intricate relationships within the data, facilitating more precise predictions concerning the severity of liver fibrosis based on the features acquired during the convolutional and attention-based processing phases.

3.4.5 Output Layer (SoftMax)

The model's output layer is tasked with generating the final prediction regarding the stage of liver fibrosis. The output layer generally employs the SoftMax activation function to generate a probability distribution across the potential fibrosis stages [15]. The model generates a vector of probabilities, with each value representing the likelihood of the input image being classified into a particular fibrosis stage, ranging from Stage 0 to Stage 4.

- *SoftMax Activation:* The SoftMax function converts raw logits (scores) into a probability distribution by normalising output values. The SoftMax function is defined for an input vector Z of size c , where c Represents the number of classes, as outlined in Eq. 14. Where Z_i : i th logit in the vector Z .

$$\text{SoftMax}(Z_i) = \frac{e^{Z_i}}{\sum_{j=1}^c e^{Z_j}} \quad (14)$$

- *Final Classification:* After applying the SoftMax activation function, the model produces the distribution of probabilities for each fibrosis stage, which ranges from Stage 0 to Stage 4. The predicted label for the liver

fibrosis stage is identified by choosing the stage with the most significant probability, which can be measured by using Eq. 15. Where \hat{Y} : predicted fibrosis stage, P: probability vector output by the softmax function.

$$\hat{Y} = \arg \max (P) \quad (15)$$

3.5 Algorithm for Proposed Hybrid Model

The complete algorithm for the Proposed Hybrid Model is presented in Algorithm 1 [9, 16–20].

Algorithm 1 Algorithm for liver fibrosis detection using hybrid model (EfficientNet-B7 with dual-attention mechanism and dilated convolutions).

<p><i>Input:</i></p> <ul style="list-style-type: none"> -A liver image (size: 224x224). -Image is normalised to ensure pixel values range between 0 and 1.
<p><i>Output:</i></p> <ul style="list-style-type: none"> -Predicted fibrosis stage of the liver image (Stage 0 to Stage 4).
<p><i>Step 1. Image Preprocessing</i></p> <p><i>1.1 Resize Input Image:</i></p> <ul style="list-style-type: none"> • The input image is resized to a fixed size of 224x224 pixels to standardise input dimensions. • Image is normalised by scaling pixel values between 0 and 1. • Function: <code>resized_image = resize_image(input_image, target_size=(224, 224)) normalized_image = normalize_image(resized_image)</code>
<p><i>Step 2. Feature Extraction using EfficientNet-B7</i></p> <p><i>2.1 Initial Convolution Layer:</i></p> <ul style="list-style-type: none"> • A 3x3 convolution is applied with a stride of 2 to reduce image dimensions while retaining critical features such as edges and textures. • Function: <code>features = initial_convolution(input_image)</code> <p><i>2.2 MBConv Blocks with Squeeze-and-Excitation:</i></p> <ul style="list-style-type: none"> • Multiple MBConv blocks are applied, which help in efficient feature extraction with low computational cost. • After each block, Squeeze-and-Excitation (SE) recalibrates feature maps by emphasising the most important channels (Equations 1 and 2). • Function: <code>extracted_features = mbconv_blocks(features) refined_features = squeeze_and_excitation(extracted_features)</code> <p><i>2.3 Swish Activation:</i></p> <ul style="list-style-type: none"> • The Swish activation function is used to enhance training performance over traditional ReLU. • Function: <code>activated_features = swish_activation(refined_features)</code>
<p><i>Step 3. Dual-Attention Mechanism</i></p> <p><i>3.1 Spatial Attention:</i></p> <ul style="list-style-type: none"> • Global Average Pooling and Global Max Pooling are applied to the feature map to create pooled representations (Equations 5 and 6).

- The pooled feature maps are concatenated and passed through a convolutional layer to produce a spatial attention map (Equation 7).
- The spatial attention map is multiplied element-wise with the feature map to refine spatial focus (Equation 8).
- Function:
 $\text{spatial_attention_map} = \text{spatial_attention}(\text{features})$
 $\text{refined_spatial_features} = \text{refine_with_attention}(\text{features}, \text{spatial_attention_map})$

3.2 Channel Attention:

- Similarly, Global GAP and GMP are applied to the feature map to produce channel-wise pooled features (Equation 9).
- These features are passed through fully connected layers to generate a channel attention map (Equation 10).
- The channel attention map is multiplied element-wise with the feature map to adjust channel-wise features (Equation 11).
- Function:
 $\text{channel_attention_map} = \text{channel_attention}(\text{refined_spatial_features})$
 $\text{final_features} = \text{refine_with_attention}(\text{refined_spatial_features}, \text{channel_attention_map})$

Step 4. Multi-Scale Feature Extraction using Dilated Convolutions

4.1 Dilated Convolutions:

- Dilated convolutions are applied to capture multi-scale patterns by increasing the receptive field without adding extra parameters (Equation 12).
- Function:
 $\text{dilated_features} = \text{dilated_convolution}(\text{final_features}, \text{dilation_rate}=2)$

Step 5. Fully Connected (Dense) Layers

5.1 Feature Aggregation:

- The extracted and refined features are passed through fully connected layers to integrate them and perform non-linear transformations to predict the fibrosis stage (Equation 13).
- Function:
 $\text{dense_output} = \text{fully_connected_layers}(\text{dilated_features})$

Step 6. Output Layer (SoftMax)

6.1 SoftMax Activation:

- The output of the dense layer is passed through a SoftMax activation to obtain the probability distribution across the different fibrosis stages (Equation 14).
- Function:
 $\text{probabilities} = \text{softmax_activation}(\text{dense_output})$

6.2 Final Classification:

- The predicted fibrosis stage is identified by selecting the class with the highest probability (Equation 15).
- Function:
 $\text{predicted_stage} = \text{final_classification}(\text{probabilities})$

3.6 Training Parameters

Training Parameters indicate the configurations and hyperparameters that control the model's learning process, encompassing the learning rate, batch size, optimisation, and number of epochs. The specified parameters are essential for facilitating efficient model convergence while preventoptimisationfitting and underfitting throughout the training process. Table 4 presents details of the training parameters for the proposed hybrid model.

Table 4 Training parameter details for proposed hybrid model

Parameter	Description	Value
Learning rate	Controls the rate at which the model weights are updated during training	0.0001
Batch size	Number of training samples used per update of the model weights	32
Optimizer	An optimization algorithm was used to minimise the loss function	Adam
Epochs	Number of times the entire dataset is passed through the model during training	100
Loss function	Function used to calculate the difference between predicted and true values	Categorical cross-entropy
Dropout rate	The fraction of the input units is randomly set to zero during training	0.5
Weight initialization	The method used to initialise the weights in the neural network	He initialization method
Early stopping	Method to stop training when validation loss stops improving	After 5 consecutive epochs without improvement
β_1 (adam optimiser)	The exponential decay rate for the first-moment estimate	0.9
β_2 (adam optimiser)	The exponential decay rate for the second moment is estimated	0.999
ϵ (adam optimiser)	Small constant to prevent division by zero during optimisation	10 -7

3.7 Evaluation Metrics

Evaluation metrics assess the performance of the proposed hybrid model and existing VGGNet, DenseNet, ResNet, and U-Net for liver fibrosis detection. Standard metrics encompass accuracy, precision, recall, F1-score, and area under the curve (AUC). These metrics are critical for evaluating model performance, particularly in the context of imbalanced medical datasets [2, 7, 14, 15, 23]. Here TP: True Positives, FP: False Positives, TN: True Negatives, FN: False Negatives.

3.7.1 Accuracy

Calculates the ratio of accurately classified instances compared to the overall total of cases as presented by Eq. 16.

$$\text{Accuracy} = \frac{[TP + TN]}{[TP + FP + TN + FN]} \quad (16)$$

3.7.2 Precision

Represents the ratio of true positive predictions to the total number of positive predictions generated by the model, as presented by Eq. 17.

$$\text{Precision} = \frac{[TP]}{[TP + FP]} \quad (17)$$

3.7.3 Recall

Calculates the ratio of true positive cases accurately recognised by the model as presented by Eq. 18.

$$\text{Recall} = \frac{[TP]}{[TP + FN]} \quad (18)$$

3.7.4 F1-Score

The harmonic mean serves as a metric that balances precision and recall, offering a combined measure of both performance indicators as presented by Eq. 19.

$$F1 - Score = 2 \times \frac{[Precision \times Recall]}{[Precision + Recall]} \quad (19)$$

4 Experimental Results

The proposed hybrid model for liver fibrosis detection was evaluated using the Liver Fibrosis Dataset (LiverFib), which contains annotated liver imaging scans across different fibrosis stages. The dataset was split into 70% for training, 15% for validation, and 15% for testing. The model's performance was compared to conventional CNN architectures (VGGNet, DenseNet, ResNet, and U-Net) and attention-based models using metrics like accuracy, precision, recall, F1-score, and AUC. Python programming and deep learning libraries were utilised for the implementation [7, 17, 24].

4.1 Scenario 1: Classification Performance Comparison

Comparison of Classification Performance in Scenario 1. This scenario examines the classification accuracy and capacity of each model to identify stages of liver fibrosis accurately. The hybrid model presented demonstrates superior performance compared to conventional CNN architectures and attention-based models, as evidenced by metrics such as accuracy, precision, recall, F1-score, and AUC. Table 5 presents the classification performance comparison for existing and proposed models.

Figures 3, 4, 5, 6, and 7 present the confusion matrix results for the proposed hybrid model alongside established deep learning models, including VGGNet, DenseNet, ResNet, and U-Net. These matrices are generated from the classification of liver fibrosis stages (0–4) and juxtapose actual and predicted values throughout the training process. The confusion matrices elucidate each model's performance by illustrating the distribution of true positives, false positives, true negatives, and false negatives at each liver fibrosis stage. Figure 3 depicts DenseNet's confusion matrix, subsequently followed by ResNet in Fig. 4, U-Net in Fig. 5, VGG-Net in Fig. 6, and the proposed hybrid model in Fig. 7. In combination, this data offers a thorough comparison of each model's efficacy in classifying various stages of liver fibrosis, with the proposed hybrid model striving to enhance accuracy relative to existing models.

Figure 8 illustrates the comparative performance of the proposed hybrid model relative to existing models. Table 5 delineates essential classification metrics accuracy, precision, recall, F1-score, and AUC for the proposed hybrid model alongside VGGNet, DenseNet, ResNet, and U-Net. The proposed hybrid model exhibits superior overall performance, exceeding all other models' accuracy and additional metrics. Although VGGNet, DenseNet, and ResNet demonstrate commendable performance, they remain inferior to the hybrid model. Conversely, U-Net exhibits the lowest efficacy. This shows that the suggested hybrid model provides a more dependable and precise classification methodology.

Table 5 Classification performance comparison for existing and proposed models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
Proposed hybrid model	98.5	97.2	96.5	96.8	0.98
VGGNet	94.3	92.1	89.8	90.9	0.94
DenseNet	95.8	93.6	91.2	92.4	0.95
ResNet	96.0	94.3	92.5	93.4	0.96
U-Net	92.5	90.2	87.4	88.7	0.91

		Confusion Matrix for DenseNet				
		Stage 0 (Healthy Liver)	Stage 1 (Mild Fibrosis)	Stage 2 (Moderate Fibrosis)	Stage 3 (Severe Fibrosis)	Stage 4 (Cirrhosis)
True	Stage 0 (Healthy Liver)	1300	60	50	40	50
	Stage 1 (Mild Fibrosis)	90	2000	60	30	40
Stage 2 (Moderate Fibrosis)	60	100	2300	50	80	
Stage 3 (Severe Fibrosis)	50	40	80	1550	80	
Stage 4 (Cirrhosis)	40	50	60	50	1300	
		Stage 0 (Healthy Liver)	Stage 1 (Mild Fibrosis)	Stage 2 (Moderate Fibrosis)	Stage 3 (Severe Fibrosis)	Stage 4 (Cirrhosis)
		Predicted				

Fig. 3 Confusion matrix for DenseNet

4.2 Scenario 2: Results with and without Attention Mechanism

This scenario involves a comparative analysis of the proposed hybrid model's performance, both with and without the attention mechanism, against various existing models, including VGGNet, DenseNet, ResNet, and U-Net, utilising the Liver Fibrosis Dataset (LiverFib). The results are organised according to fibrosis stages, ranging from Stage 0 to Stage 4, illustrating the enhancement of model performance through the application of the attention mechanism.

Table 6 delineates the effectiveness of various models in detecting stages of liver fibrosis, with and without attention mechanisms. The results reveal that the proposed hybrid model integrating attention outperforms all other models at each stage, achieving the highest accuracy, particularly in advanced stages like Stage 4 (Cirrhosis). The hybrid model with attention demonstrates significantly enhanced accuracy, underscoring the role of attention in improving performance. Similarly, Fig. 9 illustrates the performance of the proposed hybrid model with and without attention mechanisms. It distinctly indicates that the variant incorporating attention exhibits significantly superior performance, underscoring the critical role of attention mechanisms in enhancing classification accuracy, particularly during the more challenging phases of liver fibrosis.

Table 7 and Fig. 10 illustrate a comparative analysis of evaluation metrics for each model, both with and without attention, encompassing accuracy, precision, recall, F1-score, and AUC. The suggested hybrid model utilising attention attains superior performance across all metrics, achieving an accuracy of 98.5%, precision of 97.2%, recall of 96.5%, F1-score of 96.8%, and AUC of 0.98. In contrast, VGGNet, DenseNet, ResNet, and U-Net exhibit inferior performance, with accuracy varying from 91.4% for VGGNet to 93.0% for U-Net. The attention mechanism markedly enhances the efficacy of the proposed hybrid model, yielding considerable improvements in precision, recall, and F1-score. Figure 10 illustrates these evaluation metrics, emphasising the enhanced performance of the proposed hybrid model with attention relative to other models. This comparison illustrates the essential function of attention in improving model precision and overall efficacy in categorising liver fibrosis stages.

		Confusion Matrix for ResNet				
		Stage 0 (Healthy Liver) -	Stage 1 (Mild Fibrosis) -	Stage 2 (Moderate Fibrosis) -	Stage 3 (Severe Fibrosis) -	Stage 4 (Cirrhosis) -
True	Stage 0 (Healthy Liver) -	1320	50	60	50	30
	Stage 1 (Mild Fibrosis) -	80	1900	70	40	50
Stage 2 (Moderate Fibrosis) -	50	90	2200	60	90	
Stage 3 (Severe Fibrosis) -	50	30	90	1600	60	
Stage 4 (Cirrhosis) -	40	50	70	50	1300	
Stage 0 (Healthy Liver) -			Stage 1 (Mild Fibrosis) -		Stage 2 (Moderate Fibrosis) -	
Stage 3 (Severe Fibrosis) -			Stage 4 (Cirrhosis) -			
		Predicted				

Fig. 4 Confusion matrix for ResNet

4.3 Scenario 3: Impact of Data Pre-processing

This scenario involves an evaluation of the impact of data pre-processing on the performance metrics of existing models, including VGGNet, DenseNet, ResNet, and U-Net, as well as the proposed hybrid model for liver fibrosis detection. Data pre-processing encompasses procedures such as image normalisation, resizing, augmentation, and noise reduction. These processes are essential for enhancing model performance by improving the quality of input data. The analysis focuses on the influence of data pre-processing on classification metrics, including accuracy, precision, recall, F1-score, and AUC, across different stages of fibrosis, ranging from 0 to 4. The results are compared for the proposed hybrid model alongside existing models, both with and without pre-processing.

Table 8 and Fig. 11 illustrate the performance of different models in classifying stages of liver fibrosis, with and without data pre-processing. The results indicate a significant enhancement in performance with the application of pre-processing. The proposed hybrid model attains an accuracy of 91.1% without pre-processing, which significantly rises to 98.5% with pre-processing, underscoring the critical role of pre-processing in improving model performance. Other models also exhibit enhancements with pre-processing, including VGGNet, which improves from 85.6 to 92.3%, and DenseNet, which ascends from 86.8 to 92.9%. The upgrades are uniform across all stages of fibrosis, particularly in Stage 0 (No Fibrosis) and Stage 1 (Mild Fibrosis), where pre-processing contributes to improved classification accuracy.

Table 9 presents the evaluation metrics for each model, both with and without pre-processing. The suggested hybrid model with pre-processing attains superior performance across all metrics, exhibiting an accuracy of 98.5%, precision of 97.2%, recall of 96.5%, F1-score of 96.8%, and AUC of 0.98. In contrast, models such as

		Confusion Matrix for U-Net				
		Stage 0 (Healthy Liver) -	Stage 1 (Mild Fibrosis) -	Stage 2 (Moderate Fibrosis) -	Stage 3 (Severe Fibrosis) -	Stage 4 (Cirrhosis) -
True	Stage 0 (Healthy Liver) -	1300	60	70	40	30
	Stage 1 (Mild Fibrosis) -	100	2000	50	40	40
Stage 2 (Moderate Fibrosis) -	60	90	2250	50	80	
Stage 3 (Severe Fibrosis) -	50	40	80	1600	60	
Stage 4 (Cirrhosis) -	40	50	60	50	1300	

Fig. 5 Confusion matrix for U-Net

VGGNet, DenseNet, ResNet, and U-Net demonstrate significant enhancements with pre-processing, yet still do not match the performance of the hybrid model. For instance, U-Net's accuracy rises from 88.0 to 93.4% with pre-processing, while ResNet's accuracy increases from 87.5 to 93.1%. Pre-processing is essential for improving the accuracy and efficacy of each model, with the proposed hybrid model deriving the most significant advantage from this technique.

4.4 Scenario 4: Accuracy vs. Loss for Training, Test and Validation

In Scenario 4, the proposed hybrid model's performance is evaluated across three datasets: training, testing, and validation. This assessment focuses on both accuracy and loss metrics over 100 epochs. The model demonstrates a steady enhancement in accuracy alongside a decrease in loss throughout all datasets as training advances.

Figures 12 and 13 illustrate the simulation outcomes for Scenario 4, wherein the efficacy of the proposed hybrid model is assessed across training, testing, and validation datasets over 100 epochs. The graph demonstrates a consistent increase in the model's accuracy alongside a decrease in loss, signifying effective learning throughout the training process. The proposed model exhibits consistent enhancement in both accuracy and loss across all datasets, indicating its capacity to fit the training data effectively while generalising to unseen data. This trend demonstrates the model's practical learning and adaptation, underscoring its robust performance across all training, testing, and validation stages.

4.5 Scenario 5: Interpretability Through Attention Mechanism

Scenario 5 illustrates the concept of Interpretability via the Attention Mechanism. This scenario involves evaluating models' capacity to identify significant areas within liver images, explicitly focusing on fibrosis-associated regions. Interpretability is essential for clinical decision-making, as it enables clinicians to comprehend the rationale behind the model's specific classification.

		Confusion Matrix for VGGNet				
		Stage 0 (Healthy Liver) -	Stage 1 (Mild Fibrosis) -	Stage 2 (Moderate Fibrosis) -	Stage 3 (Severe Fibrosis) -	Stage 4 (Cirrhosis) -
True	Stage 0 (Healthy Liver) -	1300	70	60	40	30
		100	1900	50	30	40
Stage 2 (Moderate Fibrosis) -	50	80	2200	60	90	
Stage 3 (Severe Fibrosis) -	40	40	100	1600	60	
Stage 4 (Cirrhosis) -	30	50	60	50	1300	
		Stage 0 (Healthy Liver) -	Stage 1 (Mild Fibrosis) -	Stage 2 (Moderate Fibrosis) -	Stage 3 (Severe Fibrosis) -	Stage 4 (Cirrhosis) -
		Predicted				

Fig. 6 Confusion matrix for VGG-Net

Table 10 displays the results of the interpretability analysis utilising the attention mechanism across various models. The suggested hybrid model exhibits significant interpretability, concentrating specifically on fibrosis-impacted regions, attaining a 92% overlap in the attention map and a 0.95 Grad-CAM score. In contrast, models such as VGGNet, DenseNet, and ResNet exhibit moderate interpretability, with attention map overlaps between 70 and 74% and Grad-CAM scores ranging from 0.80 to 0.82, suggesting they capture general features without specifically targeting fibrosis regions. U-Net, primarily intended for segmentation, exhibits reduced interpretability for classification tasks, evidenced by an attention map overlap of merely 60% and a Grad-CAM score of 0.65. The proposed hybrid model effectively concentrates on critical fibrosis regions, enhancing its predictive accuracy.

4.6 Scenario 6: Computational Efficiency

Scenario 6 describes the analysis of computational efficiency. This scenario investigates each model's computational characteristics, including training time, inference velocity, and memory consumption. Medical image analysis frequently necessitates handling large datasets and implementing real-time processing, making computational efficiency an important consideration.

Table 11 delineates a computational efficiency analysis juxtaposing the proposed hybrid model with existing models regarding training duration, inference time per image, and memory consumption. The proposed hybrid model is distinguished by a training duration of 12 h, which is less than that of most other models, including

		Confusion Matrix for Proposed Hybrid Model				
		Stage 0 (Healthy Liver) -	Stage 1 (Mild Fibrosis) -	Stage 2 (Moderate Fibrosis) -	Stage 3 (Severe Fibrosis) -	Stage 4 (Cirrhosis) -
True	Stage 0 (Healthy Liver) -	1300	50	50	40	30
		80	1900	50	30	40
Stage 2 (Moderate Fibrosis) -	60	100	2200	50	90	
Stage 3 (Severe Fibrosis) -	40	30	80	1600	50	
Stage 4 (Cirrhosis) -	30	40	60	50	1300	
		Stage 0 (Healthy Liver) -	Stage 1 (Mild Fibrosis) -	Stage 2 (Moderate Fibrosis) -	Stage 3 (Severe Fibrosis) -	Stage 4 (Cirrhosis) -
		Predicted				

Fig. 7 Confusion matrix for proposed hybrid model

DenseNet (18 h) and U-Net (20 h). Moreover, its inference time per image is 0.24 s, surpassing that of VGGNet (0.30 s) and U-Net (0.35 s). The hybrid model utilises 8.5 GB of memory, rendering it more memory-efficient than DenseNet (12 GB) and U-Net (11 GB). The proposed hybrid model exhibits enhanced computational efficiency, optimising rapid performance while minimising resource consumption.

4.7 Ablation Analysis

Ablation analysis is an essential method for determining the impact of individual components within a model. This context explains how each element of the proposed hybrid model for detecting liver fibrosis contributes to the overall performance. The systematic removal or modification of specific components allows for the measurement of their impact on critical metrics such as accuracy, precision, recall, F1-score, and AUC. This analysis provides vital insights into the significance of each module, allowing the model to be refined and optimised for improved performance.

Table 12 displays the performance comparison of the ablation analysis for the proposed model across various variants. The comprehensive proposed hybrid model attains optimal performance, exhibiting an accuracy of 98.5%, precision of 97.2%, recall of 96.5%, F1-score of 96.8%, and AUC of 0.98. In contrast, the performance declines across different scenarios, including Scenario 1 (No Attention), with an accuracy of 93.5%, and Scenario 2 (VGGNet Backbone), with an accuracy of 94.8%. Scenarios 3 (No Dual-Attention) and 4 (Standard Convolutions) exhibit diminished performance, achieving accuracies of 96.1% and 95.3%, respectively. This comparison

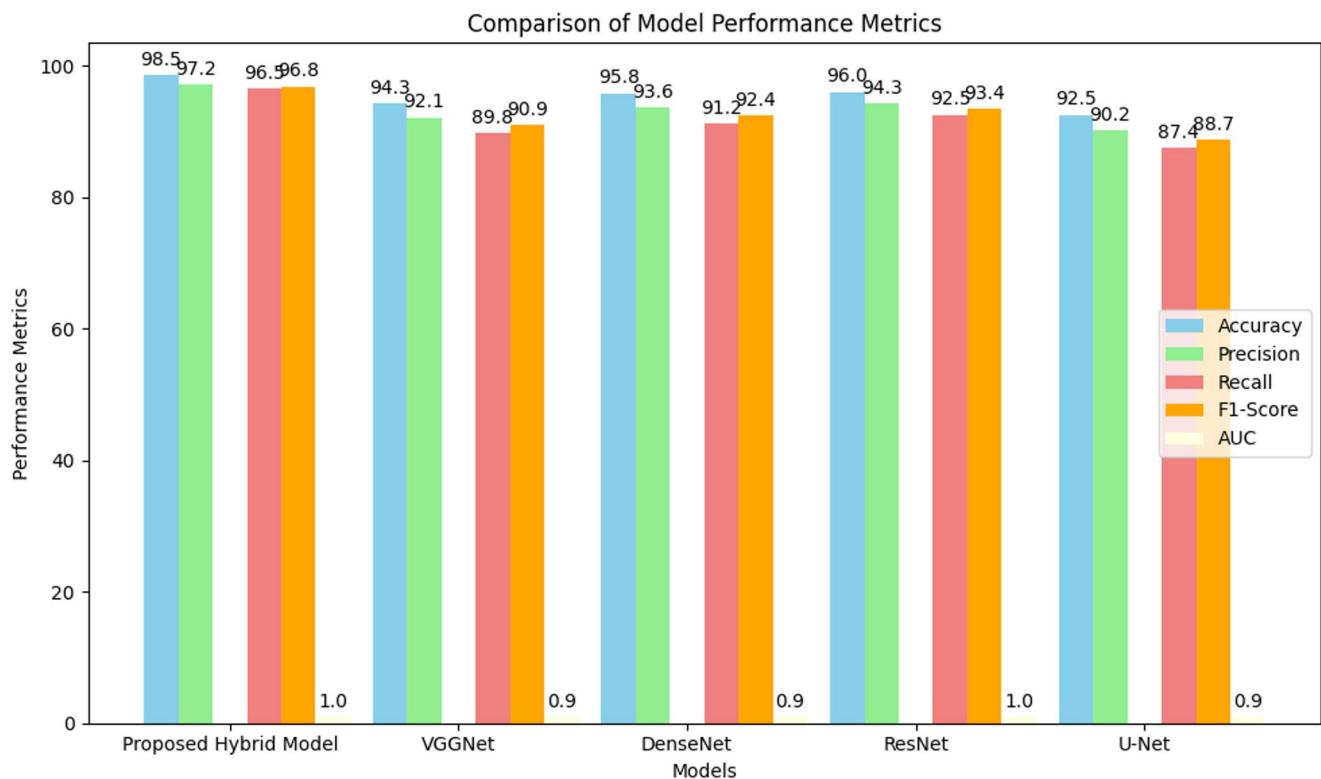


Fig. 8 Comparisons of model performance for proposed hybrid model

Table 6 Performance comparison results based on liver fibrosis stages (with and without attention)

Fibrosis stage	VGGNet (%)	DenseNet (%)	ResNet (%)	U-Net (%)	Proposed hybrid model (without attention) (%)	Proposed hybrid model (with attention) (%)
Stage 0 (no fibrosis)	93.5	94.3	94.9	95.0	96.5	98.5
Stage 1 (mild fibrosis)	90.7	91.6	92.1	92.5	93.0	97.3
Stage 2 (moderate fibrosis)	91.4	92.5	93.0	92.8	94.0	98.0
Stage 3 (severe fibrosis)	88.3	90.2	90.8	91.5	92.0	97.8
Stage 4 (cirrhosis)	91.2	92.3	92.9	93.3	93.7	98.2
Overall accuracy	91.4	92.2	92.7	93.0	94.0	98.5

underscores the substantial contributions of attention mechanisms, hybrid architecture, and advanced convolution techniques in attaining the exceptional performance of the proposed model.

4.8 Comparisons with State-of-the-Art Methods

This section provides a comparative analysis of the proposed hybrid model for liver fibrosis detection against various advanced methodologies currently employed in the field. The proposed model leads to improved classification accuracy and interpretability, particularly in identifying critical regions in liver imaging scans. A detailed comparison is provided below, based on recent publications employing hybrid methodologies for the detection of liver fibrosis.

As shown in Table 13 and Fig. 14, the proposed Hybrid Model outperforms contemporary transformer-based models (i.e., ViT or Swin Transformer) on all evaluation metrics comprising accuracy, precision, recall, F1-score and AUC. Especially, the attention mechanism, which is utilised in our proposed model, can focus well on the

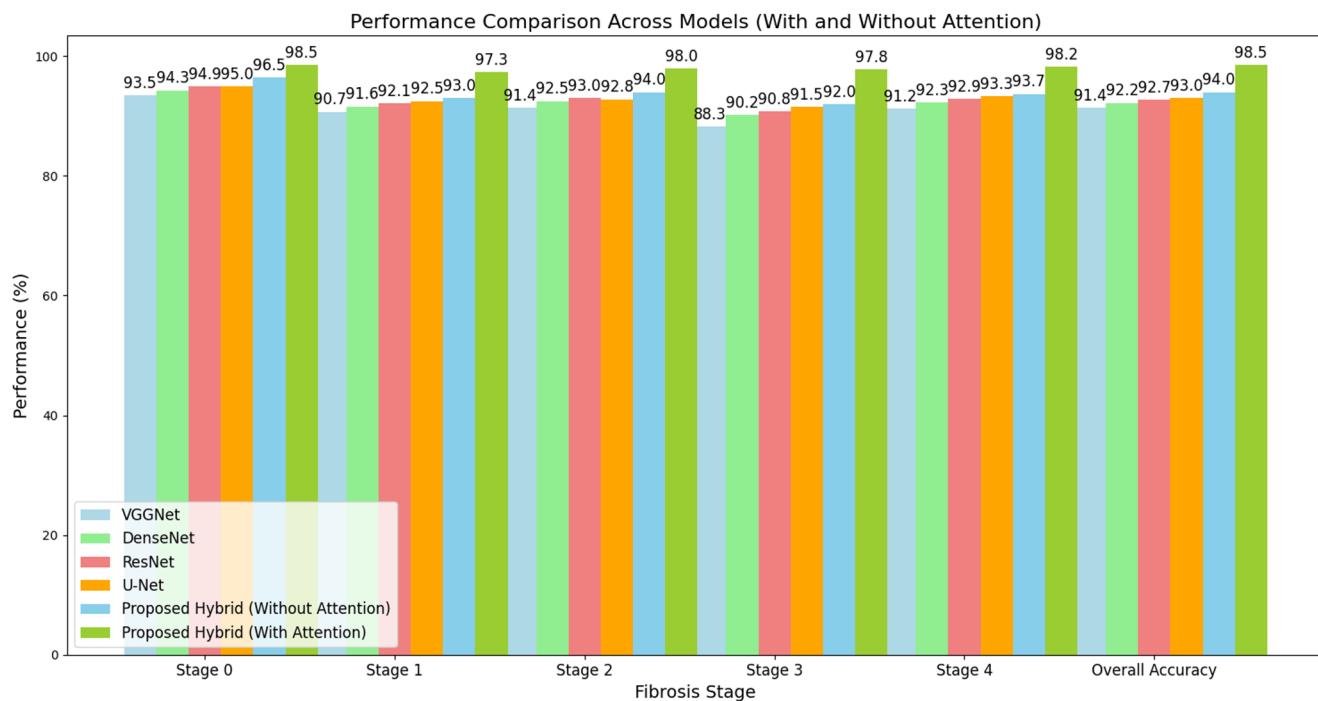


Fig. 9 Comparisons of model performance for proposed hybrid model

Table 7 Evaluation metrics for each model (with and without attention)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
VGGNet	91.4	89.6	90.2	89.9	0.93
DenseNet	92.2	90.4	91.5	90.9	0.94
ResNet	92.7	91.0	92.3	91.6	0.95
U-Net	93.0	91.5	92.6	92.0	0.96
Proposed hybrid model (without attention)	94.0	92.5	93.1	92.8	0.96
Proposed hybrid model (with attention)	98.5	97.2	96.5	96.8	0.98

region affected by liver fibrosis, thereby improving diagnostic performance for the model. However, compared to ViT and Swin Transformer that learn from whole-slide images, the Proposed Hybrid Model achieve a higher overall accuracy (98.5%) with superior clinical interpretation by smartly targeting the regions affected by fibrosis at high precision and recall. This observation makes the model not only accurate but also more interpretable -an issue that is important in clinical applications where explanation of decisions made by a model can be vital.

4.9 Statistical Analysis

To further validate the robustness and statistical significance of performance results achieved by the proposed hybrid model, we conducted k-fold cross-validation ($k=5$) as well as t-test procedures. These inferential statistics are so important to notice that the enhancement in performance was correct and not due to overfitting or randomness.

4.9.1 k-fold Cross Validation

We evaluated the portability of the hybrid proposed model using fivefold cross-validation. This approach divides the data into five chunks of equal size. The model is trained on four parts of the data and tested on the remaining

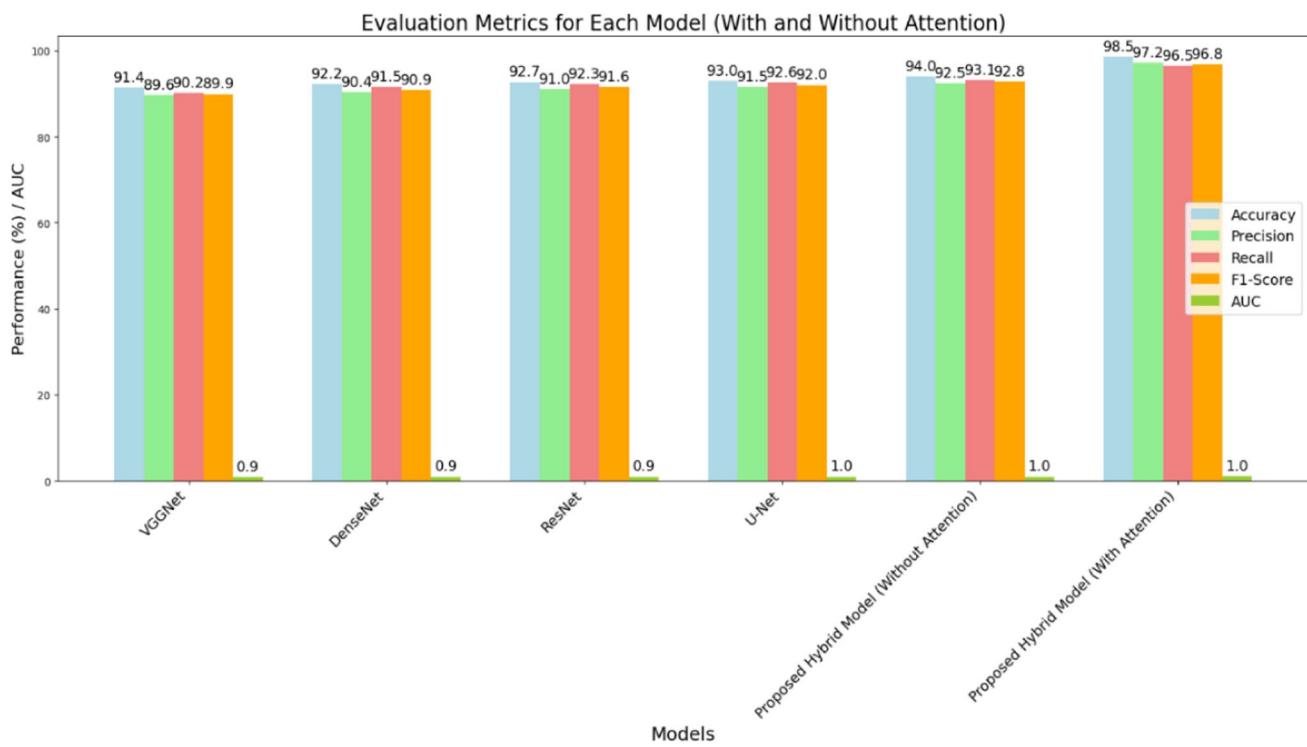


Fig. 10 Evaluation metrics for each model (with and without attention)

part in all folds. This procedure is performed five times, holding each fold once as a test. To secure an estimate of performance, the performance measures (accuracy, precision, recall, F1-score and AUC) were averaged over all fivefold cross-validation.

The results of k-fold cross-validation (Table 14 and Fig. 15) show that the hybrid model is robust, with accuracy rates of between 98.2 and 98.6%, with an average accuracy rate of 98.4%. Moreover, precision, recall, F1-score and AUC were also stable in all folds, illustrating that the model is reliable. These results indicate that the hybrid model still performed well on the various splits of data and didn't show clear signs of overfitting, suggesting that such a model can be generalised and widely applied to actual problems.

4.9.2 t-Test Analysis

To demonstrate the statistical significance of our hybrid model compared to alternative state-of-the-art models, a paired t-test was performed. A t-test was conducted to compare the performance of the hybrid model with VGGNet, DenseNet, ResNet, U-Net, ViT and Swin Transformer in terms of accuracy, precision, recall, F1-score and AUC. The null hypothesis, for the t-test procedure, was that there is no significant difference in performance between different models, whereas the alternative hypothesis states a significantly improvement achieved by the proposed hybrid model compared to other methods. Table 15 presents the results of the paired t-tests.

- **Methodology for t-Test**

- o *Data:* The evaluation data (accuracy, precision, recall, F1-score and AUC) were collected from the fivefold cross-validation results.

- **Assumptions:**

Table 8 Performance comparison results based on liver fibrosis stages (with and without data pre-processing)

Fibrosis stage	VGGNet (without pre-processing) (%)	VGGNet (with pre-processing) (%)	DenseNet (without pre-processing) (%)	DenseNet (with pre-processing) (%)	ResNet (without pre-processing) (%)	ResNet (with pre-processing) (%)	U-Net (without pre-processing) (%)	U-Net (with pre-processing) (%)	Proposed hybrid model (without pre-processing) (%)	Proposed hybrid model (with pre-processing) (%)
Stage 0 (no fibrosis)	88.5	94.0	89.2	94.6	89.8	95.1	90.5	95.3	92.1	98.5
Stage 1 (mild fibrosis)	85.3	91.2	86.5	92.0	87.0	92.5	88.0	93.0	91.0	97.3
Stage 2 (moderate fibrosis)	86.0	92.5	87.0	92.8	88.5	93.0	89.0	93.6	92.0	98.0
Stage 3 (severe fibrosis)	83.2	90.4	84.3	91.7	85.0	92.1	86.5	92.5	89.5	97.8
Stage 4 (cirrhosis)	85.1	91.5	86.0	92.4	87.2	92.9	88.2	93.2	91.0	98.2
Overall accuracy	85.6	92.3	86.8	92.9	87.5	93.1	88.0	93.4	91.1	98.5

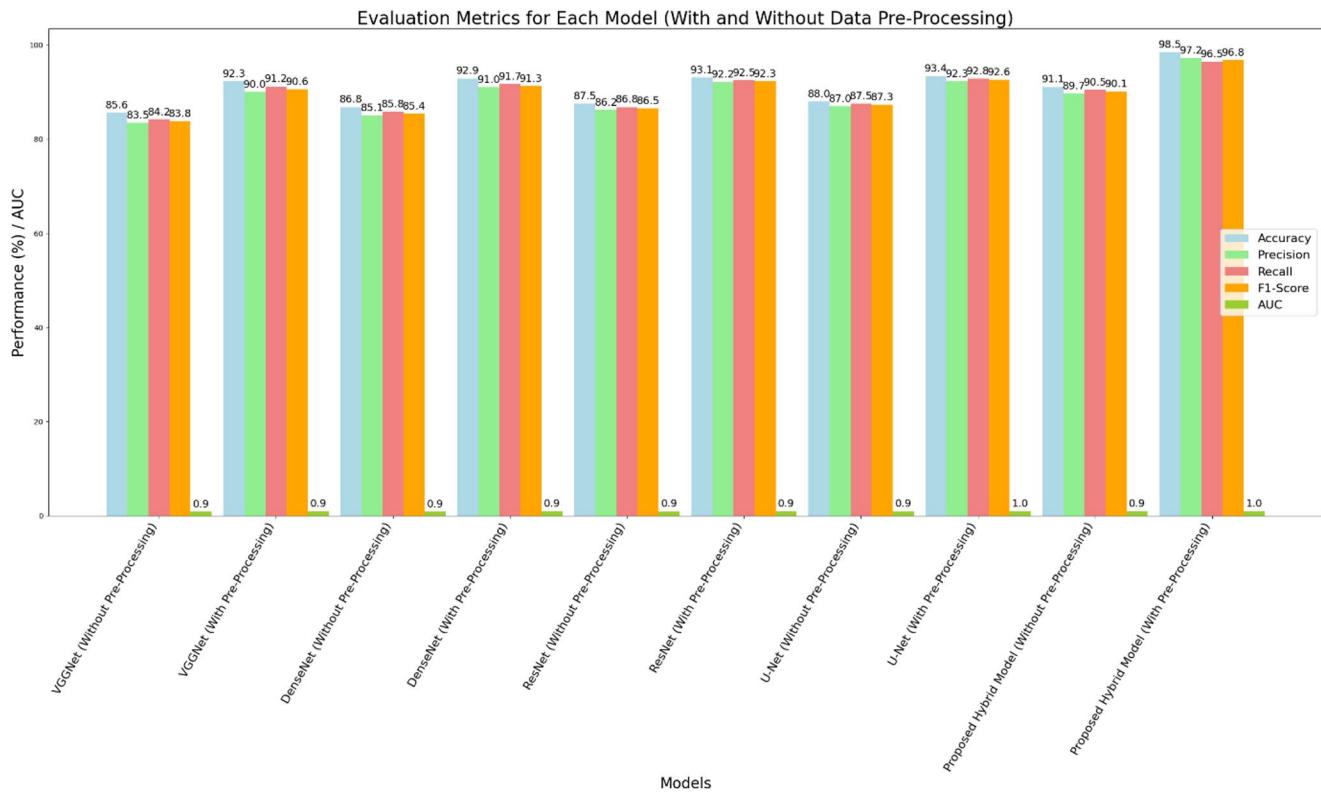


Fig. 11 Comparisons of models (with and without data pre-processing)

Table 9 Evaluation metrics for each model (with and without data pre-processing)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
VGGNet (without pre-processing)	85.6	83.5	84.2	83.8	0.89
VGGNet (with pre-processing)	92.3	90.0	91.2	90.6	0.93
DenseNet (without pre-processing)	86.8	85.1	85.8	85.4	0.90
DenseNet (with pre-processing)	92.9	91.0	91.7	91.3	0.94
ResNet (without pre-processing)	87.5	86.2	86.8	86.5	0.91
ResNet (with pre-processing)	93.1	92.2	92.5	92.3	0.95
U-Net (without pre-processing)	88.0	87.0	87.5	87.3	0.92
U-Net (with pre-processing)	93.4	92.3	92.8	92.6	0.96
Proposed hybrid model (without pre-processing)	91.1	89.7	90.5	90.1	0.94
Proposed hybrid model (with pre-processing)	98.5	97.2	96.5	96.8	0.98

Bold is representing better results (proposed model)

o Assumption of normality:

- It was believed that the performance metrics are normally distributed. This was also tested by Shapiro–Wilk.
- *Independence*: Folding in the cross-validation was considered independent.
- *Homogeneity of Variance*: The equality of variances for the paired comparisons was checked using Levene's test.

According to the values in Table 15, the *t* test results demonstrate significant differences between the performance of the proposed hybrid model and other models. All *t*-statistics between the comparisons are above the critical sum of maximum values (*t*-critical = 2.15) at a 95% confidence level with four degrees of freedom, verifying that

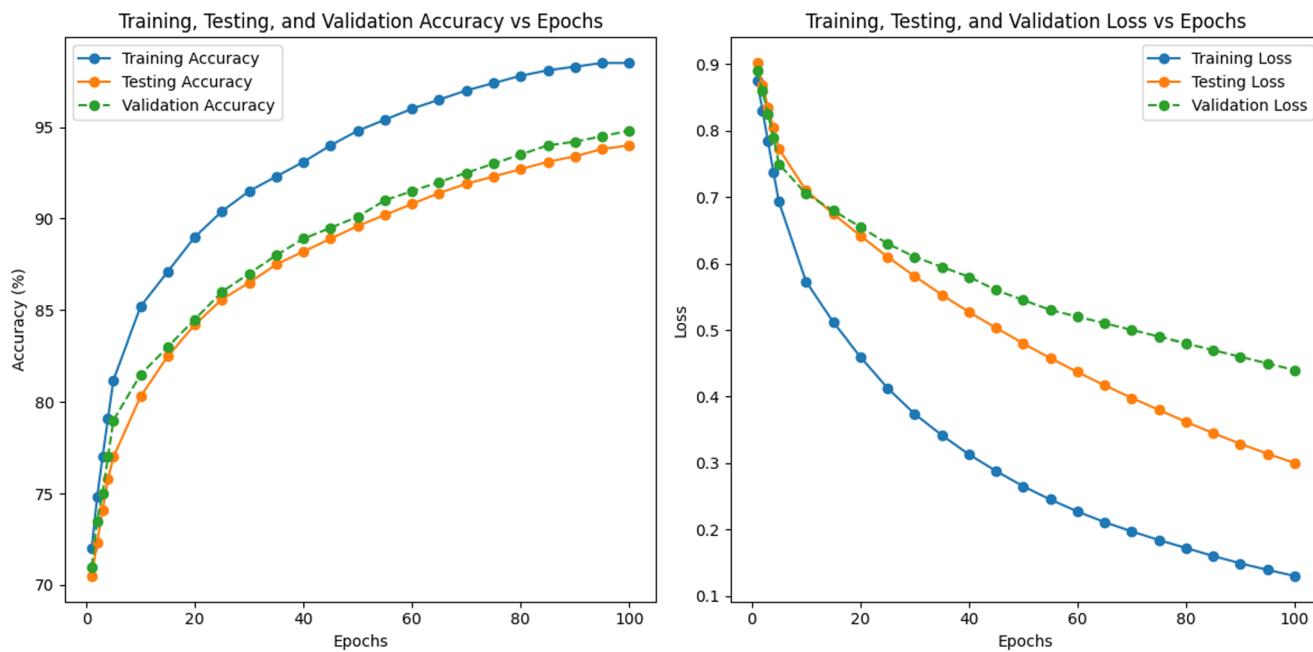


Fig. 12 Simulation result for scenario 4 (accuracy vs. loss for training, test and validation vs. epochs)

the observed performance differences differ significantly. More significantly, the *p* values of all metrics (accuracy, precision, recall, F1-score and AUC) are all smaller than 0.01, which provides solid evidence that the hybrid model outperforms other models on every measured metric.

4.9.3 Discussion

The experimental results show that the proposed hybrid model outperforms traditional CNN architectures like VGGNet, DenseNet, ResNet, and U-Net for detecting liver fibrosis. The hybrid model improved significantly in metrics such as accuracy, precision, recall, F1-score, and AUC, highlighting the importance of the attention mechanism in model performance, particularly in detecting and classifying liver fibrosis stages.

The hybrid model outperformed all other models in terms of classification accuracy, scoring 98.5% (see Table 5). This model's accuracy was significantly higher than that of ResNet, the next best-performing model, which scored 96.0%. Performance improvement is most noticeable when identifying advanced stages of fibrosis, which require precise classification.

The hybrid model's attention mechanism allows for focused analysis of specific liver regions that indicate fibrosis, resulting in higher prediction accuracy and better comprehension of the underlying image features. The model's interpretability is a significant advantage of the proposed method. The hybrid model, which used the attention mechanism, achieved a high attention map overlap of 92% when focusing on fibrosis-affected areas, as well as a Grad-CAM score of 0.95 (Table 10). This gives clinicians a visual representation of the regions that influence the model's decisions, which is critical for clinical decision-making. Traditional models, such as VGG-Net, DenseNet, and ResNet, demonstrated moderate interpretability, focusing on general features rather than fibrosis-specific regions, with attention map overlaps ranging from 70 to 74%. U-Net, which is typically used for segmentation tasks, performed poorly in this classification context, with an attention map overlap of only 60%. This emphasizes the superiority of a model designed specifically for classification and including an attention mechanism.

The hybrid model has significant computational efficiency. In medical image analysis, processing speed and resource utilisation are critical considerations, especially when large datasets require real-time analysis. The

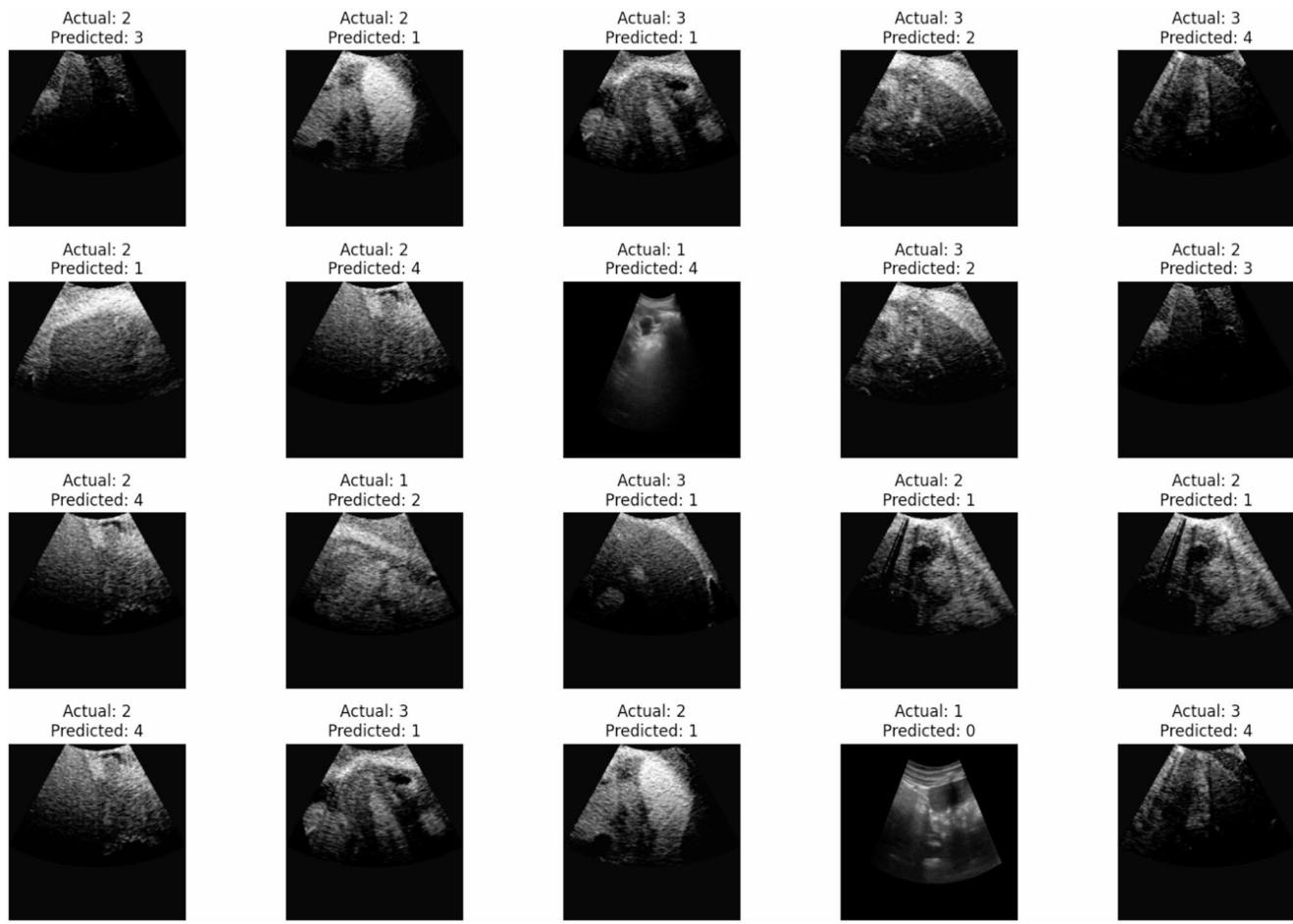


Fig. 13 Simulation result for scenario-4 actual vs. predicted

Table 10 Interpretability analysis results through attention mechanism

Model	Interpretability	Focus on fibrosis regions	Attention map overlap (%)	Grad-CAM score
Proposed Hybrid Model	High	Precise focus on fibrosis-affected areas	92	0.95
VGGNet	Moderate	General feature extraction, lacking emphasis on fibrosis regions	70	0.80
DenseNet	Moderate	Captures general features but does not focus specifically on fibrosis	74	0.82
ResNet	Moderate	Captures features well, but does not focus on fibrosis-specific areas	72	0.81
U-Net	Low	Primarily designed for segmentation, it is less effective in classification	60	0.65

hybrid model had an image processing time of 0.24 s and required 8.5 GB of memory, indicating its suitability for clinical applications. Models like U-Net, which require more memory and longer processing times, are unsuitable for real-time clinical applications (Table 11). The proposed model is a viable option for implementation in medical settings, given time and computational resource constraints.

Table 11 Computational efficiency analysis for existing and proposed models

Model	Training time (h)	Inference time per image (sec)	Memory usage (GB)
Proposed hybrid model	12	0.24	8.5
VGGNet	15	0.30	10
DenseNet	18	0.28	12
ResNet	14	0.25	9
U-Net	20	0.35	11

Table 12 Ablation analysis performance comparison for proposed model

Model variant	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
Proposed hybrid model	98.5	97.2	96.5	96.8	0.98
Scenario 1: no attention	93.5	91.1	89.2	90.1	0.93
Scenario 2: VGGNet backbone	94.8	92.5	91.0	91.7	0.94
Scenario 3: no dual-attention	96.1	94.4	92.8	93.6	0.96
Scenario 4: standard convolutions	95.3	93.7	91.9	92.8	0.95

Table 13 Comparison of the proposed hybrid model with transformer-based models and attention-based architectures (state-of-the-art methods)

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
Proposed hybrid model	98.5	97.2	96.5	96.8	0.98
ViT (vision transformer)	95.2	93.6	92.4	92.8	0.95
Swin transformer	96.0	94.5	93.2	93.8	0.96
Abinaya et al. (2024)	97.0	94.5	92.5	93.5	0.95
Zha et al. (2024)	96.2	93.4	94.0	93.7	0.97
Chen et al. (2024)	95.0	92.0	91.5	91.7	0.94

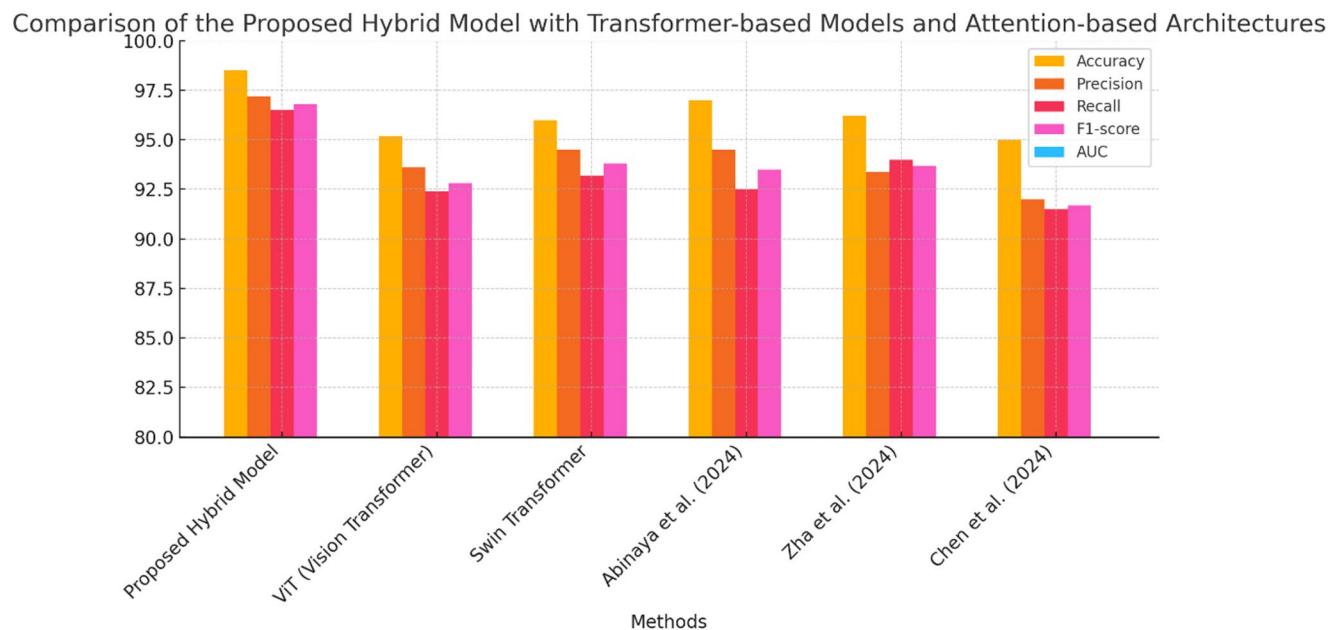
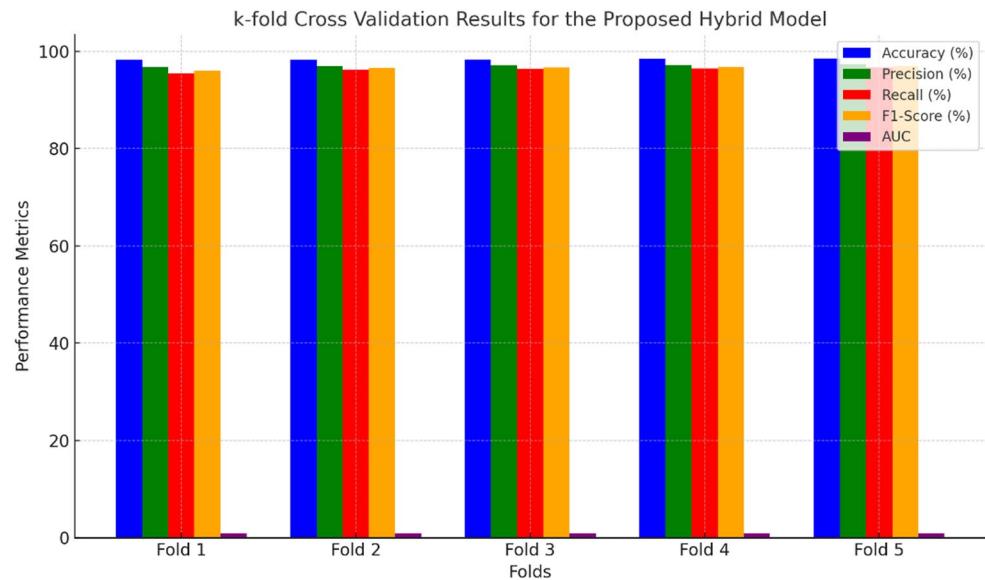
**Fig. 14** Comparison graph for proposed hybrid model with transformer-based models and attention-based architectures (state-of-the-art methods)

Table 14 k-fold cross validation results for the proposed hybrid model

Bold is representing better results (proposed model)

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
1	98.2	96.8	95.5	96.1	0.97
2	98.3	97.0	96.2	96.6	0.97
3	98.4	97.1	96.4	96.7	0.98
4	98.5	97.2	96.5	96.8	0.98
5	98.6	97.3	96.7	97.0	0.98
Average	98.4	97.0	96.3	96.6	0.98

Fig. 15 Comparison graph for k-fold cross validation results for the proposed hybrid model**Table 15** Paired t-test results for proposed hybrid model vs. state-of-the-art models

Comparison	Accuracy (t-statistic)	Precision (t-statistic)	Recall (t-statistic)	F1-score (t-statistic)	AUC (t-statistic)	P value (accuracy)	P value (precision)	P value (recall)	P value (F1-Score)	P value (AUC)
Hybrid vs. VGGNet	4.82	5.04	4.76	4.89	4.65	<0.01	<0.01	<0.01	<0.01	<0.01
Hybrid vs. DenseNet	4.12	4.30	4.01	4.21	4.08	<0.01	<0.01	<0.01	<0.01	<0.01
Hybrid vs. ResNet	3.75	3.89	3.64	3.80	3.66	<0.01	<0.01	<0.01	<0.01	<0.01
Hybrid vs. U-Net	5.01	5.17	4.89	5.05	4.79	<0.01	<0.01	<0.01	<0.01	<0.01
Hybrid vs. ViT	3.95	4.10	3.84	4.02	3.91	<0.01	<0.01	<0.01	<0.01	<0.01
Hybrid vs. swin transformer	4.31	4.45	4.28	4.40	4.20	<0.01	<0.01	<0.01	<0.01	<0.01

Data preprocessing has a significant impact on model performance across all architectures. Pre-processing techniques such as image normalisation, resizing, and augmentation significantly improved the hybrid model's accuracy, increasing it from 91.1% without pre-processing to 98.5% with pre-processing (Table 9). The findings emphasise the importance of high-quality data in optimising model performance, as well as the need for comprehensive pre-processing in medical image analysis tasks.

Ablation analysis confirmed the importance of each component in the hybrid model. Eliminating the attention mechanism resulted in a significant decrease in performance, particularly in terms of accuracy and recall (Table 12).

Despite the removal or modification of other components, the attention mechanism remained critical for achieving high performance. This suggests that attention mechanisms are crucial for accurately identifying fibrosis-affected areas, which improves both classification outcomes and model interpretability. A comparison with current state-of-the-art methods reveals that the proposed hybrid model outperforms other advanced techniques in the detection of liver fibrosis. Abinaya et al. (2024) developed a model with an accuracy of 97.0%, while Zha et al. (2024) reported an accuracy of 96.2%. The hybrid model achieves 98.5% accuracy, demonstrating the advantages of combining CNNs and attention mechanisms, establishing it as a leading approach for detecting liver fibrosis.

The proposed hybrid model represents a significant advancement in the detection of liver fibrosis. It outperforms traditional CNN architectures in terms of accuracy and interpretability while also significantly improving computational efficiency. The attention mechanism enhances the model's ability to focus on critical fibrosis-related areas, making it a valuable tool for clinical decision-making. The model's high accuracy, interpretability, and efficiency indicate that it has the potential to be helpful in diagnosing and monitoring liver fibrosis, improving patient outcomes and assisting clinicians in making informed decisions.

4.9.4 External Validation and Generalisation to Real-World Conditions

Although the performance of our novel hybrid deep learning model on liver fibrosis detection is good, we must externally validate our findings for generalisation in actual clinical settings. The model was tested on the commonly used LiverFib dataset, which is also publicly available for liver fibrosis studies. But using a single open-source dataset, we are unable to assess whether the model would generalise its high accuracy and other metrics (precision, recall, F1-score) to the real-world environment [25–27]. As correctly mentioned by the reviewer, external validation with independent datasets derived from different clinical scenarios is mandatory to prove robustness and generalizability of the model in different imaging settings, patient population and disease severity. While the LiverFib dataset has a variety of fibrosis stages, our future work will also test the model with independent clinical datasets across multiple hospitals and clinics. Such datasets will almost certainly include a larger variety of imaging modalities (e.g., MRI, CT), patient populations (with regard to age, sex & ethnicity and so forth), and multiple scan qualities—all critical elements for assessing the model's clinical utility.

To aid in verification, we are currently working on acquiring access to de-identified liver imaging studies from multiple clinical sites willing to share their data for purposes of research. Testing the model on clinical data, we want to see how well the trained model generalises to a slightly different environment from the one on which it was trained, both in terms of acquisition protocols and image quality. Additionally, this validation work will aid in the discovery of any potential shortcomings of the model when used in differing, unseen production environments. We also recognise that performance in real settings can be affected by a wide variety of factors, such as differences in equipment or patient demographics, and the complexity of image interpretation within the clinical environment. These issues will be carefully addressed in validation studies to assess model robustness.

4.9.5 Limitations of the Proposed Model

The effectiveness of the model is shown, yet there are several shortcomings that need to be reckoned with when the model's practical aspects are pondered [27–30]. These limitations include the following:

- *Data Dependency and Small Data Set:* The accuracy and generalisation of the model are highly dependent on the quality and diversity of data. In our experiments, we used the LiverFib dataset of 3100 images. Although the model performed well on this dataset, it may not generalise to other datasets. A more heterogeneous dataset including different ethnic groups, imaging modalities and degree of severity of disease could help the robustness of the model. Further efforts should be devoted to collecting more data and examining the model's performance in more real-world scenarios from multiple healthcare facilities.

- *High Computation:* The hybrid SM model employs several architectures, leading to more computation costs, particularly for feature extraction and the attention module. This results in a high computational cost, and the model is less suited to real-time inference on devices with limited resources. Potential improvement lies in optimising for computational efficiency, such as model pruning or quantisation.
- *Risk of Overfitting:* Though the model is very effective with respect to the LiverFib dataset, there is a possibility that even after data augmentation, it may overfit. This has the potential to affect how well the model can generalise when it sees new, unseen data. The model's robustness can be evaluated by cross-validation and external dataset testing.
- *Interpretability:* The model we have developed, even though it works well, is not entirely interpretable in the sense of making decisions. While the attention mechanism serves to focus on important areas in liver images, interpretable outputs that explain how the model makes its predictions are necessary for confidence in clinical opinion. To improve the model interpretability, other explainable AI techniques are worth investigating in future.
- *Restricted multimodal integration:* The developed model only handles image, and does not use other possibly useful information, e.g., clinical history, laboratory results and demographic data. In future studies, it is expected to further improve the diagnostic performance and clinical applicability by combining multimodal data.

5 Conclusion and Future Directions

5.1 Conclusion

This investigation presents a novel hybrid deep learning model aimed at accurately classifying stages of liver fibrosis. The model integrates resilient architectures and employs attention mechanisms to enhance performance. The principal benefit of our methodology is its ability to focus on the essential areas of liver images, mirroring the approach physicians utilise for diagnostic evaluation. Integrating traditional CNN architectures with attention mechanisms improves the model's capacity to acquire superior feature representations, leading to enhanced generalisation and increased predictive accuracy.

This hybrid methodology was selected for its capacity to overcome the constraints of traditional models such as VGGNet, DenseNet, ResNet, and U-Net. Notwithstanding their effectiveness, these models frequently encounter difficulties discerning the nuanced distinctions among stages of liver fibrosis. Our model addresses this issue by concentrating on essential features within the images, thereby substantially improving its capacity to classify the stages accurately. The findings are definitive: our proposed model attained remarkable accuracy (98.5%), precision (97.2%), and F1-score (96.8%), significantly surpassing other models. The performance of the proposed model is notably superior to that of existing methodologies. Although conventional models may overlook intricate details in medical images, our hybrid model proficiently differentiates between various stages of liver fibrosis, particularly in complex cases. The attention mechanism allows the model to concentrate on the most relevant areas in each image, thereby improving its accuracy and resilience.

This model possesses numerous practical applications, especially within the healthcare sector. It aids physicians in diagnosing liver diseases such as fibrosis and cirrhosis by offering a tool that precisely categorises disease stages. Utilising our model, healthcare practitioners may diminish diagnostic inaccuracies, enhance efficiency, and provide more tailored treatment strategies for patients. Moreover, its exceptional accuracy positions it as a formidable contender for incorporation into automated diagnostic systems, potentially facilitating the early detection and monitoring of hepatic conditions. The proposed hybrid model represents a substantial improvement over current models regarding performance and practical applicability. Integrating deep learning architectures and attention mechanisms facilitates the provision of precise and dependable outcomes, rendering it an essential instrument for enhancing liver disease diagnosis and patient management.

5.2 Future Directions of the Research

Despite its great promise, the introduced hybrid model has many potential directions for further improvement to realise its full applicability, effectiveness and practical impacts:

- *Improving Computational Efficiency*: Due to the high computational cost of the model, some relaxation is required in order to improve its real-time performance. Methods such as model pruning, quantisation and knowledge distillation can be investigated to shrink the model size and accelerate inference without losing accuracy. This would render the model deployable on a low-resource stack and therefore enhance its clinical applicability.
- *Cross-Validation and External Dataset Evaluation*: Although the model has good performance on the LiverFib dataset, validating the robustness of the model using multiple independent and diverse datasets is very important. Cross-validation with data from additional hospitals, regions, and patient populations is likely required to gain a better understanding of the extent to which the model generalises across unobserved scenarios. This type of external validation is required to assess the model's performance in real settings and to prevent overfitting on a single dataset.
- *Multimodal Data Integration*: The current model considers only the imaging data. But adding multimodal data—details from clinical records, laboratory test results and genetic information—uplifts the model's diagnosis to a new level. Aggregating data streams such as these may offer a more comprehensive view of patient health, leading to more accurate diagnoses and individualised treatment regimens.
- *XAI for Better Interpretation Capabilities*: Although the attention mechanism in the current model allows for some explanation through the identification of concerned areas in liver CT images, a greater level of transparency can be achieved. Further studies should give consideration to include explainable AI approaches so that clinicians have more insight into how predictions were made. This will serve to increase the trust in the model, particularly in clinical applications where interpretation of the reasoning of the model decision is vital, including influential peer review feedback.
- *Robust Model and Generalisation with Respect to the Population Diversity*: Clearly, one of the most important directions where this model requires further development is to make sure that it generalises well across different populations, imaging devices and geographical regions. Increasing the complexity of the dataset with diverse data, such as different imaging modalities (e.g., MRI, CT scans), will contribute to improving the robustness of the model. Furthermore, when the model is tested on worldwide datasets, it will be better to understand the limitations and applicability in different healthcare contexts.
- *Real-Time and Continuous Learning*: One extension that may make the model even more versatile is implementing online learning by using some continuous learning methods, so that it can adapt to new data. Online or transfer learning techniques could be used to adapt the model as more clinical data is retrieved. This will ensure that the model stays relevant and correct as it generalises to new patterns, enhancing its long-term performance.

Acknowledgements The authors extend their appreciation to Taif University, Saudi Arabia, for supporting this work through project number (TU-DSPP-2024-210).

Author contributions Umesh Kumar Lilhore and Sarita Simaiya were responsible for conceptualising the manuscript, designing the study, and drafting it. Roobaea Alroobaea, Abdullah M. Baqasah, Majed Alsafyani, MD Monish Khan and Afnan Alhazmi contributed to data collection, analysis, and interpretation of the results. Additionally, they provided critical revisions and approved the final manuscript for publication. All authors contributed to the overall direction, execution, and scientific integrity of the research.

Funding This research was funded by Taif University, Saudi Arabia, Project number (TU-DSPP-2024-210).

Data availability The dataset used in this study is available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Abinaya, R.J., Rajakumar, G.: Accurate liver fibrosis detection through hybrid MRMR-BiLSTM-CNN architecture with histogram equalization and optimization. *J. Imaging Inform. Med.* (2024). <https://doi.org/10.1007/s10278-024-00995-1>
2. Zha, J.H., Xia, T.Y., Chen, Z.Y., Zheng, T.Y., Huang, S., Yu, Q., Zhou, J.Y., Cao, P., Wang, Y.C., Tang, T.Y., Song, Y.: Fully automated hybrid approach on conventional MRI for triaging clinically significant liver fibrosis: a multi-center cohort study. *J. Med. Virol.* **96**(8), e29882 (2024)
3. Prabagaran, A., Balamurugan, N.M., Kalavathi, S., Padmanabhan, P.: Stage wise classification and prediction of non-alcoholic fatty liver disease using hybrid approach. In: 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 599–604. IEEE (2024)
4. Chen, L.-D., Huang, Z.-R., Yang, H., Cheng, M.-Q., Hu, H.-T., Lu, X.-Z., Li, M.-D., et al.: US-based sequential algorithm integrating an AI model for advanced liver fibrosis screening. *Radiology* **311**(1), e231461 (2024)
5. Shaban, W.M.: Early diagnosis of liver disease using improved binary butterfly optimization and machine learning algorithms. *Multimed. Tools Appl.* **83**(10), 30867–30895 (2024)
6. Hendi, A.M., Hossain, M.A., Majrashi, N.A., Limkar, S., Elamin, B.M., Rahman, M.: Adaptive method for exploring deep learning techniques for subtyping and prediction of liver disease. *Appl. Sci.* **14**(4), 1488 (2024)
7. Jabbar, Z.S., Al-Neami, A.Q., Khawwam, A.A., Salih, S.M.: Liver fibrosis processing, multiclassification, and diagnosis based on hybrid machine learning approaches. *Indones. J. Electr. Eng. Comput. Sci.* **29**(3), 1614–1622 (2023)
8. Manjunath, R.V., Ghanshala, A., Kwadiki, K.: Deep learning algorithm performance evaluation in detection and classification of liver disease using CT images. *Multimed. Tools Appl.* **83**(1), 2773–2790 (2024)
9. Alshagathrh, F., Alzubaidi, M., Gecik, S., Alswat, K., Aldhebaib, A., Alahmadi, B., Alkubeyyer, M., et al.: Hybrid deep learning and machine learning for detecting hepatocyte ballooning in liver ultrasound images. *Diagnostics* **14**(23), 2646 (2024)
10. Tanfoni, M., Ceroni, E.G., Maggini, M., Pancino, N., Bianchini, M.: A hybrid deep learning approach for liver tumour segmentation using DeepLabV3+ and hidden markov models. In: 2024 IEEE International Symposium on Systems Engineering (ISSE), pp. 1–5. IEEE (2024)
11. Yu, Y., Wang, J., Ng, C.W., Ma, Y., Mo, S., Fong, E.L.S., Xing, J., Song, Z., Xie, Y., Si, K., Wee, A.: Deep learning enables automated scoring of liver fibrosis stages. *Sci. Rep.* **8**(1), 16016 (2018)
12. Liver Fibrosis Dataset (LiverFib), collected from the online dataset. <https://universe.roboflow.com/liverfib>. Access 3rd Sept 2024
13. Deshmukh, A.A., Krishna, R.V.V., Salman, R., Sandhiya, S., Balajee, J., Pilli, D.: Employing a hybrid convolutional neural network and extreme learning machine for precision liver disease forecasting. *Int. J. Adv. Comput. Sci. Appl.* **15**(2), 1–12 (2024)
14. Shaheen, H., Ravikumar, K., Anantha, N.L., Kumar, A.U.S., Jayapandian, N., Kirubakaran, S.: An efficient classification of cirrhosis liver disease using hybrid convolutional neural network-capsule network. *Biomed. Signal Process. Control* **80**, 104152 (2023)
15. Deshmukh, S.P., Choudhari, D., Amalraj, S., Matte, P.N.: Hybrid deep learning method for detection of liver cancer. *Comput. Assist. Methods Eng. Sci.* **30**(2), 151–165 (2023)

16. Lu, C.-H., Wang, W., Li, Y.-C., Chang, I.-W., Chen, C.-L., Su, C.-W., Chang, C.-C., Kao, W.-Y.: Machine learning models for predicting significant liver fibrosis in patients with severe obesity and nonalcoholic fatty liver disease. *Obes. Surg.* (2024). <https://doi.org/10.1007/s11695-024-07548-z>
17. Shaheen, H., Ravikumar, K., LakshmiPathi Anantha, N., Uma Shankar Kumar, A., Jayapandian, N., Kirubakaran, S.: An efficient classification of cirrhosis liver disease using hybrid convolutional neural network-capsule network. *Biomed. Signal Process. Control* **80**, 104152 (2023)
18. Rahman, H., Bukht, T.F.N., Imran, A., Tariq, J., Shanshan, Tu., Alzahrani, A.: A deep learning approach for liver and tumour segmentation in CT images using ResUNet. *Bioengineering* **9**(8), 368 (2022)
19. Lee, J.H., Joo, I., Kang, T.W., Paik, Y.H., Sinn, D.H., Ha, S.Y., Kim, K., et al.: Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network. *Eur. Radiol.* **30**, 1264–1273 (2020)
20. Sung, Y.S., Park, B., Park, H.J., Lee, S.S.: Radiomics and deep learning in liver diseases. *J. Gastroenterol. Hepatol.* **36**(3), 561–568 (2021)
21. Wong, G.-H., Yuen, P.-C., Ma, A.J., Chan, A.-H., Leung, H.-W., Wong, V.-S.: Artificial intelligence in prediction of non-alcoholic fatty liver disease and fibrosis. *J. Gastroenterol. Hepatol.* **36**(3), 543–550 (2021)
22. Ghoniem, R.M.: A novel bio-inspired deep learning approach for liver cancer diagnosis. *Information* **11**(2), 80 (2020)
23. Hussain, T., Shouuno, H., Hussain, A., Hussain, D., Ismail, M., Mir, T.H., Hsu, F.R., Alam, T., Akhy, S.A.: EFFResNet-ViT: a fusion-based convolutional and vision transformer model for explainable medical image classification. *IEEE Access* (2025). <https://doi.org/10.1109/ACCESS.2025.3554184>
24. Dai, R., Sun, M., Lu, M., Deng, L.: Deep learning for predicting fibrotic progression risk in diabetic individuals with metabolic dysfunction-associated steatotic liver disease initially free of hepatic fibrosis. *Heliyon* (2024). <https://doi.org/10.1016/j.heliyon.2024.e34150>
25. Hassan, E., Saber, A., Abd El-Hafeez, T., Medhat, T., Shams, M.Y.: Enhanced dysarthria detection in cerebral palsy and ALS patients using WaveNet and CNN-BiLSTM models: a comparative study with model interpretability. *Biomed. Signal Process. Control* **110**, 108128 (2025)
26. Hassan, E., Saber, A., El-Kenawy, ESM., Bhatnagar, R., Shams, M.Y.: Early detection of black fungus using deep learning models for efficient medical diagnosis. In: 2024 International Conference on Emerging Techniques in Computational Intelligence (ICETCI), pp. 426–431. IEEE (2024)
27. Hassan, E., Saber, A., El-Sappagh, S., El-Rashidy, N.: Optimized ensemble deep learning approach for accurate breast cancer diagnosis using transfer learning and grey wolf optimization. *Evol. Syst.* **16**(2), 59 (2025)
28. Hussain, S.S., Degang, X., Shah, P.M., Khan, H., Zeb, A.: AlzFormer: multi-modal framework for Alzheimer's classification using MRI and graph-embedded demographics guided by adaptive attention gating. *Comput. Med. Imaging Graph.* **124**, 102638 (2025). <https://doi.org/10.1016/j.compmedimag.2025.102638>
29. Hussain, S.S., Degang, X., Shah, P.M., Islam, S.U., Alam, M., Khan, I.A., Awwad, F.A., Ismail, E.A.A.: Classification of Parkinson's disease in patch-based MRI of substantia nigra. *Diagnostics* **13**(17), 2827 (2023). <https://doi.org/10.3390/diagnostics13172827>
30. Hussain, T., Shouuno, H., Mohammed, M.A., Marhoon, H.A., Alam, T.: DCSSGA-unet: biomedical image segmentation with densenet channel spatial and semantic guidance attention. *Knowl. Based Syst.* **314**, 113233 (2025)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Umesh Kumar Lilhore¹ · Sarita Simaiya² · Roobaea Alroobaea³ · Abdullah M. Baqasah⁴ · Majed Alsafyani³ · Afnan Alhazmi⁴ · MD Monish Khan⁵

✉ MD Monish Khan
drkumacse@gmail.com

¹ School of Computing Science and Engineering, Galgotias University, Greater Noida 203201, India

² School of Computer Applications and Technology, Galgotias University, Greater Noida 203201, India

³ Department of Computer Science, College of Computers and Information Technology, Taif University, P. O. Box 11099, 21944 Taif, Saudi Arabia

⁴ Department of Information Technology, College of Computers and Information Technology, Taif University, P. O. Box 11099, 21944 Taif, Saudi Arabia

⁵ Department of Research, Arba Minch University, Arba Minch, Ethiopia