Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 1, January 2026, pp: 284-303 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Classification of Ultrasound Images Using ResNet-50 with a Convolutional Block Attention Module (CBAM)

Bagus Tegar Zahir Afif, Wiharto, and Umi Salamah

Department of Informatics, Faculty of Information Technology and Data Science, University of Sebelas Maret, Surakarta, Indonesia

Corresponding author: Wiharto (e-mail: wiharto@staff.uns.ac.id), **Author(s) Email**: Bagus Tegar Zahir Afif (e-mail: bagustegar@student.uns.ac.id), Umi Salamah (e-mail: umisalamah@staff.uns.ac.id)

Abstract **Liver fibrosis staging is a crucial component in the clinical management of chronic liver disease because it directly affects prognosis, therapeutic decision-making, and long-term patient monitoring. Ultrasound imaging is widely used as a noninvasive diagnostic modality due to its safety, low cost, and broad accessibility. Nevertheless, ultrasound-based fibrosis assessment remains challenging because liver parenchymal echotexture often exhibits low contrast, speckle noise, and subtle inter-stage variations, particularly among adjacent METAVIR stages. These characteristics frequently limit the effectiveness of conventional convolutional neural networks, which tend to emphasize dominant global patterns while suppressing weak but clinically meaningful texture cues. This study presents a task-oriented integration of a Convolutional Block Attention Module into a ResNet-50 backbone to enhance feature discrimination for five-stage liver fibrosis classification using heterogeneous B-mode ultrasound images. Rather than introducing a new attention mechanism, the contribution lies in the systematic insertion of CBAM after residual outputs across multiple network stages, enabling repeated channel and spatial recalibration from low-level texture descriptors to higher-level semantic representations. To further improve robustness and reduce prediction variance, a stratified 5-fold training strategy is combined with logit-level ensemble inference, where logits from independently trained fold models are averaged prior to Softmax normalization. Experiments were conducted on a publicly available dataset comprising 6,323 ultrasound images acquired from two tertiary hospitals using multiple ultrasound systems, with fibrosis stages labeled from F0 to F4 according to histopathology-based METAVIR scoring. The proposed framework achieves a test accuracy of 98.34%and consistently high precision, recall, and F1 scores across all fibrosis stages, with the most pronounced improvement observed for intermediate stages. Statistical analysis based on paired fold-wise comparisons confirms that the performance gain over the baseline ResNet 50 model is statistically significant. These results demonstrate that combining lightweight attention-based feature refinement with logit ensemble inference effectively addresses the inherent challenges of ultrasound-based liver fibrosis staging and provides a reliable noninvasive decision support framework with strong potential for clinical application and future multicenter validation.**

Keywords **Liver fibrosis; Ultrasound; ResNet-50; CBAM**.

I. Introduction

Chronic liver disease continues to exert a profound impact on global health, accounting for a significant proportion of deaths worldwide [1]. Its development is linked to various primary causes, including chronic hepatitis B and C infections, alcohol associated hepatic disease, and the increasingly prevalent non-alcoholic fatty liver disease (NAFLD), driven in part by the worldwide rise in obesity [2]. Among the pathological manifestations of long-term liver injury, fibrosis is one of the most prevalent, arising from continuous hepatocellular insult and characterized by excessive collagen-rich extracellular matrix deposition [3]. Without early diagnosis and appropriate therapeutic intervention, fibrosis may advance to end-stage liver complications such as cirrhosis, liver failure, and hepatocellular carcinoma (HCC), which are linked to poor patient outcomes [4]. A comprehensive meta-analysis including 168,571 participants across 19 studies revealed that non-cirrhotic MASH patients exhibited a significantly higher prevalence of HCC (38.0%) compared with non-cirrhotic individuals with other liver disease etiologies (14.2%; $p < 0.001$) [5], [6]. Therefore, evaluating liver fibrosis is a fundamental aspect of chronic liver disease (CLD) management, facilitating prognosis, individualized therapeutic and

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

surveillance strategies, and assessment of treatment response over time [7].

Given these challenges, early detection and accurate staging of liver fibrosis are critically important in clinical practice [7]. Reliable fibrosis staging not only guides prognosis but also informs therapeutic decisions and enables effective monitoring of treatment response [8]. For decades, liver biopsy has been regarded as the gold standard for determining the severity of fibrosis [9]. However, biopsy presents substantial limitations, including its invasive nature, risk of complications such as bleeding, relatively high cost, and susceptibility to sampling error, which can lead to inaccurate or non-representative assessments [10].

In response to these limitations, non-invasive diagnostic alternatives have been increasingly developed, with ultrasound (US) imaging becoming a prominent option [11]. Since its introduction as a diagnostic technique in the mid twentieth century, ultrasound has evolved into a compact and sophisticated modality characterized by enhanced image quality and advanced features, including artificial intelligence applications [12]. Early hepatobiliary applications date back to 1958, when only A-mode ultrasound was available, followed by the introduction of real-time B-mode imaging in the mid 1970s [13]. Increasing reliance on ultrasound imaging highlights the importance of understanding its physical principles, including sound propagation, tissue interactions, and image construction, to improve diagnostic accuracy and reduce artifacts [14]. Ultrasound is extensively utilized as the first line modality for liver evaluation due to its wide availability, cost-effectiveness, and safety advantages, such as the absence of ionizing radiation and contrast agents [15], [16]. However, the interpretation of ultrasound images is strongly influenced by operator experience, resulting in variability and limiting the consistency of diagnostic assessments [17]. Recent studies further emphasize that machine learning and deep learning techniques offer significant advantages in reducing inter-observer variability and improving diagnostic reliability in ultrasound-based liver assessment [18], [19]. These findings show that automated feature extraction can enhance fibrosis staging performance even when ultrasound quality varies due to operator dependency.

The rapid evolution of artificial intelligence (AI) in recent years, particularly through advances in deep learning techniques, has significantly expanded the possibilities for addressing complex problems in medical image analysis [20], [21]. Among these approaches, Convolutional Neural Networks (CNNs) have gained considerable attention for their outstanding performance across numerous medical imaging applications, including classification, anomaly detection, and segmentation. Their effectiveness primarily stems from the ability to automatically learn rich hierarchical representations of spatial and textural information from imaging data, thereby reducing reliance on manually designed features [22]. Beyond classical CNN architectures, recent research has explored frequency domain ultrasound features and one-dimensional CNNs for liver fibrosis assessment, demonstrating improved robustness when addressing heterogeneous tissue patterns and radiofrequency characteristics [23].

A comprehensive study [24] evaluated multiple CNN architectures for five-stage liver fibrosis classification (METAVIR F0–F4) using heterogeneous ultrasound images. Their results showed that VGGNet achieved 83.17% accuracy, ResNet-50 reached 85.92%, DenseNet 84.17%, EfficientNet 85.17%, and Vision Transformer 83.42%. Backbone selection. We adopt ResNet-50 as the backbone for three reasons. First, on the same heterogeneous dataset and evaluation setting reported by Joo et al. [24], ResNet-50 provided a strong baseline among commonly used backbones, making it an appropriate reference for isolating the contribution of attention. Second, residual learning supports stable optimization with moderate model capacity, which is important for medical ultrasound data, where sample size, acquisition variability, and label noise can limit the safe use of larger backbones without overfitting. Third, ResNet-50 is widely adopted in medical image analysis, enabling clearer comparison and easier replication across studies. We acknowledge that deeper residual networks, EfficientNet families, and transformer-based models can be competitive under specific training regimes. However, fixing a well-understood backbone allows observed gains to be attributed more directly to CBAM insertion and the proposed training and ensembling strategy rather than to changes in network scale. However, a detailed analysis of every class revealed a meaningful limitation: ResNet-50 excelled at identifying extreme fibrosis stages (F0 and F4) but struggled to differentiate intermediate stages (F1–F3), which exhibit subtle textural variations and indistinct anatomical boundaries. Because ResNet-50 treats all feature channels equally, fine-grained diagnostic cues essential for early-stage identification often remain underemphasized [25]. These challenges are further exacerbated by an imbalance in data distribution across fibrosis stages. Recent work employing contrastive fusion strategies on pure B-mode ultrasound images has demonstrated promising results in non-invasive liver fibrosis staging, improving stability and accuracy over traditional CNN baselines [26]. Beyond image quality limitations, conventional CNN architectures are often less effective for intermediate liver fibrosis staging because diagnostically relevant echotexture cues in ultrasound images are weak, diffuse, and overlap across adjacent stages. Standard

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

convolutional and pooling pipelines progressively reduce spatial resolution and uniformly aggregate feature responses, which can attenuate subtle parenchymal texture differences that separate stages F1 to F3. During this process, weak but clinically meaningful fibrosis patterns are frequently suppressed, while stronger yet nonspecific variations such as speckle statistics, scanner-dependent intensity changes, and acquisition-related artifacts, dominate feature representations. These structural characteristics of conventional CNNs limit sensitivity to borderline fibrosis stages and lead to unstable separation between neighboring classes, as also reported in prior imaging studies [27], [28]. Consequently, attention-based feature refinement is required to adaptively reweight informative channels and spatial locations, preserving subtle fibrosis cues while suppressing irrelevant responses.

Context on attention mechanisms: Channel recalibration modules, such as squeeze-and-excitation, are lightweight and effective when discriminative information is primarily encoded in channel responses, but they do not explicitly model where informative regions are located. Non-local blocks and transformer-style self-attention capture long-range dependencies and global context, yet they introduce higher computational overhead and may require stronger regularization or larger datasets to remain robust under heterogeneous ultrasound acquisition. We therefore adopt CBAM, which sequentially applies channel and spatial attention with low overhead, enabling the network to emphasize subtle parenchymal echotexture cues while suppressing speckle-dominated and scanner-dependent responses in B-mode ultrasound. This property is particularly valuable for liver fibrosis staging, where diagnostically relevant textural differences are weak and often overlap across adjacent METAVIR stages. Specifically, channel attention adaptively reweights feature channels to strengthen discriminative responses, while spatial attention highlights informative regions that conventional CNN pipelines may attenuate through uniform aggregation and downsampling. Prior medical imaging studies have reported improved performance after integrating CBAM into ResNet backbones, for example, increasing AUC from 0.772 to 0.866 in a ResNet-50-based classifier [32]. Another study reported an increase in classification accuracy from 74.42% to 95.74% after integrating CBAM into a ResNet-based framework [33]. These findings support the hypothesis that CBAM enhances fine-grained feature extraction essential for distinguishing closely related fibrosis stages.

Building on this evidence, the present study introduces a CBAM-enhanced ResNet-50 model for classifying liver fibrosis severity from ultrasound images. Through its channel attention mechanism, CBAM prioritizes the most diagnostically meaningful features, ensuring optimal use of minority class information, while its spatial attention component reinforces sensitivity to subtle patterns often overlooked by conventional CNNs. This integrated approach is expected to contribute to the advancement of non-invasive clinical tools for liver fibrosis assessment by improving diagnostic accuracy and consistency.

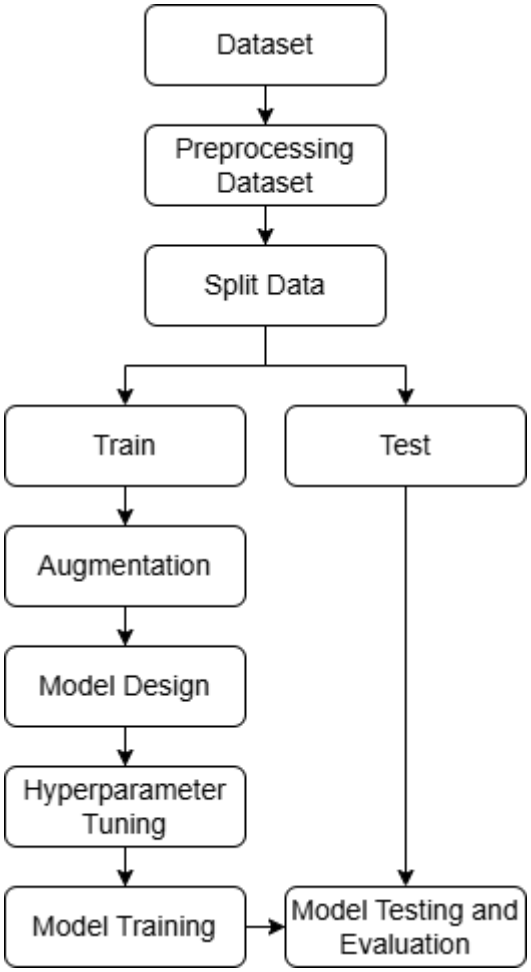


Fig. 1. Flow of the proposed research.

Motivated by these challenges, this study proposes an accurate and interpretable deep learning framework for classifying liver fibrosis stages from ultrasound images. The key contributions of this work can be summarized as follows:

1) Proposing a CBAM-enhanced ResNet-50 architecture designed to improve fine-grained texture recognition in liver ultrasound images.
2) Conducting systematic hyperparameter tuning involving learning rate, weight decay, activation function, loss function, and attention mechanism;

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

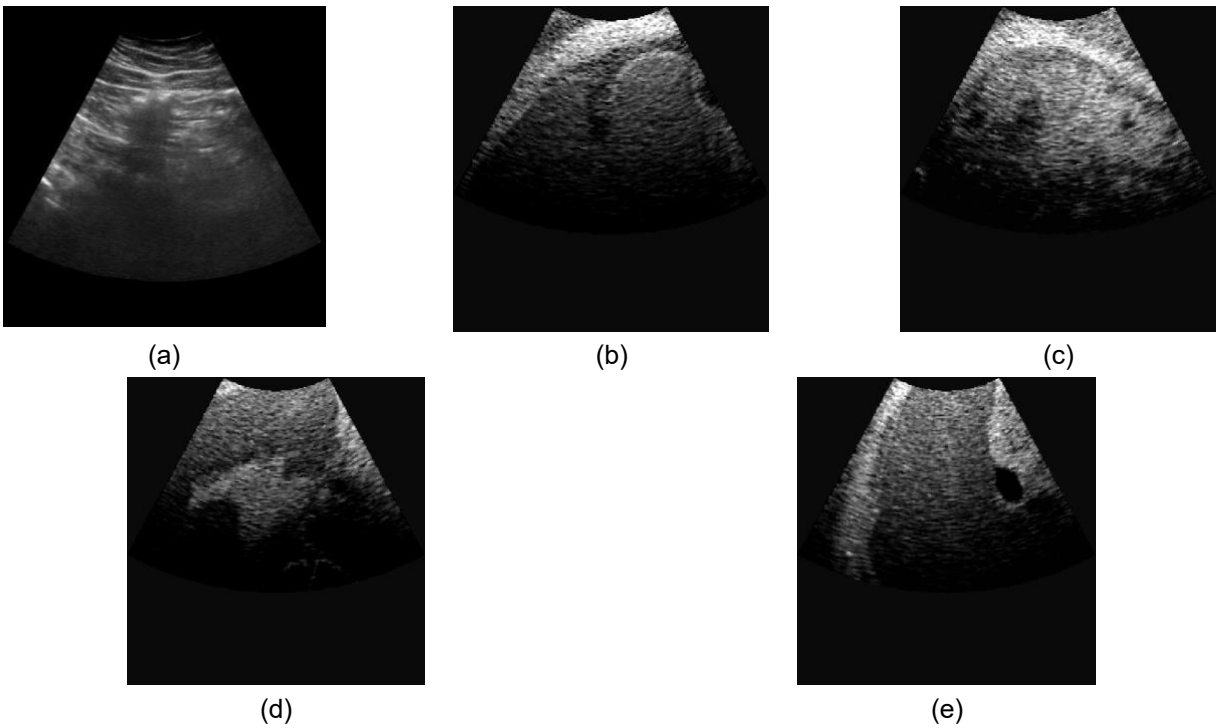


Fig. 2. Sample dataset for each class, (a) F0, (b) F1, (c) F2, (d) F3, (e) F4.

3) Implementing Stratified 5-Fold Cross Validation with ensemble testing to ensure robust and reliable performance.
4) A thorough comparison of previously published state-of-the-art models is also conducted. This paper is organized into five sections: Section II introduces the dataset, data preprocessing steps, and network architecture; Section III reports the experimental framework and findings; Section IV discusses and interprets the results; and Section V summarizes the study and suggests avenues for future research.

II. Method

Fig. 1 illustrates the methodological framework implemented in this research. The workflow includes six major components: dataset preparation, preprocessing, model construction, hyperparameter tuning, training, and testing with performance evaluation.

A. Dataset

This study used a publicly available liver ultrasound dataset on kaggle reported by Joo et al. [24]. The images were acquired at two tertiary university hospitals in South Korea, namely Seoul St. Mary's Hospital and Eunpyeong St. Mary's Hospital, using eight B-mode ultrasound systems from multiple vendors. The original frames had a resolution of 800 × 600 pixels and a sector-shaped field of view before being resized for model input, which introduces realistic inter-device and inter-protocol variability relevant for assessing generalization. Liver fibrosis stages were assigned according to METAVIR criteria (F0- F4) using histopathology as the reference standard, and the staging follows the routine pathology reporting described in [24]. Ultrasound examinations were performed within three months before biopsy or surgical resection to minimize temporal mismatch between imaging appearance and tissue staging. Fig. 2 presents representative examples for each stage, including F0 (no fibrosis), F1 (portal fibrosis without septa), F2 (portal fibrosis with few septa), F3 (numerous septa without cirrhosis), and F4 (cirrhosis).

A total of 6,323 ultrasound images were used in this work. The class distribution comprises 2,114 images for F0, 861 for F1, 793 for F2, 857 for F3, and 1,698 for F4, indicating a non-uniform distribution where F0 and F4 account for more than half of the samples, while intermediate stages are relatively underrepresented. This imbalance reflects practical clinical data characteristics and contributes to the difficulty of separating adjacent intermediate stages with subtle echotexture differences. To support fair evaluation, stratified splitting and stratified cross-validation were applied so that the original class proportions were preserved across training, validation, and test partitions.

B. Preprocessing Dataset

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

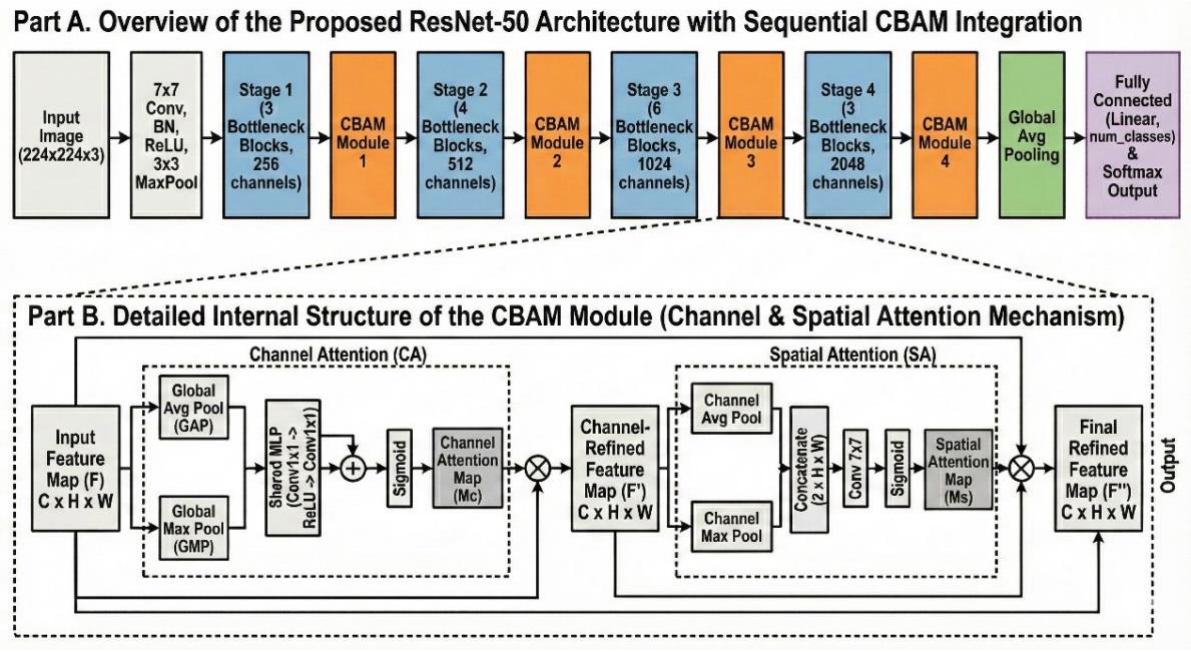


Fig. 3. Resnet 50 Architecture with CBAM.

All ultrasound images underwent preprocessing by resizing to 224 × 224 pixels to align with the ResNet-50 input format. Following resizing, ImageNet-based mean and standard deviation normalization was applied to standardize pixel intensity distributions throughout the training process.

Table 1. The dataset was divided into training, validation, and test sets.

Class	F0	F1	F2	F3	F4	Sum
Train and Val. (80%)	1,691	688	634	685	1,358	5,056
Test (20%)	423	173	159	172	340	1,267

The dataset was then divided using a stratified split, with 80% used for the training–validation set and 20% reserved for an independent test for each fibrosis class (F0–F4), as shown in Table 1. This stratification ensured that the class proportions in both subsets matched those of the original distribution. After this division, the training–validation split was further processed using 5-fold cross-validation. In this scheme, the training-validation set was partitioned into five balanced folds based on the class distribution. At each iteration, one subset was reserved for validation, while the remaining four formed the training data. Employing this approach strengthens the consistency of model evaluation and helps prevent overfitting, and allows the model to be assessed across multiple variations of the training subset. The test set was excluded from the K-Fold procedure and used only once at the final stage for independent performance evaluation. After Stratified 5-Fold training, five independently trained ResNet-50 plus CBAM models were obtained. During inference on the held-out test set, logits from the five models were averaged per class, and Softmax was applied to the averaged logits to obtain the final prediction. All headline metrics are reported for this logit-level ensemble, while fold-wise scores are provided to summarize variability across the five independently trained models. Several common augmentation operations were considered, but a conservative augmentation strategy was adopted to avoid distorting fibrosis-related echotexture cues. Following the dataset characteristics described in [24], the original ultrasound frames are sector-shaped, and overly aggressive geometric transforms may alter the anatomical context or introduce unrealistic boundaries. Therefore, we applied Random Horizontal Flip with probability 0.5 to increase geometric variability while preserving clinically plausible appearance. We did not apply random cropping because it may exclude diagnostically relevant parenchymal regions, and we avoided large rotations or strong intensity jitter because brightness and orientation can be acquisition dependent in ultrasound and may confound fibrosis texture interpretation. Validation and test images were processed only with resizing and normalization to ensure that the evaluation reflects performance on unmodified images. Meanwhile, the validation data in each fold, as well as the test data, underwent only resizing and normalization without additional augmentation to ensure that model evaluation

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

accurately reflected performance on unmodified images.

C. Model Design

The model design stage outlines the architecture employed to extract discriminative features and perform liver fibrosis stage classification from ultrasound images. In this study, a ResNet-50 architecture integrated with the Convolutional Block Attention Module (CBAM) is proposed, with the overall processing flow illustrated in Fig. 3. ResNet-50 is adopted as the backbone network due to its residual learning mechanism, which enables effective training of deep neural networks while mitigating the vanishing gradient problem [34]. In a standard residual block, the output feature map is obtained by combining the input feature map with the output of a residual mapping through skip connections. This residual learning operation is defined in Eq. (1) [34].

$$y = F(x, W) + x \quad (1)$$

where $x \in R^{H \times W \times C}$ denotes the input feature map of the residual block with spatial dimensions $H \times W$ and C channels, $F(x, W)$ represents the residual mapping learned by a series of convolutional layers parameterized by weights W, and y is the resulting output feature map. This formulation allows the network to preserve gradient flow by learning residual functions rather than direct mappings, facilitating deeper feature representation learning.

Despite its effectiveness, previous work reported indicates that the standard ResNet-50 architecture encounters difficulties in distinguishing intermediate liver fibrosis stages (F1–F3) [24]. This limitation is mainly attributed to the uniform treatment of feature channels, which may result in insufficient emphasis on subtle texture patterns critical for fibrosis assessment. To address this issue, CBAM modules are integrated into the ResNet-50 backbone, as illustrated in Fig. 3. Specifically, CBAM-module is inserted after the output of each residual stage, enabling stage-level feature recalibration before propagation to the subsequent stage. This design applies both channel-wise and spatial-wise attention across four semantic levels, ranging from low-level texture descriptors to high-level semantic representations, while maintaining a light weight architectural modification.

The internal structure of CBAM consist of sequential channel attention and spatial attention mechanisms. Given an intermediate feature map $F \in R^{H \times W \times C}$, the channel attention map $M_C(F)$ is computed as defined in Eq. (2) [25].

$$M_C(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F) \quad (2)$$

where $AvgPool(\cdot)$ and $MaxPool(\cdot)$ denote global average pooling and global max pooling operations applied along the spatial dimensions, respectively, producing two channel-wise descriptors of size $1 \times 1 \times C$. The function $MLP(\cdot)$ represents a shared multi-layer perceptron used to model inter-channel dependencies, and $\sigma(\cdot)$ denotes the sigmoid activation function. The resulting channel attention map $Mc(F)$ adaptively reweights each feature channel to emphasize informative channels while suppressing less ones. Following channel attention, spatial attention is applied to the channel refined feature map F′ to identify diagnostically relevant spatial regions. The spatial attention mechanism is defined in Eq. (3) [25].

$$M_S(F') = \sigma(\text{f\^7x7}([AvgPool(F'); MaxPool(F')])) \quad (3)$$

where $AvgPool(\cdot)$ and $MaxPool(\cdot)$ are applied along the channel dimension to generate two spatial feature maps of size $H \times W \times 1$, which are then concatenated along the channel axis, denoted by $[\cdot;\cdot]$. The function $f^{7\times7}(\cdot)$ represents a convolution operation with a 7×7 kernel size, and $\sigma(\cdot)$ is the sigmoid activation function. This spatial attention mechanism enables the network to focus on diagnostically relevant spatial regions associated with fibrosis-related patterns.

By integrating CBAM into the residual learning framework, the output of the enhanced residual block is formulated as shown in Eq. (4) [25].

$$y = CBAM(F(x, W) + x \quad (4)$$

where $CBAM(\cdot)$ denotes the sequential application of channel attention and spatial attention operations. This integration allows the network to selectively enhance discriminative features while retaining the advantages of residual learning, thereby improving the representation capability for liver fibrosis stage classification. As shown in Fig. 3, the input ultrasound image with a resolution of 224 × 224 × 3 pixels is first processed by an initial convolutional layer with a 7 × 7 kernel and a stride of 2, followed by batch normalization, a ReLU activation function, and a 3 × 3 max pooling layer. The resulting feature maps are then propagated through four residual stages (conv2_x to conv5_x), with CBAM embedded at each stage to enhance feature refinement. Finally, the extracted feature maps are aggregated using Global Average Pooling (GAP) and passed to a fully connected layer with five output neurons corresponding to fibrosis stages F0–F4. During training, the CrossEntropyLoss function is employed, which implicitly incorporates a Softmax operation to produce class probability distributions. Within CBAM, Channel Attention emphasizes important feature channels by integrating spatial cues obtained from average and max pooling. Spatial Attention, on the other hand, captures informative spatial regions by applying channel-wise pooling, followed by a 7 × 7 convolution to calculate the spatial attention features. Through the combined use of channel-wise and spatial attention, CBAM improves the model's focus on informative features, thereby

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

increasing its responsiveness to subtle fibrosis characteristics. Following the CBAM residual blocks, the extracted feature maps undergo Global Average Pooling (GAP) before being passed to a fully connected layer featuring five output nodes representing fibrosis stages F0 through F4. The training process adopts CrossEntropyLoss, which implicitly includes a Softmax operation to generate class probability scores.

D. Hyperparameter Tuning

Hyperparameter tuning was conducted as a controlled ablation-style grid evaluation. The learning rate was tested in 1e−3, 1e−4, 1e−5, and weight decay in 1e−4, 1e−5; activation functions ReLU, GeLU, and loss functions CrossEntropyLoss, Sparse Categorical Crossentropy were compared; and attention variants none and CBAM were evaluated under the same Stratified 5-Fold protocol. The final configuration was selected based on consistently strong fold-wise performance and stable convergence behavior rather than a single peak result. To achieve this, several critical parameters, including learning rate, weight decay, activation functions, loss functions, and attention mechanism variants, are iteratively adjusted. The tuning strategy also incorporates comparative evaluation between the baseline model and the attention-enhanced architecture to determine the most performant setup. This comparative approach ensures that performance improvements can be attributed to the integration of attention mechanisms rather than to incidental parameter variations.

E. Model Evaluation

In this study, the evaluation process employs the test set from each dataset to assess the classification performance. Four primary metrics are used to measure model effectiveness: Accuracy, Precision, Recall, and F1-Score.

1. Accuracy

Accuracy measures the overall proportion of correct predictions over all test samples. As shown in Eq. (5) [12], the numerator $\sum_{c=1}^{C} n_{cc}$ sums the diagonal elements of the confusion matrix, i.e., the total number of correctly classified samples across all classes. Dividing by N yields the fraction of correct predictions among all test images. Accuracy provides a global summary of classification correctness, but it can be influenced by class imbalance; therefore, we also report macro and weighted class-wise metrics to evaluate performance more fairly across stages.

$$Accuracy = \frac{\sum_{c=1}^{C} n_{cc}}{N} \quad (5)$$

2. Precision

Precision evaluates the model's tendency toward over-prediction, measured as the proportion of correctly predicted samples for class c among all samples predicted as class c, correctly identified out of all voxels classified as positive. [35]. Precision evaluates the accuracy of positive predictions by measuring the proportion that are genuinely relevant, as defined in Eq. (6) [12]. The macro precision metric is implemented to achieve a balanced assessment, ensuring that every class is equally represented in performance metrics, as described in Eq. (7) [12]. When the dataset exhibits class imbalance, weighted precision, defined in Eq. (8) [12], is more appropriate because it accounts for the number of samples in each class.

$$Precision_C = \frac{TPc}{TPc + FPc} \quad (6)$$

$$Precision_{\text{macro}} = \frac{1}{5}\sum_{c=1}^{5} Precision_c \quad (7)$$

$$Precision_{\text{weighted}} = \frac{1}{N}\sum_{c=1}^{5} n_c \; Precision_c \quad (8)$$

3. Recall

Recall (also called sensitivity or true positive rate) measures the ability of the model to correctly detect samples that truly belong to a given class. In fibrosis staging, recall is important for quantifying the classifier's tendency to miss a stage (under-detection), especially for intermediate stages that are visually subtle. Class-wise recall in Eq. (9) [12] is defined using TP_C as the number of classes-c samples predicted correctly and FN_C as the number of true classes-c samples misclassified as other classes, so it measures the proportion of ground-truth class-c samples that the model successfully retrieves.

$$Recall_c = \frac{TP_c}{TP_c + FN_c} \quad (9)$$

Macro recall (Eq. (10) [12]) is the unweighted mean of recall across all classes, giving each stage equal importance. This is essential for assessing whether the classifier consistently detects minority and intermediate classes rather than achieving high performance only on frequent classes.

$$Recall_{\text{macro}} = \frac{1}{5}\sum_{c=1}^{5} Recall_c \quad (10)$$

$$Recall_{\text{weighted}} = \frac{1}{\mathrm{N}}\sum_{\mathrm{c=1}}^{5} \mathrm{n_c} \; \mathrm{Recall_c} \quad (11)$$

Weighted recall (Eq. (11) [12]) weights each class recall by n_c and normalizes by N, producing a recall estimate that reflects the empirical distribution of the dataset. Like weighted precision, it can be dominated by larger classes, so it is reported together with macro recall.

4. F1-Score

The F1-Score provides a single class-wise measure

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

that balances precision and recall. This is particularly useful in multi-class medical classification because a model can achieve high precision by being conservative (but miss many true cases), or high recall by being permissive (but produce many false positives). F1 penalizes such imbalanced behavior. Class-wise F1 in Eq. (12) [12] is the harmonic mean of $Precision_C$ and $Recall_C$. Because it is a harmonic mean, the score becomes high only when both precision and recall are high; if either one is low, $F1_C$ decreases substantially.

$$F1_c = \frac{2 \cdot Precision_c \cdot Recall_c}{Precision_c + Recall_c} \quad (12)$$

Macro F1 (Eq. (13) [12]) is the unweighted mean of $F1_C$, across all classes. This metric assigns equal importance to each class, regardless of the number of samples in each category. As a result, macro F1 is particularly suitable for imbalanced medical datasets, where minority classes, such as intermediate fibrosis stages, are clinically important and should not be overshadowed by the majority classes.

$$F1_{\text{macro}} = \frac{1}{5}\sum_{c=1}^{5} F1_c \quad (13)$$

Weighted F1 (Eq. (14) [12]) computes a weighted average of the class-wise F1-Scores $F1_C$, where each class is weighted by its number of samples n_C and normalized by the total number of samples N. This metric reflects class distribution and overall performance under imbalance, but is more influenced by majority classes and may underrepresent minority fibrosis stages.

$$F1_{\text{weighted}} = \frac{1}{N}\sum_{c=1}^{5} n_c \; F1_c \quad (14)$$

III. Result

A. Hyperparameter Tuning

Hyperparameter tuning was performed to determine the most effective training configuration for the proposed ResNet-50 + CBAM model. This optimization stage aimed to ensure stable convergence, maximize generalization, and identify hyperparameters that consistently improved performance across all evaluation folds. Several key components, including learning rate, weight decay, activation function, loss function, and attention mechanism, were systematically examined to assess their individual and collective influence on model performance [36]. The tuning process was conducted in a controlled, reproducible manner to minimize the risk of biased model selection and to ensure that observed performance gains were attributable to architectural and parameter choices rather than random fluctuations. Each hyperparameter was evaluated while keeping other settings fixed to isolate its specific impact on training stability and classification accuracy. This strategy enabled a fair comparison across configurations across all validation folds. In addition, the selection of the final hyperparameter set considered not only peak performance but also consistency across folds, reflecting the model's robustness. Such a systematic tuning procedure is essential for deep learning models applied to medical imaging tasks, where reliable generalization is critical for practical applicability. Thus, the final hyperparameters were fixed for all subsequent experiments to ensure a fair and reproducible evaluation.

The first experiment jointly evaluated the influence of learning rate and weight decay, as both parameters play a critical role in controlling the magnitude of weight updates and regularization strength during training. Three learning rates (1e−3, 1e−4, 1e−5) and two weight decay values (1e−4, 1e−5) were tested. The results showed that a learning rate of 1e−3 produced unstable validation accuracy and oscillatory loss, while 1e−4 improved stability but still exhibited minor fluctuations. A learning rate of 1e−5 yielded the most stable convergence and the lowest validation loss. For weight decay, 1e−4 offered superior regularization, yielding smoother curves and better generalization compared to 1e−5. Based on these observations, the optimal combination selected for the final model was a learning rate of 1e−5 and a weight decay of 1e−4. The experimental results are shown in Fig. 4. Insights from tuning.

Table 2. Comparative analysis of activation functions and loss functions

Activation Fuction	Macro average (%)	Weighted average (%)	Accuracy (%)
ReLU	97.64	98.34	**98.34**
GeLU	93.67	95.41	95.42
Loss Fuction	Macro average (%)	Weighted average (%)	Accuracy (%)
Sparse Categorical Crossentropy	94.40	95.63	95.66
Cross Entropy Loss	98.20	98.34	**98.34**

The tuning study indicates that learning rate is the most sensitive hyperparameter for this task. A larger rate produces oscillatory validation loss, consistent with noisy gradients when learning from speckle-dominated ultrasound textures, whereas a smaller rate yields

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

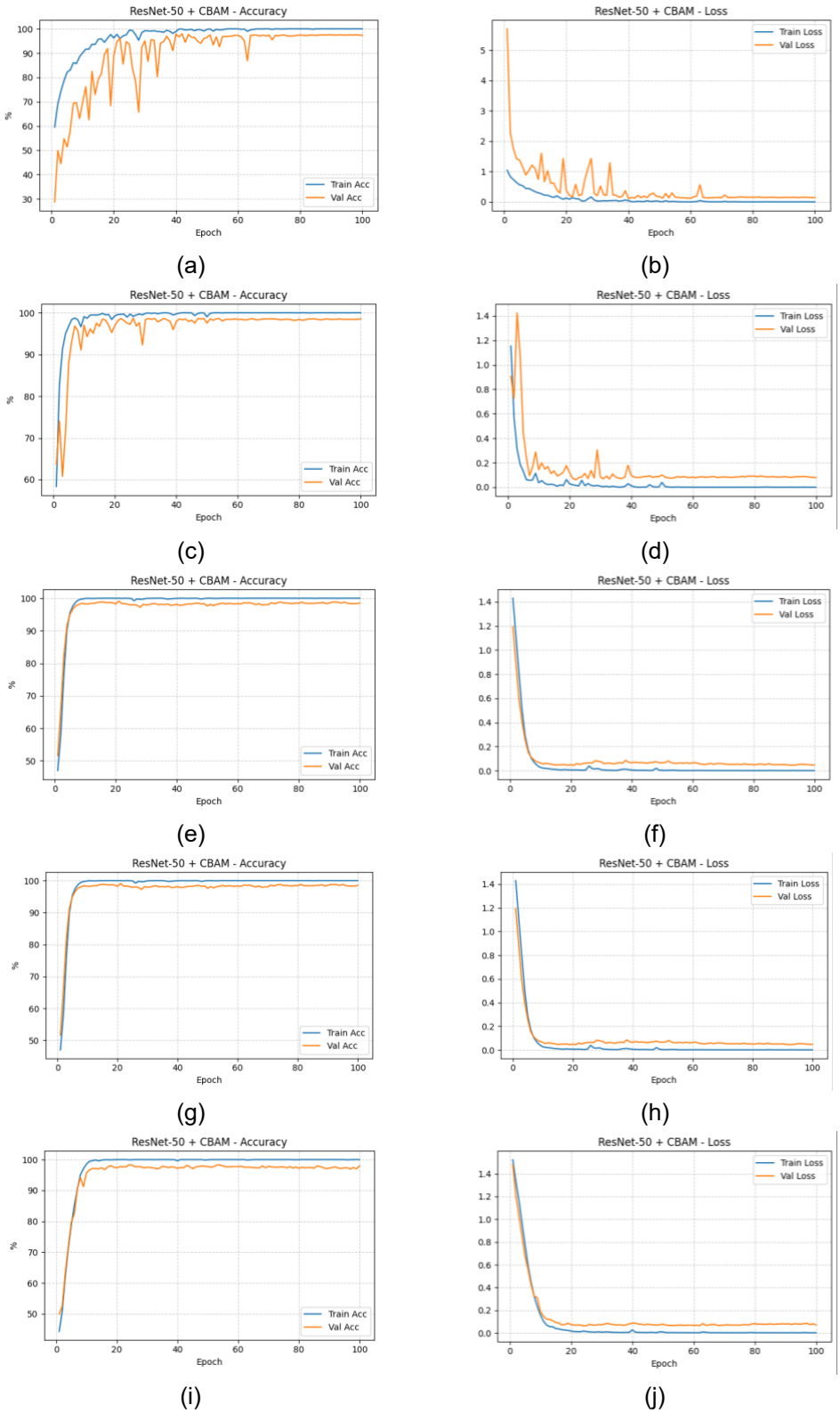


Fig. 1. Hyperparameter tuning curves: (a) accuracy at LR = 1e−3, (b) loss at LR = 1e−3, (c) accuracy at LR = 1e−4, (d) loss at LR = 1e−4, (e) accuracy at LR = 1e−5, (f) loss at LR = 1e−5, (g) accuracy at weight decay = 1e−4, (h) loss at weight decay = 1e−4, (i) accuracy at weight decay = 1e−5, and (j) loss at weight decay = 1e−5.

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

Table 3. Model experiment results.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet 50	84.98	F0: 0.921 F1: 0.760 F2: 0.801 F3: 0.824 F4: 0.852	F0: 0.858 F1: 0.814 F2: 0.861 F3: 0.824 F4: 0.864	F0: 0.888 F1: 0.786 F2: 0.830 F3: 0.824 F4: 0.858
ResNet 50 + CBAM	**98.34**	F0: 1.000 F1: 0.959 F2: 0.962 F3: 0.982 F4: 0.985	F0: 1.000 F1: 0.965 F2: 0.956 F3: 0.953 F4: 1.000	F0: 1.000 F1: 0.962 F2: 0.959 F3: 0.967 F4: 0.992

important role by reducing overfitting to scanner-specific intensity patterns and improving fold consistency. These observations provide practical guidance for future ultrasound-based staging studies using similar backbones.

The next experiment combined the evaluation of the activation function and the loss function, as both directly influence gradient flow and the learning behavior of the model. ReLU and GeLU were compared, with ReLU demonstrating distinctly superior performance across macro average, weighted average, and accuracy metrics. Its stability and sparsity inducing properties made it more suitable for extracting discriminative texture features from ultrasound images. Two loss functions, CrossEntropyLoss and Sparse Categorical Crossentropy, were also tested. CrossEntropyLoss consistently achieved higher accuracy and more stable convergence, attributed to its direct operation on logits and its increased robustness to imbalanced datasets. As a result, ReLU was selected as the activation function and CrossEntropyLoss as the primary loss function for all subsequent training processes. The experimental results are shown in Table 2.

The subsequent experiment evaluated the effect of integrating an attention mechanism into the ResNet-50 backbone, followed by an assessment of overall classification performance, as summarized in Table 3. The comparison focuses on the baseline ResNet-50 model and the proposed ResNet-50 enhanced with the Convolutional Block Attention Module (CBAM). As shown in the table, incorporating CBAM results in substantial improvements across all evaluation metrics. The overall accuracy increases from 84.98% to 98.34%, accompanied by consistent gains in precision, recall, and F1-score across all fibrosis stages. Notably, the most pronounced improvements are observed in the intermediate stages F1–F3, where the baseline model exhibits lower discrimination capability due to subtle and overlapping ultrasound texture patterns. The dual attention mechanism in CBAM enables adaptive refinement along both channel and spatial dimensions, allowing the network to emphasize diagnostically relevant texture cues while suppressing acquisition-related noise. Compared with the baseline ResNet-50, the CBAM-augmented model demonstrates stronger generalization behavior and more stable predictions across all fibrosis categories. These results confirm that integrating CBAM into the ResNet-50 backbone yields a robust, discriminative architecture for accurate liver fibrosis classification from ultrasound images.

Table 3. Comparison of Test Accuracy and Loss in Every Fold.

ResNet-50	Test Accuracy (%)	Test Loss
Fold 1	84.90	0.586
Fold 2	85.10	0.571
Fold 3	85.00	0.579
Fold 4	84.80	0.593
Fold 5	85.10	0.569
ResNet-50 + CBAM	**Test Accuracy (%)**	**Test Loss**
Fold 1	97.71%	0.088
Fold 2	96.92%	0.113
Fold 3	97.32%	0.102
Fold 4	97.24%	0.107
Fold 5	96.69%	0.122

To assess performance robustness across different data partitions, fold-wise accuracy results obtained from the Stratified 5-Fold Cross Validation are summarized in Table 4. The proposed ResNet-50 + CBAM model demonstrates consistently high accuracy across all folds, accompanied by a low standard deviation, indicating stable generalization behavior. In contrast, the baseline ResNet-50 exhibits larger variability across folds. These findings confirm that the observed performance improvement is not driven by a single

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

favorable split but is consistently maintained across different validation folds. In particular, Table 4 shows that the ResNet-50 + CBAM model maintains test accuracy above 96.6% in every fold, whereas the baseline ResNet-50 remains around 84.8–85.1% across folds. The corresponding test losses follow the same trend, with substantially lower values for ResNet-50 + CBAM (0.0882–0.1222) compared to the baseline (0.5694–0.5931), indicating more confident and stable predictions. The narrow spread of both accuracy and loss across folds suggests that the model behavior is not sensitive to a particular partition, supporting stable generalization under stratified sampling. This fold-consistent improvement is consistent with the intended role of CBAM, which repeatedly refines features at multiple residual stages and helps preserve subtle texture cues that can be diluted during downsampling.

Therefore, the cross-validation evidence in Table 4 provides a strong basis for the subsequent statistical test, since the observed gains reflect a reproducible pattern rather than isolated fold-specific effects. A paired test was conducted on fold-wise test accuracies to quantify statistical significance under the stratified 5-fold protocol. The baseline ResNet 50 achieved 84.98% ± 0.13% accuracy, whereas ResNet 50 with CBAM achieved 97.18% ± 0.39% accuracy, yielding a mean improvement of 12.20 percentage points with a 95% confidence interval from 11.59 to 12.80. The paired test confirms that the improvement is statistically significant, with t = 55.65 and p = 6.24 × 10^-7.

B. Experiment Result

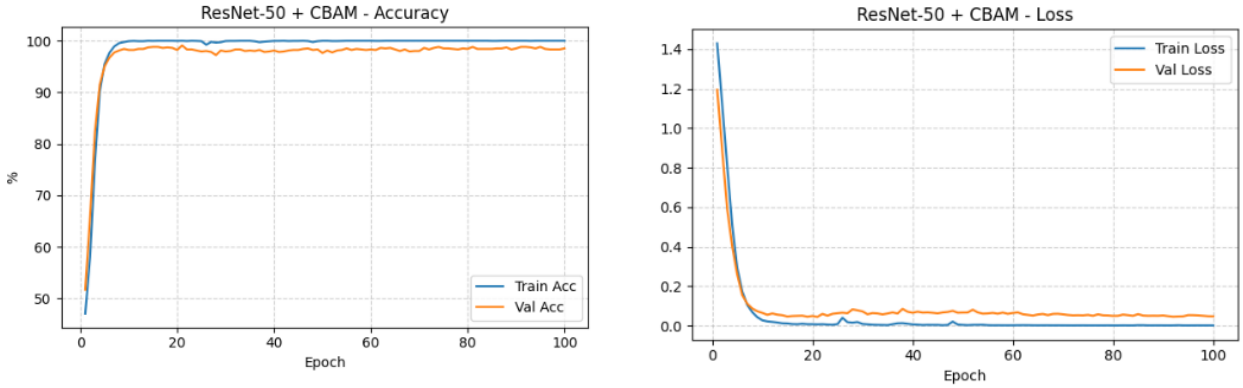
The final model training was performed using the optimal hyperparameter configuration obtained from the tuning process. This configuration was selected due to its ability to support stable convergence, control overfitting, and achieve consistently high performance across all validation folds. The finalized training configuration strikes a balance between learning efficiency and generalization, ensuring the model performs effectively across diverse patterns of liver fibrosis. The complete set of selected hyperparameters used for the final training stage is shown in Table 5.

Table 5. Hyperparameter Tuning.

Hyperparameter	Value
Batch Size	64
Epoch	100
Learning Rate	1e-5 (0.00001)
Optimizer	Adam (Adaptive Moment Estimation)
Scheduler	CosineAnnealingLR
Activation Function	ReLU
Loss Function	CrossEntropyLoss
Weight Decay	1e-4 (0.0001)
K-Fold Validation	5-Fold Stratified

Across all epochs, the ResNet-50 model integrated with the Convolutional Block Attention Module (CBAM) exhibits smooth, stable learning dynamics. The accuracy trends reveal rapid improvement in both training and validation performance at the beginning of training, followed by convergence at high accuracy levels. The close correspondence between these curves suggests strong generalization capability with negligible overfitting. This pattern suggests that incorporating the Convolutional Block Attention Module (CBAM) effectively enables the network to capture discriminative ultrasound features across distinct stages of liver fibrosis.

In addition to the accuracy curve, there are also loss curves corresponding to training and validation. The loss values decrease smoothly and maintain a narrow gap throughout training, reflecting a healthy optimization process with stable gradient updates. The validation loss does not exhibit irregular spikes or divergence, further confirming that the model maintains balanced learning



(a) (b)

Fig. 5. The best (a) accuracy curve and (b) loss curve results of ResNet-50 with CBAM.

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

dynamics, indicating a well-balanced training process free from overfitting and underfitting [37]. Overall, the accuracy and loss curves shown in Fig. 5 provide strong evidence that the ResNet-50 + CBAM architecture achieves efficient convergence and robust performance. The consistency observed across both metrics highlights the model's reliability and strengthens the conclusion that the proposed approach is well-suited for liver fibrosis classification using ultrasound imaging. Importantly, the narrow gap between the training and validation curves indicates that the model generalizes well, remembering rather than memorizing fold-specific patterns, which is critical given the heterogeneous multi-vendor acquisition conditions of the dataset. This stable optimization behavior also supports that CBAM guided feature recalibration helps the network focus on diagnostically relevant echotexture cues while suppressing noise-dominated responses, thereby reducing overfitting risk and improving robustness across different data partitions. Moreover, the selected hyperparameters support CBAM-driven feature learning, enabling stable optimization, and good generalization. Consequently, the final model shows consistent performance across the stratified folds, reinforcing its suitability for ultrasound settings. The confusion matrix in Fig. 6 (logit-level ensemble on the independent test set, N = 1,267) provides a class-wise view beyond aggregate accuracy. Correct predictions dominate the diagonal, with 1,246/1,267 samples correctly classified, leaving only 21 misclassifications (1.66%). The extreme stages F0 and F4 achieve perfect sensitivity (FN = 0 for both), indicating that the model reliably detects the absence of fibrosis and advanced cirrhosis.

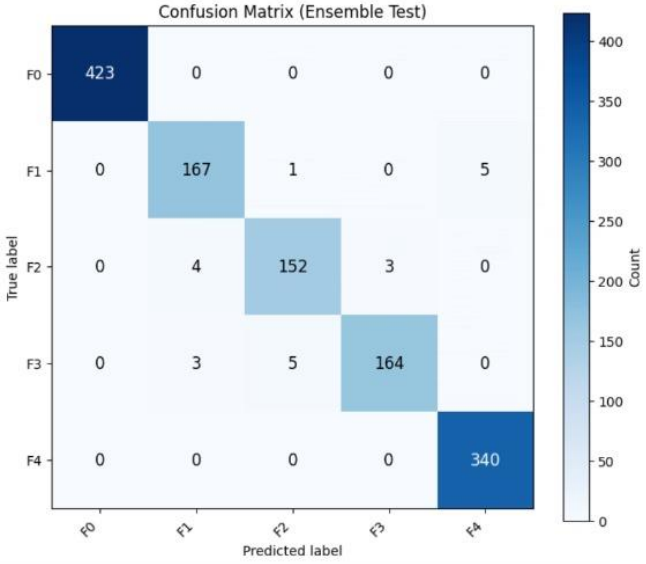


Fig. 6. Confusion matrix ensemble test ResNet-50 with CBAM.

Remaining errors are concentrated in the intermediate stages (F1–F3), consistent with the known visual ambiguity of adjacent METAVIR categories in B-mode ultrasound. Specifically, the model yields FN = 6 for F1, FN = 7 for F2, and FN = 8 for F3, showing that most residual mistakes occur within the clinically challenging mid-spectrum. False positives are also limited (FP = 7 for F1, FP = 6 for F2, FP = 3 for F3), indicating that the decision boundaries are not overly permissive for intermediate classes. Importantly, predictions as F4 remain conservative, with only 5 false positives, while maintaining FN = 0 for F4. Overall, the confusion matrix supports the claim that the proposed model reduces severe staging errors and keeps residual misclassifications largely within the intermediate range, which is clinically more plausible than extreme jumps. Because the confusion matrix is computed from a logit-level ensemble across five independently trained folds, these behaviors reflect stable prediction tendencies rather than a single favorable split.

Table 6. Fibrosis Stage Clinical Metrics.

Class	F0	F1	F2	F3	F4
Sens	1.000	0.965	0.956	0.953	1.000
Spec	1.000	0.993	0.994	0.997	0.994
PPV	1.000	0.959	0.962	0.982	0.985
NPV	1.000	0.994	0.993	0.992	1.000
TP	423	167	152	164	340
FP	0	7	6	3	5
FN	0	6	7	8	0
TN	844	1087	1102	1092	922

Clinical Metrics every Class show in Table 6. Beyond overall accuracy, stage wise clinical metrics derived from the ensemble confusion matrix provide a more meaningful assessment of diagnostic reliability for liver fibrosis staging. Sensitivity values are high at the extreme stages F0 and F4 and remain high across intermediate stages, indicating that the proposed model consistently identifies fibrosis presence and severity across the disease spectrum. Specificity values exceeding 0.99 for all classes demonstrate a low rate of false positives, while high positive and negative predictive values further confirm the reliability of class assignments at the individual stage level. These values are consistent with the confusion matrix, where false negatives are limited to intermediate stages (FN = 6 - 8 for F1 - F3) and false positives remain low across all classes (FP ≤ 7). Clinically, such errors are more likely to occur near staging boundaries within the intermediate spectrum, reinforcing that the model should be used as decision support alongside clinical assessment rather than as a standalone diagnosis.

The preservation of strong sensitivity and specificity for stages F1–F3 is particularly important given the

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

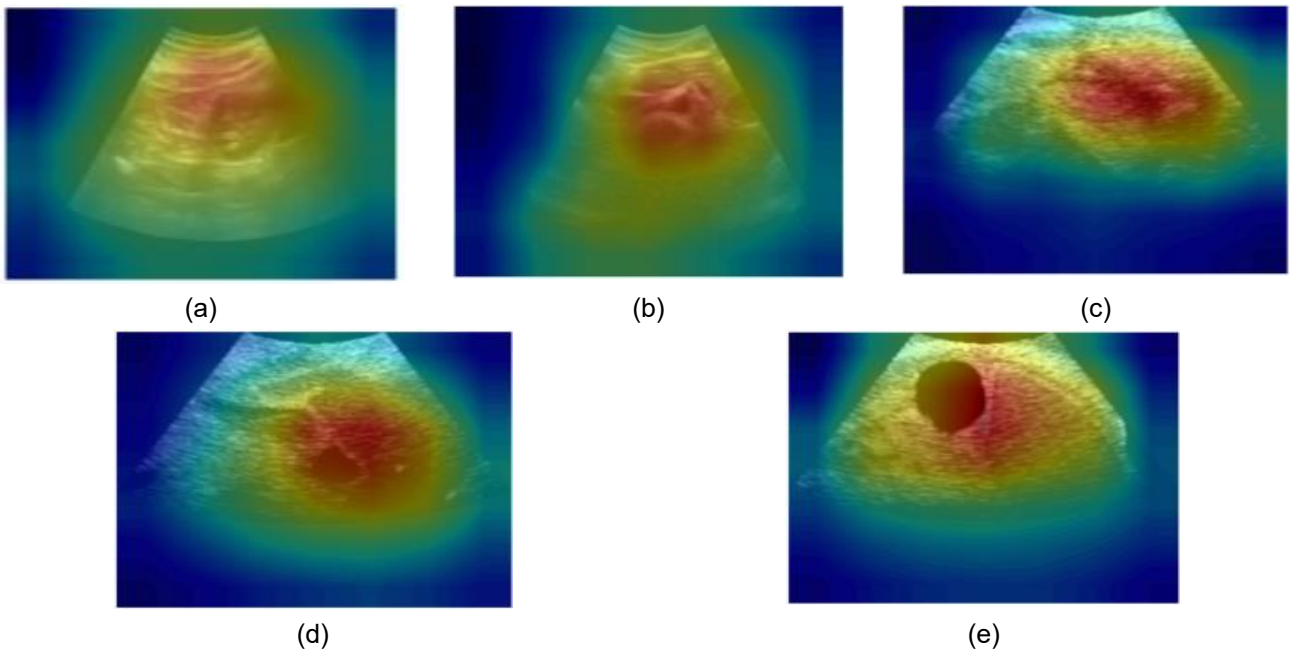


Fig. 7. GRAD-CAM visualization (a) F0, (b) F1, (c) F2, (d) F3, (e) F4.

known diagnostic ambiguity of intermediate fibrosis on ultrasound imaging. These results indicate that the proposed approach does not achieve high accuracy by favoring majority classes, but instead maintains balanced performance across all fibrosis stages. From a clinical standpoint, this behavior reduces the likelihood of systematic under staging or over staging and supports the suitability of the model as a non-invasive decision support tool for fibrosis assessment and longitudinal monitoring.

Grad-CAM visualizations are presented in Fig. 7. To provide insight into the internal decision making behavior of the proposed model, Grad-CAM visualizations were generated for representative test images across all fibrosis stages. These visualizations qualitatively indicate that the CBAM-enhanced network produces discriminative responses concentrated within the liver parenchymal region inside the ultrasound sector, rather than focusing on background areas or acquisition-related artifacts. This observation is consistent across fibrosis stages, including intermediate categories where texture differences are subtle and spatially diffuse.

It is emphasized that these visualizations are presented for qualitative transparency rather than quantitative validation, as the dataset provides image-level labels without pixel-level ground truth annotations. Consequently, the attention maps are not interpreted as precise lesion localization. Instead, they offer supporting evidence that attention-based feature refinement guides the network toward anatomically relevant regions during classification, aligning model behavior with clinical expectations while acknowledging known limitations of Grad-CAM under large receptive fields and heterogeneous ultrasound acquisition conditions.

Empirically, the influence of CBAM is reflected in the improved stability of the training process, the consistency of the accuracy and loss curves, and the final evaluation metrics. The learning curve demonstrates that the model attains near-perfect training accuracy and maintains high validation accuracy without signs of overfitting. The small discrepancy between the training and validation curves indicates that the attention mechanism successfully guides the network toward diagnostically informative features, thereby strengthening generalization performance. In addition to quantitative gains, CBAM improves the interpretability of the feature extraction process by encouraging the model to concentrate on clinically meaningful tissue patterns. Previous studies have similarly demonstrated that attention mechanisms enhance model explainability by explicitly highlighting salient anatomical regions, offering clinicians greater transparency into deep learning–based decision processes [38]. Furthermore, ensemble evaluation on the CBAM-enhanced ResNet-50 test set corroborated these benefits, achieving an accuracy of 98.34% alongside consistently high precision, recall, and F1-scores across all fibrosis stages (F0–F4), as shown in Table 7. These results collectively demonstrate that incorporating CBAM significantly improves feature extraction, increases sensitivity to fibrosis-related texture patterns, and results in more accurate and consistent classification performance in ultrasound-based liver fibrosis assessment.

C. Analysis of Attention Mechanism

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

The proposed approach outperforms previous architectures in Table 7, showing that earlier models trained on the same dataset, including VGGNet, DenseNet, EfficientNet, ViT, and the baseline ResNet 50, achieve lower accuracy. We attribute the improvement of the proposed method to three complementary design choices. First, CBAM provides sequential channel and spatial recalibration, which helps preserve weak echotexture cues that can be diluted by downsampling and uniform aggregation. Second, the conservative augmentation strategy avoids introducing unrealistic texture artifacts that can harm generalization on heterogeneous ultrasound data. Third, logit ensembling across independently trained fold models reduces variance and stabilizes predictions, particularly at borderline intermediate stages. Together, these factors explain why the proposed approach achieves a larger performance gain on this multi-center, multi-vendor dataset. Furthermore, the precision, recall, and F1-scores were consistently high across all fibrosis stages, including the more challenging intermediate categories. These results suggest that the inclusion of channel-level and spatial attention mechanisms enhance the model's ability to selectively attend to diagnostically relevant information, thereby decreasing misclassification commonly observed in fibrosis stages F1 to F3. The ability to maintain strong performance across all classes, rather than only at extreme stages (F0 and F4), is a significant advancement over earlier CNN-based approaches. Similar improvements using contrastive fusion ultrasound models have recently been reported, where deep learning achieved stable staging performance even under small sample and heterogeneous image conditions, corroborating the robustness of our attention-enhanced approach [26]. Comparable performance improvements have also been documented in other medical classification tasks, where CBAM-based architectures consistently outperform standard CNN models in extracting discriminative features from low contrast and heterogeneous imaging data [39], [40].

IV. Discussion

This study presents a detailed evaluation of the proposed ResNet-50 integrated with the CBAM and compares its performance with several established deep learning architectures reported in previous studies. Based on the comparative results shown in Table 7, earlier work by [24] demonstrated that conventional CNN models, including VGGNet, DenseNet, EfficientNet, and even Vision Transformer, achieved varying accuracy levels for liver fibrosis staging, with standard ResNet-50 yielding the highest performance at 85.92%. Although this accuracy was better than that of other architectures, their study also highlighted that ResNet-50 struggled to reliably classify intermediate fibrosis stages (F1–F3), where parenchymal texture differences are subtle and often ambiguous in ultrasound images. These findings indicate that conventional CNNs are still limited in capturing fine-grained diagnostic features required for precise fibrosis assessment [38].

The proposed approach outperforms previous architectures in Table 7, showing that earlier models trained on the same dataset, including VGGNet, DenseNet, EfficientNet, ViT, and the baseline ResNet 50, achieve lower accuracy. We attribute the improvement of the proposed method to three complementary design choices. First, CBAM provides sequential channel and spatial recalibration, which helps preserve weak echotexture cues that can be diluted by downsampling and uniform aggregation. Second, the conservative augmentation strategy avoids introducing unrealistic texture artifacts that can harm generalization on heterogeneous ultrasound data. Third, logit ensembling across independently trained fold models reduces variance and stabilizes predictions, particularly at borderline intermediate stages. Together, these factors explain why the proposed approach achieves a larger performance gain on this multi-center, multi-vendor dataset. Furthermore, the precision, recall, and F1-scores were consistently high across all fibrosis stages, including the more challenging intermediate categories. These results suggest that the inclusion of channel-level and spatial attention mechanisms enhances the model's ability to selectively attend to diagnostically relevant information, thereby decreasing misclassification commonly observed in fibrosis stages F1 to F3. The ability to maintain strong performance across all classes, rather than only at extreme stages (F0 and F4), is a significant advancement over earlier CNN-based approaches. Similar improvements using contrastive fusion ultrasound models have recently been reported, where deep learning achieved stable staging performance even under small-sample and heterogeneous image conditions, corroborating the robustness of our attention-enhanced approach [26]. Comparable performance improvements have also been documented in other medical classification tasks, where CBAM-based architectures consistently outperform standard CNN models in extracting discriminative features from low contrast and heterogeneous imaging data [39], [40].

The improvements observed in this study suggest that attention mechanisms play a crucial role in enhancing feature representation in medical ultrasound applications, where noise, variability, and subtle texture patterns pose inherent challenges. The attention modules guide the model to focus on areas within the image that contain meaningful fibrosis-related characteristics, while suppressing irrelevant or

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

Table 7. Comparison of the proposed method with other methods with the same dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
VGG Net [24]	83.17	F0: 0.863 F1: 0.784 F2: 0.860 F3: 0.738 F4: 0.842	F0: 0.863 F1: 0.775 F2: 0.750 F3: 0.705 F4: 0.916	F0: 0.863 F1: 0.779 F2: 0.801 F3: 0.721 F4: 0.879
Dense Net [24]	84.17	F0: 0.869 F1: 0.793 F2: 0.874 F3: 0.793 F4: 0.836	F0: 0.892 F1: 0.818 F2: 0.750 F3: 0.666 F4: 0.913	F0: 0.881 F1: 0.806 F2: 0.807 F3: 0.724 F4: 0.873
Efficient Net [24]	85.17	F0: 0.880 F1: 0.850 F2: 0.848 F3: 0.785 F4: 0.848	F0: 0.875 F1: 0.818 F2: 0.831 F3: 0.750 F4: 0.895	F0: 0.878 F1: 0.834 F2: 0.839 F3: 0.672 F4: 0.871
ViT [24]	83.42	F0: 0.820 F1: 0.845 F2: 0.875 F3: 0.781 F4: 0.850	F0: 0.900 F1: 0.787 F2: 0.810 F3: 0.641 F4: 0.876	F0: 0.858 F1: 0.815 F2: 0.842 F3: 0.704 F4: 0.863
ResNet-50 [24]	85.92	F0: 0.874 F1: 0.831 F2: 0.921 F3: 0.754 F4: 0.878	F0: 0.897 F1: 0.831 F2: 0.790 F3: 0.750 F4: 0.907	F0: 0.886 F1: 0.831 F2: 0.850 F3: 0.752 F4: 0.892
Proposed Method	98.34	F0: 1.000 F1: 0.959 F2: 0.962 F3: 0.982 F4: 0.985	F0: 1.000 F1: 0.965 F2: 0.956 F3: 0.953 F4: 1.000	F0: 1.000 F1: 0.962 F2: 0.959 F3: 0.967 F4: 0.992

misleading features. This selective focus contributes not only to increased accuracy but also to improved stability, as reflected by the consistent training and validation curves. These trends align with previous studies suggesting that employing attention-based deep learning techniques can considerably boost model effectiveness in tasks involving visually ambiguous or low contrast patterns [41].

Robustness under class imbalance. The dataset is imbalanced, with F0 and F4 accounting for more than half of the samples. We mitigate this by stratified splitting, stratified 5-fold training, and reporting macro averaged metrics that treat all stages equally. Importantly, improvements from CBAM are not limited to majority classes. The baseline ResNet 50 yields F1 scores of 0.786 for F1, 0.830 for F2, and 0.824 for F3, whereas ResNet 50 plus CBAM reaches 0.962, 0.959, and 0.967 for these stages, respectively. This consistent gain suggests that attention-based recalibration improves representation of intermediate stages rather than simply amplifying recognition of the easiest classes. The superior performance of the proposed model also indicates its potential suitability for clinical decision support systems. With high accuracy and low variance across classes, the ResNet-50 + CBAM architecture demonstrates robustness even when ultrasound images exhibit subtle or overlapping fibrosis patterns. Such improvements are particularly important because accurate staging of intermediate fibrosis levels is critical for determining

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

treatment strategies and monitoring disease progression. Therefore, the enhanced performance achieved in this study confirms that incorporating CBAM effectively addresses the limitations of earlier methods and offers a more reliable solution for fibrosis classification.

In summary, the comparative findings show that the proposed model not only surpasses existing approaches in accuracy but also demonstrates stronger generalization and interpretability. The consistent results across all fibrosis stages reinforce that CBAM provides meaningful enhancements to the feature extraction process. Overall, these results position the ResNet-50 + CBAM model as a promising architecture for future research and potential clinical applications in ultrasound-based liver fibrosis assessment. Although the findings are positive, several constraints should be noted. Primarily, the data were collected from a relatively small number of institutions, which may constrain the generalizability of the proposed model across diverse ultrasound devices and clinical environments. Second, the evaluation was conducted on retrospective data; real-time prospective testing is still required. Third, the model relies solely on B-mode ultrasound without integrating clinical variables that may improve staging accuracy. Considering recent successes in fusion-based ultrasound approaches, future work may benefit from integrating contrast-enhanced ultrasound or combining ultrasound with elastography or other modalities to enhance staging reliability [26].

Despite these limitations, the high accuracy and consistent performance demonstrated by the proposed model indicate strong potential for its integration into routine clinical workflows as a reliable decision support tool for radiologists and hepatologists. The model's ability to maintain robust classification results across all fibrosis stages suggests that the incorporation of attention mechanisms effectively enhances feature representation and reduces diagnostic variability commonly encountered in ultrasound-based liver assessment. This feature is highly advantageous in medical settings, where accurate and standardized assessment is required, informed clinical decision making. This indicates that the proposed approach may support more consistent fibrosis evaluation in clinical practice. Overall, these findings confirm that attention based deep learning architectures provide substantial advancements in ultrasound-based liver fibrosis assessment by improving sensitivity to subtle and heterogeneous texture patterns. The demonstrated effectiveness of the proposed approach highlights the potential of attention mechanisms to address limitations of conventional convolutional neural networks. Consequently, this study may serve as a foundation for future research aimed at developing clinically reliable, interpretable, and non-invasive diagnostic systems, as well as facilitating the broader adoption of artificial intelligence-assisted tools in hepatology practice. As a result, this study can guide and inform future investigations on attention-based models in medical ultrasound analysis. Furthermore, this study provides a reference for integrating attention mechanisms into deep learning architectures. Future studies should prioritize external multi-center validation and standardized reporting to assess whether the observed improvements remain consistent across different ultrasound devices, acquisition protocols, and clinical populations.

As summarized in Table 7, the proposed method achieves the highest accuracy (98.34%) compared with the previously reported architectures evaluated on the same dataset setting in [24]. The improvement is accompanied by consistently high precision, recall, and F1-scores across stages, indicating that the performance gain is not limited to the majority classes. This suggests that the attention-based feature refinement and the proposed evaluation protocol contribute to more reliable separation of the clinically challenging intermediate stages (F1–F3), where conventional CNN backbones tend to show higher confusion. Overall, these results highlight that integrating CBAM with a robust validation scheme enhances both classification accuracy and clinical reliability for liver fibrosis staging.

V. Conclusion

The present study demonstrates the successful application of a deep learning framework for classifying liver fibrosis stages (F0–F4) from ultrasound images by enhancing the ResNet-50 architecture with the CBAM. The integration of channel and spatial attention addressed the limitations of conventional CNNs, particularly in recognizing subtle textural variations in intermediate fibrosis stages (F1–F3). Using stratified 5-fold validation and ensemble averaging of logits from five trained models, the proposed system demonstrated highly stable performance throughout training, as evidenced by consistent accuracy and loss curves without signs of overfitting. The final evaluation on the test set showed that the CBAM-enhanced ResNet-50 achieved an accuracy of 98.34%, accompanied by exhibiting consistently high precision, recall, and F1-scores across all classes of fibrosis, with confirmation from the confusion matrix further confirming excellent classification capability, especially for classes F0 and F4, which were recognized with perfect accuracy. These results indicate that CBAM substantially strengthens ResNet-50's feature extraction, improves model generalization, and enhances the reliability of automated

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

liver fibrosis staging from ultrasound images. Notwithstanding the strong results obtained, further investigations are required to confirm the robustness and generalizability of the proposed approach. Future work should prioritize external validation on additional multi-center datasets and ultrasound devices to evaluate performance stability across different acquisition settings. In addition, exploring clinically plausible augmentation strategies beyond horizontal flipping and comparing alternative attention modules under the same training and evaluation protocol may provide deeper insight into improving ultrasound-based liver fibrosis staging.

Acknowledgment

The authors would like to express their sincere appreciation to Universitas Sebelas Maret for its continuous institutional support and the academic environment that contributed to the completion of this research. The authors also thank the Faculty of Information Technology and Data Science for providing essential resources, facilities, and administrative assistance throughout the study. This work was further supported by the guidance and encouragement of faculty members and the institution's ongoing commitment to research and innovation in information technology and data science.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not for profit sectors.

Data Availability

The dataset analyzed in this study is publicly available and was obtained from the Kaggle repository reported by Joo et al. [24]. All data preprocessing procedures and evaluation protocols used in this work are described in the Methods section to support reproducibility.

Author Contribution

Bagus Tegar Zahir Afif contributed to model implementation, experimentation, evaluation, and manuscript drafting. Wiharto supervised the study, contributed to the research design, and provided critical revisions to the manuscript. Umi Salamah contributed to methodological refinement, interpretation of results, and manuscript review. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work.

Declarations

Ethical Approval

This study used a publicly available and de-identified ultrasound image dataset reported in [24]. Therefore, no additional ethical approval was required for this work.

Consent for Publication Participants.

Consent for publication was given by all participants

Competing Interests

The authors declare no competing interests.

References

[1] S. K. Asrani, H. Devarbhavi, J. Eaton, and P. S. Kamath, "Burden of liver diseases in the world Introduction and global burden." [Online]. Available: http://www.ncdrisc.org/index.html

[2] J. Li, Q. Wang, W. Ni, C. Liu, Z. Li, and X. Qi, "Global health burden of cirrhosis and other chronic liver diseases (CLDs) due to non-alcoholic fatty liver disease (NAFLD): A systematic analysis for the global burden of disease study 2019," *Glob Transit*, vol. 5, pp. 160–169, Jan. 2023, doi: 10.1016/j.glt.2023.09.002.

[3] V. Hernandez-Gea and S. L. Friedman, "Pathogenesis of liver fibrosis," *Annual Review of Pathology: Mechanisms of Disease*, vol. 6, pp. 425–456, Feb. 2011, doi: 10.1146/annurev-pathol-011110-130246.

[4] M. Parola and M. Pinzani, "Invited review liver fibrosis in NAFLD/NASH: From pathophysiology towards diagnostic and therapeutic strategies," Feb. 01, 2024, *Elsevier Ltd*. doi: 10.1016/j.mam.2023.101231.

[5] G. Ferraioli and R. G. Barr, "Ultrasound evaluation of chronic liver disease," Mar. 01, 2025, *Springer*. doi: 10.1007/s00261-024-04568-2.

[6] G. Ferraioli *et al.*, "WFUMB Guideline/Guidance on Liver Multiparametric Ultrasound: Part 1. Update to 2018 Guidelines on Liver Ultrasound Elastography," Aug. 01, 2024, *Elsevier Inc.* doi: 10.1016/j.ultrasmedbio.2024.03.013.

[7] R. Loomba and L. A. Adams, "Advances in non-invasive assessment of hepatic fibrosis," Jul. 01, 2020, *BMJ Publishing Group*. doi: 10.1136/gutjnl-2018-317593.

[8] L. Castera, M. Friedrich-Rust, and R. Loomba, "Noninvasive Assessment of Liver Disease in Patients With Nonalcoholic Fatty Liver Disease," Apr. 01, 2019, *W.B. Saunders*. doi: 10.1053/j.gastro.2018.12.036.

[9] C. Sang *et al.*, "Diagnosis of Fibrosis Using Blood Markers and Logistic Regression in Southeast Asian Patients With Non-alcoholic Fatty Liver Disease," *Front Med (Lausanne)*,

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

vol. 8, Feb. 2021, doi: 10.3389/fmed.2021.637652.

[10] M. J. Brol, U. Drebber, J. A. Luetkens, M. Odenthal, and J. Trebicka, "'The pathogenesis of hepatic fibrosis: basic facts and clinical challenges'—assessment of liver fibrosis: a narrative review," *Dig Med Res*, vol. 5, pp. 24–24, Jun. 2022, doi: 10.21037/dmr-22-9.

[11] A. Berzigotti *et al.*, "EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis – 2021 update," *J Hepatol*, vol. 75, no. 3, pp. 659–689, Sep. 2021, doi: 10.1016/j.jhep.2021.05.025.

[12] S. A. Hicks *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-09954-8.

[13] N. Leigh and C. W. Hammill, "Liver ultrasound: Normal anatomy and pathologic findings," *Surg Open Sci*, vol. 19, pp. 178–198, Jun. 2024, doi: 10.1016/j.sopen.2024.05.002.

[14] H. Shokoohi, D. Chu, and N. Al Jalbout, "Ultrasound Physics," Nov. 01, 2024, *W.B. Saunders*. doi: 10.1016/j.emc.2024.05.002.

[15] N. Kutaiba, W. Chung, M. Goodwin, A. Testro, G. Egan, and R. Lim, "The impact of hepatic and splenic volumetric assessment in imaging for chronic liver disease: a narrative review," Dec. 01, 2024, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1186/s13244-024-01727-3.

[16] T. Duan, H. Y. Jiang, B. Song, and W. W. Ling, "Noninvasive imaging of hepatic dysfunction: A state-of-the-art review," Apr. 28, 2022, *Baishideng Publishing Group Inc*. doi: 10.3748/wjg.v28.i16.1625.

[17] H. C. Park *et al.*, "Automated classification of liver fibrosis stages using ultrasound imaging," *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12880-024-01209-4.

[18] N. S. Punn, B. Patel, and I. Banerjee, "Liver fibrosis classification from ultrasound using machine learning: a systematic literature review," Jan. 01, 2024, *Springer*. doi: 10.1007/s00261-023-04081-y.

[19] S. Hirata, A. Isshiki, D. I. Tai, P. H. Tsui, K. Yoshida, and T. Yamaguchi, "Convolutional neural network classification of ultrasound images by liver fibrosis stages based on echo-envelope statistics," *Front Phys*, vol. 11, 2023, doi: 10.3389/fphy.2023.1164622.

[20] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," Dec. 2018, doi: 10.1016/j.zemedi.2018.11.002.

[21] R. Santos, J. Pedrosa, A. M. Mendonça, and A. Campilho, "Grad-CAM: The impact of large receptive fields and other caveats," *Computer Vision and Image Understanding*, vol. 258, Jul. 2025, doi: 10.1016/j.cviu.2025.104383.

[22] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical Image Analysis using Convolutional Neural Networks: A Review," Nov. 01, 2018, *Springer New York LLC*. doi: 10.1007/s10916-018-1088-1.

[23] H. Ai, Y. Huang, D. I. Tai, P. H. Tsui, and Z. Zhou, "Ultrasonic Assessment of Liver Fibrosis Using One-Dimensional Convolutional Neural Networks Based on Frequency Spectra of Radiofrequency Signals with Deep Learning Segmentation of Liver Regions in B-Mode Images: A Feasibility Study," *Sensors*, vol. 24, no. 17, Sep. 2024, doi: 10.3390/s24175513.

[24] Y. Joo, H. C. Park, O. J. Lee, C. Yoon, M. H. Choi, and C. Choi, "Classification of Liver Fibrosis From Heterogeneous Ultrasound Image," *IEEE Access*, vol. 11, pp. 9920–9930, 2023, doi: 10.1109/ACCESS.2023.3240216.

[25] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," Jul. 2018, [Online]. Available: http://arxiv.org/abs/1807.06521

[26] X. Dong, Q. Tan, S. Xu, J. Zhang, and M. Zhou, "Ultrasound image-based contrastive fusion non-invasive liver fibrosis staging algorithm," *Abdominal Radiology*, Dec. 2025, doi: 10.1007/s00261-025-04991-z.

[27] P. Gupta *et al.*, "Advanced liver fibrosis detection using a two-stage deep learning approach on standard T2-weighted MRI," *Abdominal Radiology*, 2025, doi: 10.1007/s00261-025-05160-y.

[28] H. Zhao *et al.*, "Diagnostic performance of EfficientNetV2-S method for staging liver fibrosis based on multiparametric MRI," *Heliyon*, vol. 10, no. 15, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35115.

[29] H. C. Park *et al.*, "Automated classification of liver fibrosis stages using ultrasound imaging," *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12880-024-01209-4.

[30] K. Yang, F. Chen, A. Tian, L. Deng, and X. Mao, "A Novel Deep Learning Framework for Liver Fibrosis Staging and Etiology Diagnosis Using Integrated Liver–Spleen Elastography," *Diagnostics*, vol. 15, no. 23, p. 2986, Nov. 2025, doi: 10.3390/diagnostics15232986.

[31] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021, doi: 10.1016/j.neucom.2021.03.091.

[32] W. Islam, M. Jones, R. Faiz, N. Sadeghipour, Y. Qiu, and B. Zheng, "Improving Performance of Breast Lesion Classification Using a ResNet50

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

Model Optimized with a Novel Attention Mechanism," *Tomography*, vol. 8, no. 5, pp. 2411–2425, Oct. 2022, doi: 10.3390/tomography8050200.

[33] S. Potharaju, S. N. Tambe, S. K. Tadepalli, S. Salvadi, T. C. Manjunath, and A. Srilakshmi, "Optimizing Waste Management with Squeeze-and-Excitation and Convolutional Block Attention Integration in ResNet-Based Deep Learning Frameworks," *Journal of Artificial Intelligence and Technology*, vol. 5, pp. 211–220, Jun. 2025, doi: 10.37965/jait.2025.0709.

[34] A. Danyo, A. Dontoh, and A. Aboah, "An Improved ResNet50 Model for Predicting Pavement Condition Index (PCI) Directly from Pavement Images," Apr. 2025, [Online]. Available: http://arxiv.org/abs/2504.18490

[35] A. Al-Kababji, F. Bensaali, S. P. Dakua, and Y. Himeur, "Automated liver tissues delineation techniques: A systematic survey on machine learning current trends and future orientations," Jul. 2022, [Online]. Available: http://arxiv.org/abs/2103.06384

[36] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artif Intell Rev*, vol. 57, no. 4, Apr. 2024, doi: 10.1007/s10462-024-10721-6.

[37] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, "A Comprehensive Survey of Loss Functions in Machine Learning," *Annals of Data Science*, vol. 9, no. 2, pp. 187–212, Apr. 2022, doi: 10.1007/s40745-020-00253-5.

[38] Z. Chen, H. Zhu, Y. Liu, and X. Gao, "MSCA-UNet: multi-scale channel attention-based UNet for segmentation of medical ultrasound images," *Cluster Comput*, vol. 27, no. 5, pp. 6787–6804, Aug. 2024, doi: 10.1007/s10586-024-04292-y.

[39] I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, "Deep Convolutional Neural Networks in Medical Image Analysis: A Review," Mar. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/info16030195.

[40] Y. Zhang *et al.*, "Deep-Learning Model of ResNet Combined with CBAM for Malignant–Benign Pulmonary Nodules Classification on Computed Tomography Images," *Medicina (Lithuania)*, vol. 59, no. 6, Jun. 2023, doi: 10.3390/medicina59061088.

[41] Y. Ye *et al.*, "Image segmentation using improved U-Net model and convolutional block attention module based on cardiac magnetic resonance imaging," *J Radiat Res Appl Sci*, vol. 17, no. 1, p. 100816, Mar. 2024, doi: 10.1016/j.jrras.2023.100816.

[42] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," Sep. 07, 2022, *Elsevier B.V.* doi: 10.1016/j.neucom.2022.06.111.

Author Biography

Bagus Tegar Zahir Afif is an undergraduate student at Sebelas Maret University, Surakarta, majoring in Informatics. He has strong research interests in deep learning, machine learning, and medical image analysis. He actively engages in research activities related to computer-aided diagnosis and neural network optimization. He completed the Bangkit Academy program under the Machine Learning track, which strengthened his technical competencies in model development, data preprocessing, and applied machine learning. His current research focuses on medical image classification, attention mechanisms, and clinical decision support systems. Throughout his academic journey, he has consistently sought opportunities to expand his knowledge through scientific forums, workshops, and applied research projects. He is particularly interested in developing robust and interpretable artificial intelligence models for real-world clinical and healthcare applications in Indonesia.

Wiharto is a senior lecturer in the Department of Informatics, Faculty of Information Technology and Data Science, Sebelas Maret University, Surakarta. He has a strong academic background and extensive experience in computational science and engineering. He earned his doctoral degree from Gadjah Mada University, where he specialized in biomedical informatics. His research primarily focuses on biomedical informatics, artificial intelligence, and computational intelligence, with an emphasis on developing innovative computational approaches to solve real-world problems. His academic contributions include mentoring undergraduate and postgraduate students, publishing research articles in reputable national and international journals, and collaborating on interdisciplinary projects to advance health informatics and artificial intelligence.

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406

Umi Salamah is a lecturer in the Department of Informatics at the Faculty of Information Technology and Data Science, Sebelas Maret University, Surakarta. She earned both her Master's and Doctoral degrees from Institut Teknologi Sepuluh Nopember, Surabaya. Her research interests include medical image processing, pattern recognition, and intelligent systems for healthcare applications. She has conducted research on computer-aided diagnosis, particularly in image-based detection of infectious diseases. Beyond research, she actively contributes to curriculum development, academic activities, and student supervision. She has published her work in national and international journals and continues to advance artificial intelligence for medical imaging and health informatics to support clinical and healthcare innovation.

Manuscript Received 20 October 2025; Revised 5 January 2026; Accepted 10 January 2026; Available online 18 January 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i1.1406