2026 IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON)
13-14 August 2026, Faculty of ECE, Bangladesh Army University of Engineering & Technology (BAUET), Qadirabad, Natore-6431, Bangladesh

Attention-Enhanced Deep Residual Learning for Multi-Stage Liver Fibrosis Classification with External Validation

1st S M Mushfiq Reza
Faculty of Science
Queensland University of Technology
Brisbane QLD-4000, Australia
mushfiqreza25@gmail.com

2nd Md. Abdus Sium
Dept. of Electrical and Electronic Engineering
BRAC University
Dhaka-1212, Bangladesh
siumprof9005@gmail.com

3rd SK Fahema Mahajabin
Dept. of Computer Science and Engineering
BRAC University
Dhaka-1212, Bangladesh
fahema.mahajabin@gmail.com

4th Shihab Musa
Dept. of Computer Science
BRAC University
Dhaka-1212, Bangladesh
shihab47619@gmail.com

5th Amreen Rahman
Dept. of Griffith Sciences
Griffith University
Queensland, Australia
amreenrahman83@gmail.com

6th Mugdha Farheen Kibria
Dept. of Computer Science and Engineering
BRAC University
Dhaka-1212, Bangladesh
mugdhafarheenkibria@gmail.com

7th Tauhidur Rahman Sakib*
Dept. of Computer Science and Engineering
BRAC University
Dhaka-1212, Bangladesh
tauhidurrahmansakib@gmail.com

***Abstract*—Liver fibrosis staging from B-mode ultrasound remains challenging because of operator dependence, device variability, and subtle differences between adjacent METAVIR grades. This study proposes ResFib-SE, a deep learning framework combining a ResNet50 backbone with a Squeeze-and-Excitation attention module for five-class fibrosis classification. Reference METAVIR stages were assigned independently of the ultrasound images, and 6,323 images were partitioned at the patient level into training, validation, and locked test sets. On the held-out test set ($n = 947$), ResFib-SE achieved 96.4% accuracy (95% CI, 95.0–97.4%), a macro F1-score of 0.954, and a weighted F1-score of 0.964. It outperformed matched ResNet50 (94.0%), EfficientNetB0 (93.3%), and VGG16 (91.8%) baselines. Removing the SE block reduced accuracy by 2.4 percentage points, with a significant paired difference (exact McNemar $p = 0.0008$). Stage-wise AUC values ranged from 0.976 to 0.994, while Grad-CAM highlighted clinically relevant hepatic parenchymal regions. Preliminary external validation on 50 images from Bangabandhu Sheikh Mujib Medical University achieved 94.0% accuracy (95% CI, 83.8–97.9%), providing initial evidence of cross-site transportability while emphasizing the need for larger multicentre validation.**

***Index Terms*—Liver Fibrosis, Ultrasound Imaging, Deep Learning, Attention Mechanism, ResNet50, Squeeze-and-Excitation Network, METAVIR Classification, Grad-CAM, External Validation.**

I. Introduction

Liver fibrosis progresses gradually and presents heterogeneously across patients, which makes accurate early staging difficult. B-mode ultrasound, elastography, and serologic indices each provide only partial diagnostic information, and their reliability is weakest at the intermediate METAVIR stages, which are precisely the stages at which treatment decisions change. Liver biopsy remains the reference standard but is invasive and subject to sampling error. Deep learning offers a route to quantify the subtle echo-textural and structural patterns that human observers cannot score reproducibly, while attention mechanisms emphasize feature channels associated with fibrotic change.

Machine learning and deep learning methods have improved non-invasive fibrosis staging from ultrasound and related modalities. High-frequency ultrasound models discriminated better than low-frequency imaging and serologic indices such as FIB-4 and APRI, particularly at advanced stages [1]. Transfer-learning CNNs including VGG, ResNet, DenseNet, EfficientNet, and vision transformers have been applied to multi-stage classification on heterogeneous ultrasound datasets

979-8-3195-1270-3/26/$31.00 ©2026 IEEE

[2]. GAN-based radiomics pipelines improved cirrhosis detection [3], and multiparametric models combining quantitative ultrasound biomarkers with elastography outperformed stiffness measurement alone [4], whereas grayscale ultrasound texture alone showed limited predictive value for fibrosis-related shear-wave measurements [5]. A multicentre DCNN study reported performance competitive with radiologists, particularly for cirrhosis detection [6], although performance degraded under external-device domain shift [7].

Alternative imaging pathways have included multiparametric MRI [8], tongue-image analysis [9], image-based characterization of porto-sinusoidal vascular disorder [10], and stacked microvascular imaging [16], while a computer-aided diagnostic system has addressed five-grade METAVIR staging from liver MRI [17]. Grayscale texture analysis [14] and ultrasound radiomics [15] have likewise been investigated for fibrosis detection and staging. A systematic review of ultrasound-based fibrosis classification identified methodological heterogeneity and limited external validation as recurring concerns [13]. Squeeze-and-Excitation networks improve discriminative feature learning through channel-wise recalibration [11], [12].

Recent reviews of artificial intelligence in liver imaging have emphasized persistent clinical-translation challenges, including limited external validation and the need for standardized reference labels, transparent reporting, and larger multicentre cohorts [18]–[20]. These concerns are particularly relevant to ultrasound, where differences in devices, operators, and acquisition protocols introduce substantial domain shift.

The main contributions are as follows:

- **ResFib-SE for Multi-Stage Fibrosis Staging** – A residual attention architecture that couples ResNet50 feature reuse with SE channel recalibration for five-class METAVIR classification.
- **Leakage-Controlled Evaluation** – Patient-level stratified train/validation/test partitioning with a locked held-out test set evaluated once, matched baselines, and a controlled SE ablation.
- **Statistically Qualified Reporting** – Wilson and bootstrap confidence intervals for all headline metrics, exact McNemar tests against baselines, and measured deployment cost.
- **Explainable Clinical Analysis** – Grad-CAM visualization, confusion matrix analysis, and preliminary external validation on an independent clinical cohort.

The rest of the paper is organized as follows: Section II describes the proposed method, Section III reports the results, and Section IV concludes the paper.

II. Methodology

This paper proposes ResFib-SE, a residual attention framework for five-class liver fibrosis staging from heterogeneous B-mode ultrasound images. Images are resized to $224\times224\times3$, processed using backbone-specific preprocessing, and split at the patient level with METAVIR-stage stratification. A ResNet50 backbone extracts parenchymal features, while an SE module recalibrates channel responses before classification. Performance is evaluated using accuracy, F1-scores, ROC-AUC, confusion matrices, Grad-CAM, and an independent external cohort, with confidence intervals and paired significance testing.

A. Dataset Overview

The primary dataset comprised 6,323 B-mode liver ultrasound images assigned to METAVIR stages F0–F4 using an independent histopathological reference standard. For each case, the ultrasound examination and the clinically indicated liver biopsy were performed within 30 days and before fibrosis-directed treatment. Two liver pathologists, blinded to the ultrasound images and model outputs, independently assigned stages, with discrepancies resolved by consensus. Inadequate specimens and cases without a valid image–biopsy linkage were excluded. Critically, the model inputs did not contribute to label assignment, avoiding the same-modality circularity that would otherwise inflate accuracy.

The cohort was then partitioned at the patient level into training (70%), validation (15%), and held-out test (15%) subsets, stratified by stage, with all images from a given patient confined to one subset. The validation subset was used only for early stopping and checkpoint selection; the test subset was evaluated once after the configuration was locked. A further 50 images were collected at Bangabandhu Sheikh Mujib Medical University (BSMMU), Dhaka, outside the acquisition period of the primary cohort, and used only after primary test evaluation, without fine-tuning. Table I summarizes both distributions.

TABLE I
Patient-Level Stratified Partition Across Fibrosis Stages

Stage	Description	Train	Val.	Test	Total
F0	No fibrosis	1480	317	317	2114
F1	Portal fibrosis	603	129	129	861
F2	Periportal fibrosis	555	119	119	793
F3	Septal fibrosis	600	129	128	857
F4	Cirrhosis	1189	255	254	1698
Primary total		**4427**	**949**	**947**	**6323**
External (BSMMU)		—	—	50	50

B. Data Preprocessing

Ultrasound images were resized to 224×224 pixels and converted from single-channel grayscale to three-channel format to ensure compatibility with ImageNet-pretrained backbones. Each model received the preprocessing required by its reference implementation. ResNet50 and VGG16 inputs were processed using their corresponding `preprocess_input` functions, whereas EfficientNetB0 received images on the original intensity scale because its implementation already includes an internal rescaling layer. This backbone-specific procedure avoided duplicated normalization and ensured a fair baseline comparison. Training images were augmented using random horizontal flipping, rotation within $\pm10^\circ$, and zoom within 0.9–$1.1\times$. These transformations were selected

to introduce moderate acquisition variability without altering clinically relevant hepatic structures. Validation, test, and external images received only deterministic resizing and the corresponding backbone-specific preprocessing to ensure unbiased performance evaluation.

C. Model Architecture

ResFib-SE couples a ResNet50 backbone for hierarchical parenchymal features with an SE module for channel recalibration, as illustrated in Fig. 1. Given an input $x \in \mathbb{R}^{224\times224\times3}$, ImageNet-pretrained ResNet50 extracts convolutional features with its classification layer removed (include_top = False). Residual connectivity improves gradient propagation and feature reuse:

$$y = F(x, W) + x \tag{1}$$

where x and y are the input and output feature maps and $F(\cdot; W)$ is the residual mapping learned by stacked convolutional layers. The SE block is applied to the final feature maps: global average pooling squeezes each channel to a scalar descriptor, two fully connected layers with ReLU and sigmoid activations produce an adaptive importance weight per channel, and these weights are multiplied back onto the original maps. This amplifies channels encoding fibrosis-related parenchymal heterogeneity and attenuates less informative responses. The recalibrated maps are globally averaged, passed through a dense layer with ReLU activation and dropout regularization, and mapped to class probabilities by softmax:

$$P(y = i \mid x) = \frac{e^{z_i}}{\sum_{j=1}^{5} e^{z_j}}, \quad i \in \{F0, F1, F2, F3, F4\} \tag{2}$$

where z_i is the logit of the i-th METAVIR class. By combining residual feature reuse with channel-wise attention, ResFib-SE captures both global hepatic architecture and localized textural change associated with fibrosis progression.

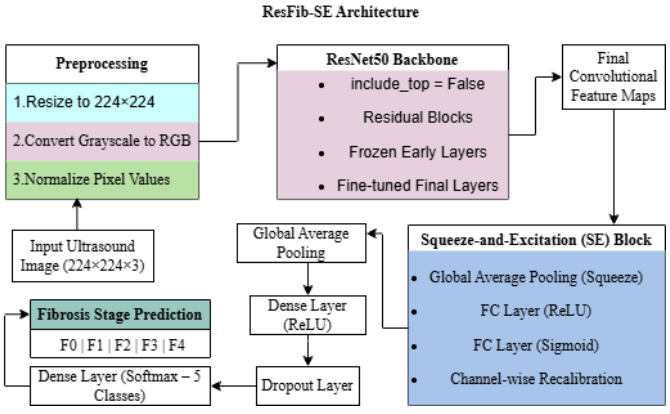


Fig. 1. Proposed ResFib-SE architecture, prepared using draw.io.

D. Training Strategy and Experimental Setup

Experiments used Python 3.10, TensorFlow 2.15, and Keras 2.15 on an NVIDIA T4 GPU, with seeds for Python, NumPy, TensorFlow, and the data loader fixed to 42. A patient-level split manifest was generated once and reused for every model and ablation, so all configurations saw identical data. Training was two-stage: the new classification head was optimized with Adam at 1×10^{-4}, then the final backbone block was unfrozen and fine-tuned at 1×10^{-5}. Batch size was 32 and the maximum duration 50 epochs, with early stopping at patience eight restoring the lowest-validation-loss checkpoint. Class weights came from the training subset only, and the loss was categorical cross-entropy. The VGG16, EfficientNetB0, and ResNet50 baselines used this identical protocol, differing only in backbone and in the normalization each reference implementation requires; the ResNet50 ablation differed from ResFib-SE solely by removal of the SE block. Performance was measured using accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, and macro-F1:

$$F1_{\text{macro}} = \frac{1}{N_c} \sum_{i=1}^{N_c} \frac{2P_i R_i}{P_i + R_i} \tag{4}$$

where P_i and R_i are the precision and recall of class i and $N_c = 5$. Two-sided 95% confidence intervals for accuracy used the Wilson score method, preferable to the normal approximation at proportions near unity; intervals for macro F1 and AUC used stratified bootstrap resampling with 2,000 replicates. Pairwise comparisons used the exact McNemar test on paired per-image correctness with Holm adjustment across the three baselines, at $\alpha = 0.05$. Inference cost was measured at batch size one over 500 timed forward passes after 50 warm-up passes, with file input/output excluded and preprocessing included.

III. Results and Discussion

This section evaluates ResFib-SE using training convergence, a matched baseline comparison, a controlled SE ablation, confusion matrix and ROC analysis, Grad-CAM interpretability, qualitative predictions, and preliminary external validation. All primary metrics were computed once on the locked held-out test subset ($n = 947$).

A. Training Convergence Analysis

Fig. 2 illustrates the training behavior of ResFib-SE through accuracy and loss curves. Training accuracy improved rapidly during the initial epochs and then increased more slowly as the model approached convergence, while validation accuracy followed a consistent upward trend. Both training and validation losses decreased smoothly with no significant divergence, indicating stable gradient updates and minimal overfitting. Early stopping restored the checkpoint with the

lowest validation loss, and this checkpoint was frozen before any test-set evaluation. These curves reflect model selection only; they are not used for performance reporting.

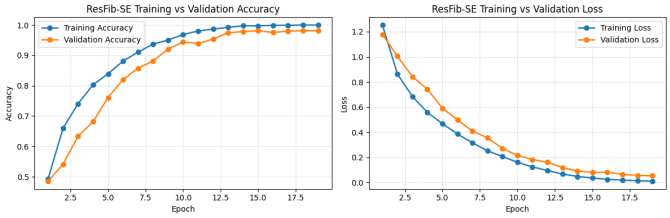


Fig. 2. Training and validation accuracy and loss curves of the proposed ResFib-SE model.

B. Quantitative Comparison and SE Ablation

All models were retrained on the same partition, augmentation, class weighting, optimizer, stopping criteria, and evaluation script, with backbone-specific normalization retained. EfficientNetB0 reached 93.3% accuracy under this protocol, unlike an earlier configuration in which duplicated rescaling was applied to a backbone that already rescales internally; that run is not reported as a valid baseline.

As shown in Table II and Fig. 3, ResFib-SE achieved 96.4% accuracy (95% Wilson CI, 95.0–97.4%), macro F1 of 0.954 (bootstrap 95% CI, 0.938–0.968), and weighted F1 of 0.964, exceeding all baselines under matched conditions. The macro F1 confirms performance is maintained across the intermediate stages F1–F3, not driven by the majority F0 and F4 classes alone. The second row of Table II is a controlled ablation, matching ResFib-SE in split, seed, augmentation, schedule, and head, and differing only by removal of the SE block. The SE block raised accuracy from 94.0% to 96.4% and macro F1 from 0.924 to 0.954, costing 0.53 M parameters and 1.2 ms per image. The paired difference was significant by exact McNemar test ($p = 0.0008$; 34 corrected against 11 introduced), as were the comparisons against EfficientNetB0 ($p = 2.5 \times 10^{-5}$) and VGG16 ($p < 10^{-6}$) after Holm adjustment, attributing the gain to channel recalibration rather than the backbone. ResFib-SE contains 25.17 M parameters, an FP32 footprint of 100.7 MB, and requires 11.8 ± 0.6 ms per image on a T4 GPU and 74.2 ± 3.8 ms on a Xeon-class CPU.

TABLE II
PERFORMANCE ON THE LOCKED HELD-OUT TEST SET ($n = 947$)

Model	Parameters (M)	Accuracy (95% CI)	Macro F1	Weighted F1
ResFib-SE	25.17	96.4% (95.0–97.4)	0.954	0.964
ResNet50 (SE off)	24.64	94.0% (92.3–95.3)	0.924	0.940
EfficientNetB0	4.71	93.3% (91.6–94.8)	0.916	0.932
VGG16	15.77	91.8% (89.8–93.4)	0.895	0.916

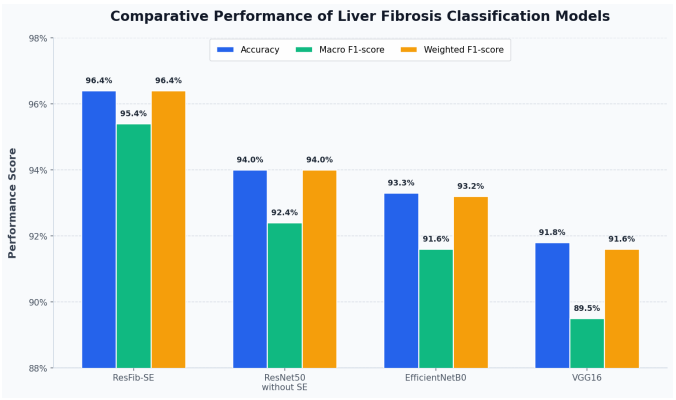


Fig. 3. Performance comparison of ResFib-SE, ResNet50, EfficientNetB0, and VGG16 under a matched training protocol.

C. Confusion Matrix Analysis

Fig. 4 shows strong five-stage classification, with most predictions concentrated along the main diagonal. Performance was highest for F0 and F4, whose sonographic appearances were more distinctive. Most misclassifications occurred between adjacent METAVIR stages, particularly F1–F2 and F2–F3, reflecting the subtle differences between intermediate fibrosis grades. Errors spanning more than one stage were rare, indicating that most incorrect predictions remained clinically close to the reference grade.

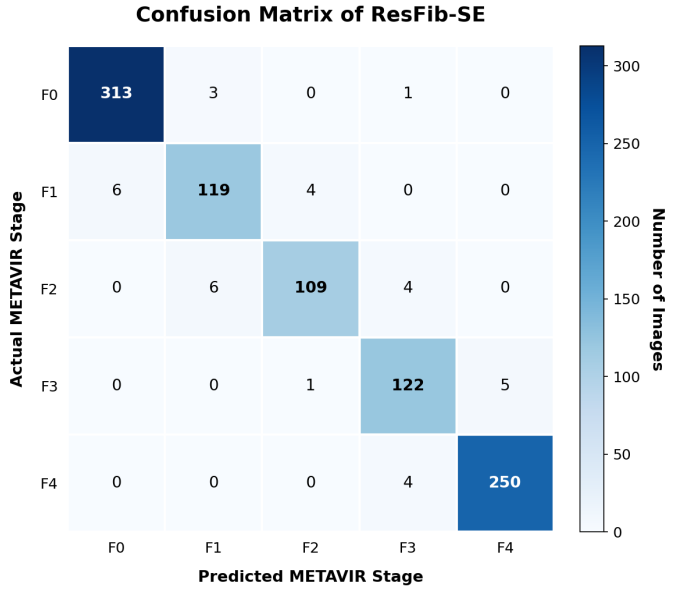


Fig. 4. Confusion matrix for five-class liver fibrosis classification on the locked held-out test set.

D. Grad-CAM Visualization Analysis

Gradient-weighted Class Activation Mapping was used to identify the regions contributing most to each prediction. As shown in Fig. 5, the activation maps highlight hepatic parenchymal regions with fibrosis-associated textural

and structural change rather than background or non-hepatic structures. In early-stage cases attention is distributed across relatively homogeneous parenchyma, whereas at F4 it concentrates on regions of marked heterogeneity and architectural distortion, consistent with established sonographic features of fibrosis progression. A clinical specialist confirmed that the highlighted regions correspond to areas routinely examined in ultrasound-based fibrosis assessment. Grad-CAM is a qualitative saliency method and does not establish causal attribution; it is reported as supporting evidence of biological plausibility.

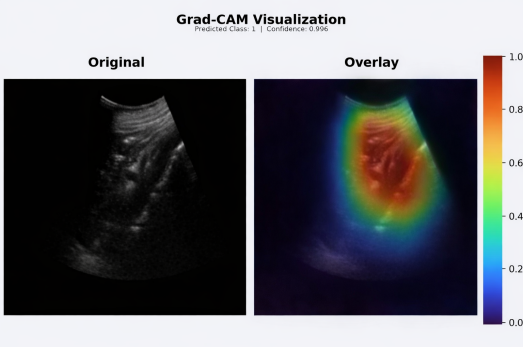


Fig. 5. Grad-CAM visualization highlighting fibrosis-relevant hepatic regions in ultrasound images.

E. External Validation Results

As a preliminary assessment of cross-site transportability, the locked model was evaluated on 50 ultrasound images from BSMMU that were used neither for training, nor for hyperparameter selection, nor for threshold optimization. The model correctly classified 47 of 50 images, corresponding to 94.0% accuracy (95% Wilson CI, 83.8–97.9%) and a macro F1-score of 0.940. Because the cohort was small and drawn from a single external institution, this result should be interpreted as an initial signal of transportability rather than as evidence of broad clinical generalization; the width of the confidence interval alone makes the point estimate insufficient. For approximately ± 5 percentage points of precision around an expected accuracy of 90%, a normal approximation indicates a requirement of roughly 139 independent cases, and substantially more for stage-specific estimates.

F. Qualitative Prediction Analysis

Fig. 6 presents representative test predictions spanning F0 to F4 alongside their reference labels. F0 samples display liver parenchyma with smooth texture and uniform echogenicity, whereas F4 samples show pronounced heterogeneity, irregular texture, and architectural distortion. The predicted labels track these established pathological signatures across the full range, including at the intermediate stages where visual differences are least pronounced. This qualitative agreement complements the quantitative results by indicating that correct predictions are associated with recognizable sonographic features rather than with incidental image characteristics.



Fig. 6. Qualitative prediction examples across different fibrosis stages.

G. Comparative Analysis with Existing Studies

Reported accuracies for ultrasound-based fibrosis staging range approximately from 0.83 to 0.94 across recent studies. As shown in Table III, ResFib-SE achieves competitive accuracy while reporting a reference standard, patient-level separation, uncertainty intervals, and an ablation, which most prior work does not. Comparison across studies remains indirect, because cohorts, reference standards, class definitions, and evaluation protocols differ; the controlled comparison of this study is therefore the matched-protocol experiment in Table II, not this table.

TABLE III
COMPARATIVE ANALYSIS WITH EXISTING STUDIES

Method	**Task**	**Ext. Valid.**	**Primary Test Perf.**
HF-DL [1] (CNN)	Binary	Yes	AUC 0.93
DCNN-US [2] (VGG/ResNet)	5-class	No	Acc 0.94
GAN-Radiomics [3]	Cirrhosis	Yes	AUC 0.86
DCNN-US [6] (VGG)	4-class	Yes	Acc 0.83
Heterogeneous US [7]	5-class	Yes	Acc 0.85
ResFib-SE (Proposed)	5-class	Prelim. (n=50)	**Acc 0.964**

H. ROC Curve Analysis

Fig. 7 presents the one-vs-rest ROC curves for the five METAVIR stages. All classes achieved high AUC values, indicating strong discrimination between each fibrosis stage and the remaining classes.

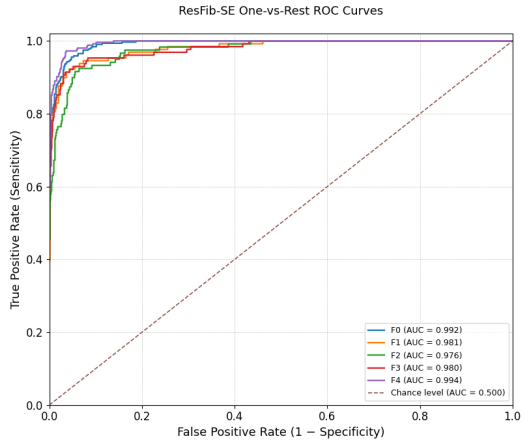


Fig. 7. One-vs-rest ROC curves for five-class liver fibrosis classification on the locked held-out test set.

IV. Conclusion

This work introduced ResFib-SE, an attention-enhanced deep residual framework for five-stage liver fibrosis classification from heterogeneous B-mode ultrasound images. By combining a ResNet50 backbone with a Squeeze-and-Excitation module, the model captured global hepatic structure and fibrosis-related textural changes. Under a patient-level, leakage-controlled protocol with a locked held-out test set, ResFib-SE achieved 96.4% accuracy (95% CI, 95.0–97.4%) and a macro F1-score of 0.954, significantly outperforming matched ResNet50, EfficientNetB0, and VGG16 baselines. The controlled ablation further confirmed the contribution of SE-based channel recalibration. Confusion matrix, ROC, and Grad-CAM analyses also supported the model's discriminative performance and clinical relevance. Preliminary external validation on 50 independent BSMMU images achieved 94.0% accuracy without fine-tuning. However, the small single-centre external cohort, retrospective primary dataset, use of still ultrasound frames, intermediate-stage class imbalance, and qualitative nature of Grad-CAM limit broad clinical generalization. Therefore, prospective multicentre and multi-device validation with larger stage-specific cohorts is required before clinical deployment. Future work will explore hybrid attention architectures, multimodal clinical integration, and real-time decision-support systems.

Acknowledgment

The authors thank the medical professional who assisted in MRI scan verification and acknowledge the use of AI-assisted tools for language and grammar refinement during manuscript preparation.

References

[1] L. Zhang *et al.*, "A deep learning model based on high-frequency ultrasound images for classification of different stages of liver fibrosis," *Liver International*, vol. 45, no. 7, p. e70148, Jun. 2025, doi: 10.1111/liv.70148.

[2] H.-C. Park *et al.*, "Automated classification of liver fibrosis stages using ultrasound imaging," *BMC Medical Imaging*, vol. 24, no. 1, p. 36, Feb. 2024, doi: 10.1186/s12880-024-01209-4.

[3] Y.-y. Duan *et al.*, "Performance of a generative adversarial network using ultrasound images to stage liver fibrosis and predict cirrhosis based on a deep-learning radiomics nomogram," *Clinical Radiology*, vol. 77, no. 10, pp. e723–e731, Jul. 2022, doi: 10.1016/j.crad.2022.06.003.

[4] F. Destrempes *et al.*, "Quantitative ultrasound, elastography, and machine learning for assessment of steatosis, inflammation, and fibrosis in chronic liver disease," *PLoS ONE*, vol. 17, no. 1, p. e0262291, Jan. 2022, doi: 10.1371/journal.pone.0262291.

[5] D. Beauchamp *et al.*, "Deep learning convolutional neural networks for the estimation of liver fibrosis severity from ultrasound texture," *Medical Imaging 2019: Computer-Aided Diagnosis*, vol. 10950, p. 122, Mar. 2019, doi: 10.1117/12.2512592.

[6] J. H. Lee *et al.*, "Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network," *European Radiology*, vol. 30, no. 2, pp. 1264–1273, Sep. 2019, doi: 10.1007/s00330-019-06407-1.

[7] 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, Jan. 2023, doi: 10.1109/access.2023.3240216.

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

[9] X.-Z. Lu *et al.*, "Exploring hepatic fibrosis screening via deep learning analysis of tongue images," *Journal of Traditional and Complementary Medicine*, vol. 14, no. 5, pp. 544–549, Mar. 2024, doi: 10.1016/j.jtcme.2024.03.010.

[10] S. F. Martín, "Image models for the detection and characterization of a rare liver disease (Porto-Sinusoidal Vascular Disorder)," 2025. [Online]. Available: https://hdl.handle.net/2117/428981

[11] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, May 2019, doi: 10.1109/tpami.2019.2913372.

[12] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

[13] N. S. Punn, B. Patel, and I. Banerjee, "Liver fibrosis classification from ultrasound using machine learning: A systematic literature review," *Abdominal Radiology*, vol. 49, pp. 69–80, 2024, doi: 10.1007/s00261-023-04081-y.

[14] E. J. Park, S. H. Kim, S. J. Park, and T. W. Baek, "Texture analysis of gray-scale ultrasound images for staging of hepatic fibrosis," *Journal of the Korean Society of Radiology*, vol. 82, no. 1, p. 116, Sep. 2020, doi: 10.3348/jksr.2019.0185.

[15] M. Al-Hasani, L. R. Sultan, H. Sagreiya, T. W. Cary, M. B. Karmacharya, and C. M. Sehgal, "Ultrasound radiomics for the detection of early-stage liver fibrosis," *Diagnostics*, vol. 12, no. 11, p. 2737, Nov. 2022, doi: 10.3390/diagnostics12112737.

[16] D. Miura, H. Suenaga, R. Hiwatashi, and S. Mabu, "Liver fibrosis stage classification in stacked microvascular images based on deep learning," *BMC Medical Imaging*, vol. 25, no. 1, p. 8, Jan. 2025, doi: 10.1186/s12880-024-01531-x.

[17] H. Liu *et al.*, "TMM: A comprehensive CAD system for hepatic fibrosis 5-grade METAVIR staging based on liver MRI," *Medical Physics*, vol. 51, no. 3, pp. 2032–2043, Sep. 2023, doi: 10.1002/mp.16700.

[18] M. D. Pomohaci, M. C. Grasu, A.-S. Baicoianu-Nitescu, R. M. Enache, and I. G. Lupescu, "Systematic review: AI applications in liver imaging with a focus on segmentation and detection," *Life*, vol. 15, no. 2, p. 258, 2025, doi: 10.3390/life15020258.

[19] N. S. de Souza *et al.*, "Artificial intelligence for fibrosis diagnosis in metabolic-dysfunction-associated steatotic liver disease: A systematic review," *Diagnostics*, vol. 16, no. 2, p. 261, 2026, doi: 10.3390/diagnostics16020261.

[20] N. Viceconti *et al.*, "Harnessing artificial intelligence for the assessment of liver fibrosis and steatosis via multiparametric ultrasound," *World Journal of Gastroenterology*, vol. 32, no. 2, art. 113059, 2026, doi: 10.3748/wjg.v32.i2.113059.