

Article

LivSCP: Improving Liver Fibrosis Classification Through Supervised Contrastive Pretraining

Yogita Dubey ^{1,*}, Aditya Bhongade ^{1,†} and Punit Fulzele ^{2,†}

¹ Department of Electronics & Telecommunication Engineering, Yeshwantrao Chavan College of Engineering, Nagpur 441110, India; adityabhongade24@gmail.com

² Directorate of Research & Innovation, Sharad Pawar Dental College, Datta Meghe Institute of Higher Education & Research, Wardha 442001, India; punitr007@gmail.com

* Correspondence: yogeetadubey@yahoo.co.in

† These authors contributed equally to this work.

Abstract

Background: Deep learning models have been used in the past for non-invasive liver fibrosis classification based on liver ultrasound scans. After numerous improvements in the network architectures, optimizers, and development of hybrid methods, the performance of these models has barely improved. This creates a need for a sophisticated method that helps improve this slow-improving performance. **Methods:** We propose *LivSCP*, a method to train liver fibrosis classification models for better accuracy than the traditional supervised learning (SL). Our method needs no changes in the network architecture, optimizer, etc. **Results:** The proposed method achieves state-of-the-art performance, with an accuracy, precision, recall, and F1-score of 98.10% each, and an AUROC of 0.9972. A major advantage of *LivSCP* is that it does not require any modification to the network architecture. Our method is particularly well-suited for scenarios with limited labeled data and computational resources. **Conclusions:** In this work, we successfully propose a training method for liver fibrosis classification models in low-data and computation settings. By comparing the proposed method with our baseline (Vision Transformer with SL) and multiple models, we demonstrate the state-of-the-art performance of our method.

Keywords: contrastive learning; liver fibrosis; pretraining; supervised contrastive learning; vision transformer



Academic Editors: Ghazanfar Latif and Jaafar Alghazo

Received: 6 November 2025

Revised: 10 December 2025

Accepted: 12 December 2025

Published: 17 December 2025

Citation: Dubey, Y.; Bhongade, A.; Fulzele, P. LivSCP: Improving Liver Fibrosis Classification Through Supervised Contrastive Pretraining. *Diagnostics* **2025**, *15*, 3226. <https://doi.org/10.3390/diagnostics15243226>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Liver diseases, particularly non-alcoholic fatty liver disease (NAFLD) and fibrosis, are a global health concern. Liver fibrosis is characterized by the excessive accumulation of extracellular matrix proteins, causing chronic liver diseases.

Globally, liver disease contributes for approximately 2 million deaths per year worldwide, with 1 million due to liver cirrhosis and 1 million due to viral hepatitis [1]. The global prevalence of NAFLD among adults is estimated to be around 32%, and higher (40%) in males compared to females (26%) [2]. The global prevalence of alcohol-related liver diseases (ARLDs) was 4.8%, as reported in [3]. From the analysis a 70 million population from 38 countries, the global prevalence of NAFLD was estimated to be 30.2% [4]. In India, the prevalence of NAFLD is estimated at 38%, with rates reaching more than 60% in urban areas [5], with 1.59% higher risk of developing the disease in males than females, as per the study reported in [6].

The disease burden also extends beyond health outcomes, with patients and families experiencing financial stress due to chronic liver disease management [7]. The analysis carried out for the period from 1990 to 2021 shows that NAFLD-related cirrhosis is rising, with about 1.43 million deaths due to cirrhosis and chronic liver diseases in 2021 [8]. These findings emphasize the need for automated tools for liver disease detection and classification.

Liver biopsy is the gold standard for fibrosis assessment; however, it is invasive, costly, and prone to sampling errors, patient discomfort, and inter-observer variability [9,10]. This has motivated the development of non-invasive methods such as elastography (FibroScan) and serological biomarkers [11,12]. However, these methods have the issue of limited reproducibility and performance across diverse populations. Ultrasound imaging is safe, low-cost, and widely accessible, making it an ideal modality for developing automated diagnostic systems for liver evaluation. However, its diagnostic sensitivity for early-stage fibrosis remains modest and is affected by speckle noise, low contrast, and inter-device variability.

Recently, deep learning (DL) has been widely used for liver disease classification [13–15]. Convolutional neural networks (CNNs) have shown strong potential in learning discriminative features directly from raw medical images. Kagadis et al. [16] used a DL network with fine-tuned elastography image sequences for liver disease assessment. A DL-based framework for liver cancer histopathology image classification is presented in [17] using CNNs trained with global labels only. A multi-scale CNN for liver disease classification in ultrasound images proposed by [18] achieved an accuracy above 90% and an AUROC of 97.8%. The effect of dataset size on histopathology-based hepatocellular carcinoma (HCC) classification is investigated by [19]. The DL classifier achieved over 90% accuracy, sensitivity, and specificity in HCC classification.

Subsequent studies have increasingly integrated advanced AI methodologies with ultrasound imaging. Ultrasound radiomics was used to detect early-stage fibrosis, demonstrating the utility of handcrafted imaging biomarkers with improved accuracy [20]. A generative adversarial network (GAN) was investigated by [21] for ultrasound-based fibrosis classification, combining synthetic data generation with a radiomics nomogram. An ensemble of deep learning networks was used on heterogeneous ultrasound images acquired across imaging devices to classify liver fibrosis [22]. A modified Faster R-CNN framework is proposed by [23] for ultrasound-based diagnosis of liver diseases, which achieved strong diagnostic performance but required high computational resources.

Ai et al. [24] used frequency-domain features by applying a one-dimensional CNN model to ultrasound radiofrequency (RF) spectra. Park et al. [25] evaluated CNNs for the automated classification of fibrosis in B-mode ultrasound, reporting accuracies above 90% and reasonable AUROC. These advancements demonstrate the increasing maturity of CNN-based solutions for fibrosis classification. More recently, hybrid frameworks have been widely used. Le et al. [26] explored radiomics as a paradigm shift in liver disease detection and classification. Xia et al. [27] proposed a method for diagnosing fibrotic non-alcoholic steatohepatitis (NASH) using an ultrasound radiomics-based logistic regression model.

Despite significant progress, several challenges remain. Histopathology-based approaches achieve strong accuracy but are invasive and impractical for repeated monitoring. CNN-based ultrasound models show excellent performance but are sometimes limited by generalizability across institutions and imaging devices. Radiomics approaches enhance interpretability but often rely on handcrafted features, which may limit scalability. GAN-based augmentation helps mitigate data scarcity but raises questions about clinical reliability.

In this study, we are interested in a situation where labelled data is available in relatively small quantities, and we are computation-constrained as well. To combat this, we propose a framework, *LivSCP*, that is useful in such data- and compute-constrained cases. Also, our method aims to boost the slow-improving performance of the liver fibrosis classification models with time.

Supervised learning (SL), unsupervised learning (UL), reinforcement learning (RL), and self-supervised learning (SSL) are common paradigms to learn model(s) from the available data. The most common, regular SL in data-constrained conditions might cause overfitting or data ill-separability. For the situations where, although in lower quantities, labelled data is available, using UL will not make the best use of the available labels. RL has many limitations; the most significant one is computation. RL requires multiple models and good computing power to post-train a model; however, it is not feasible to use it in the training phase. SSL is an interesting paradigm with which we can pre-train models on a corpus of data without labels. For vision models, methods like DINO, MAE, etc., are common choices. However, SSL is also not suitable for our case, where we are limited both by data and computational capability. Also, learning by reconstruction, like in MAE, does not lead to very well-transferable weights [28].

2. Materials and Methods

We use the Vision Transformer-based classification model illustrated in Figure 1 for this study. The methods, data, and architectures are detailed well in this section.

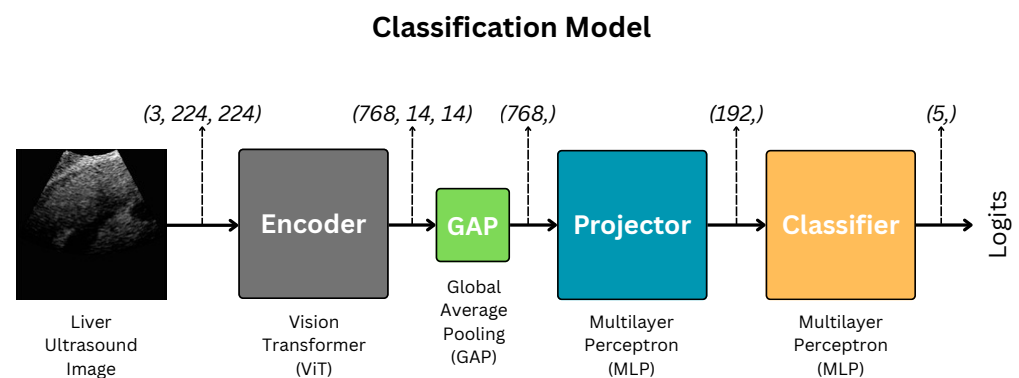


Figure 1. Architecture of the liver fibrosis classification model. GAP indicates Global Average Pooling.

2.1. Data

The METAVIR score is used for fibrosis classification. The stages are classified as F0, F1, F2, F3, and F4. F0 is the normal tissue with no fibrosis and no architectural distortion. F1 is portal fibrosis without septa. It is symptomatic and may be reversible. F2 is termed periportal fibrosis. Here, fibrosis extends beyond portal tracts, forming thin septa. This indicates significant fibrosis associated with a mild increase in liver stiffness. F3 is termed septal fibrosis with fibrotic septa linking portal tracts and central veins. In this stage, there is substantial distortion of hepatic architecture; however, cirrhosis not fully developed. F4 is the last stage of cirrhosis, with extensive fibrosis with regenerative nodules and complete architectural distortion. It is associated with complications like portal hypertension, liver failure, and hepatocellular carcinoma. The sample liver images across fibrosis stages (F0–F4) are shown in Figure 2.

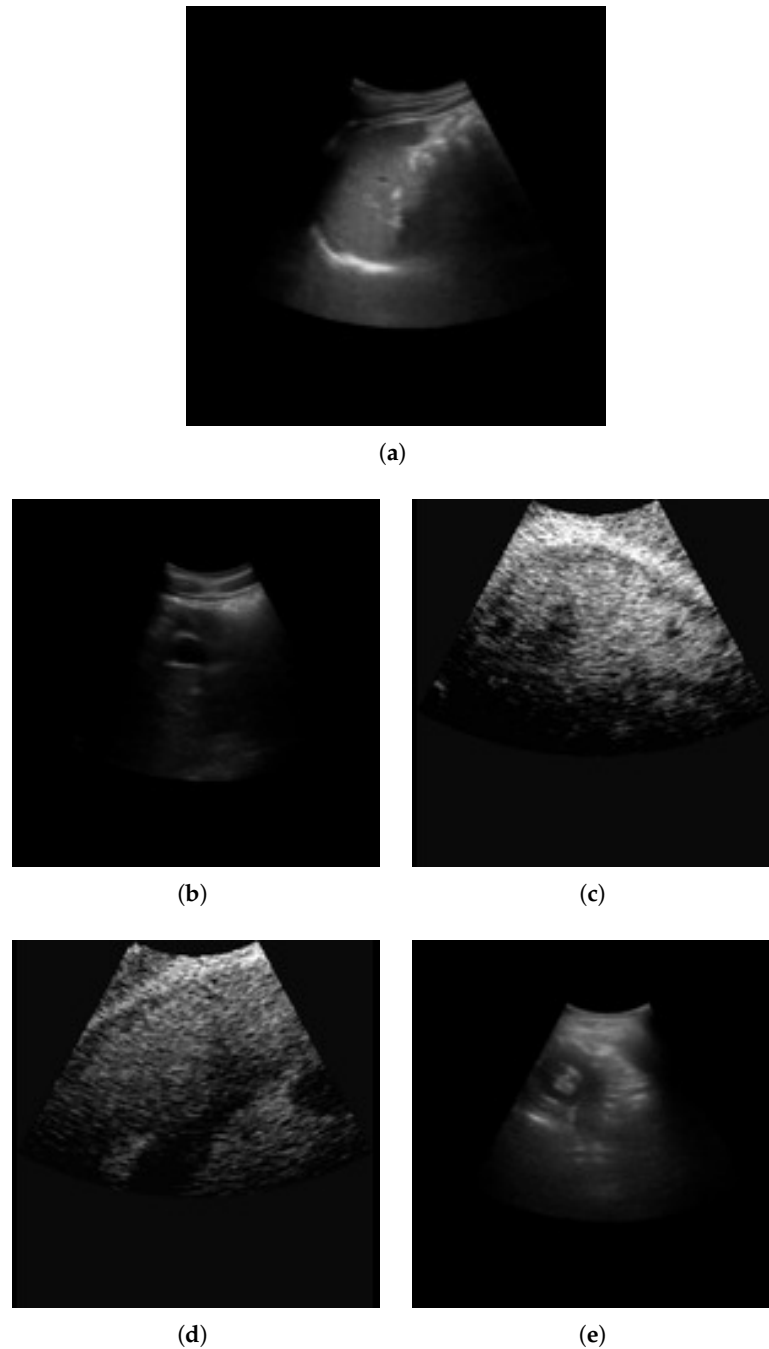


Figure 2. Sample liver images across fibrosis stages (F0–F4). (a) F0; (b) F1; (c) F2; (d) F3; (e) F4.

The liver images for fibrosis classification are sourced from Kaggle [29]. The dataset contains a total of 6323 images. The distribution of images among the stages is detailed in Table 1.

Table 1. Liver fibrosis stages and corresponding number of images.

Stage	# Images
F0	2114
F1	861
F2	793
F3	857
F4	1698

The images are resized to a uniform size of (224, 224) through bicubic interpolation [30] and normalized via Z standardization. After this, we split the data into three sets—training, validation, and testing—at a ratio of 8:1:1. Data in the same quantities and proportions were used during training and testing. We use no additional class-balancing methods since our method handles this inherently.

Unlike other methods, we use much simpler augmentations, specifically 90° rotation, horizontal and vertical flipping, color jitter, blurring, and noising. Classic SupCon [31] uses RandAugment [32] to create augmentations of the training images.

2.2. Encoder: Vision Transformer

We use Vision Transformer (ViT) [33] as the feature backbone to build the classification models. ViT is based on the transformer architecture [34], which was originally proposed for language applications. Transformers use multihead attention to learn long-range dependencies in the data. They are excellent at learning a robust global representation of the data and better at generalizability than Convolutional Neural Networks (CNNs). In ViT, the input image is first divided into non-overlapping patches; this process is called patch embedding. Considering input $X \in \mathbb{R}^{H \times W \times C}$, the patched image $X_p \in \mathbb{R}^{N \times (P^2 \cdot C)}$ is obtained as follows:

$$X_p = \text{PatchEmbedding}(X) \quad (1)$$

where PatchEmbedding is a convolutional layer (in practice) and N is the number of patches given by the following:

$$N = \frac{H \cdot W}{P^2} \quad (2)$$

where P is the size of a square patch.

The model card from [35] is used to carry out all our experiments. We use the ViT encoder only, discarding the rest, and build smaller projection and classifier heads. The pre-trained encoder is completely frozen after SCP, which is different in the case of SupCon [31], as it keeps the encoder trainable throughout.

We prefer ViT over CNN-based models for the encoder due to its superior continual learning performance [36–38]. This enables better knowledge retention from SCP. As we aim to improve the fibrosis classification performance through SCP, it is crucial to choose an encoder that demonstrates better continual learning performance, which the ViT does. Hence, it is preferred for this study.

2.3. Projection and Classification Heads

The encoding process can be formulated as follows:

$$f = \text{Encoder}(x) \quad (3)$$

where $f \in \mathbb{R}^{768 \times 14 \times 14}$ signifies the encoded feature map, Encoder signifies the ViT encoder, and $x \in \mathbb{R}^{3 \times 224 \times 224}$ is the input image.

Then, f after Global Average Pooling, GAP, is fed to the projection head, Projector, to obtain projected features m . Mathematically,

$$m = \text{Projector}(\text{GAP}(f)) \quad (4)$$

as per our implementation, $\text{GAP}(f) \in \mathbb{R}^{768}$, and $m \in \mathbb{R}^{192}$.

Finally, m is fed to the classifier to obtain the logit z .

$$z = \text{Classifier}(m) \quad (5)$$

where $z \in \mathbb{R}^C$. C is the number of classes; thus, $z \in \mathbb{R}^5$ in our case.

Projector and Classifier are shallow multilayer perceptrons (MLPs). Their architectures are illustrated in Figure 3.

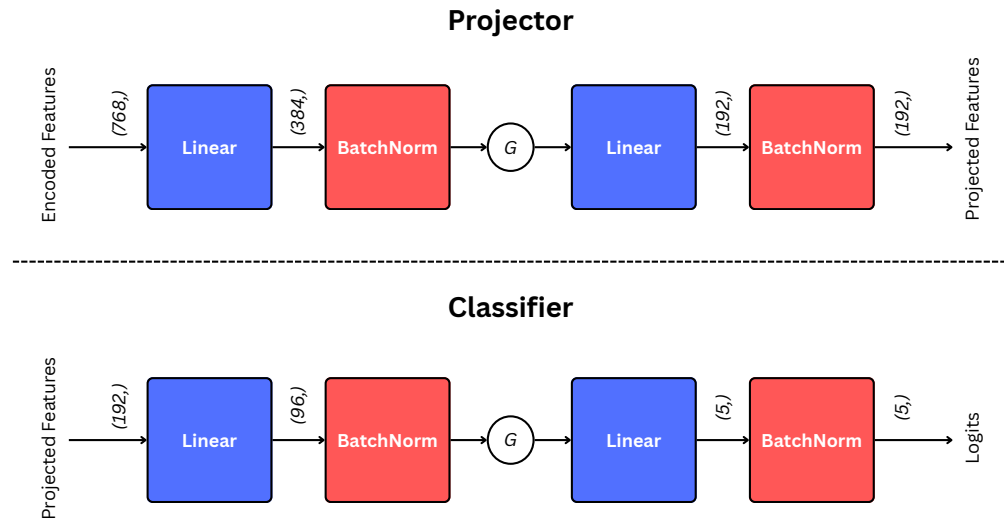


Figure 3. Architectures of projection and classification Heads. The G indicates GELU activation [39]. The encoded features (input to the projector) are first pooled through GAP, as illustrated in Equation (4).

We freeze the encoders while training the projection and classification heads.

2.4. Supervised Contrastive Learning

SCL [31] aims to maximize the inter-class distance and minimize the intra-class distance in the embedding space. This is achieved by using a Supervised Contrastive Loss function for a single sample i given by the following:

$$\mathcal{L}_i^{\text{sup}} = \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_a / \tau)} \quad (6)$$

where $P(i)$ is the set of indices of all possible positive samples, $A(i)$ is the set of all possible contrastive samples, τ is temperature, z_i is the feature embedding vector of the anchor sample, z_p is the feature embedding vector of a positive sample, and z_a is the feature embedding vector of any sample $a \in A(i)$. And, it is averaged over all samples in the batch as follows:

$$\mathcal{L}^{\text{sup}} = \frac{1}{N} \sum_{i=1}^N \mathcal{L}_i^{\text{sup}} \quad (7)$$

The encoder is pretrained with SCL in LivSCP. To combat the problem of class imbalance, as mentioned in Table 1, we use Class-Weighted Supervised Contrastive Loss (CW-SCL). CW-SCL extends the standard SCL by incorporating class-specific weights w_{y_i} to handle class imbalance. This is defined as follows:

$$\mathcal{L}^{\text{cw-sup}} = \frac{1}{N} \sum_{i=1}^N \frac{w_{y_i}}{\bar{w}} \mathcal{L}_i^{\text{sup}} \quad (8)$$

where $\bar{w} = \frac{1}{N} \sum_{i=1}^N w_{y_i}$ is the mean class weight, and $\mathcal{L}_i^{\text{sup}}$ is the supervised contrastive loss for sample i given by the following:

$$\mathcal{L}_i^{\text{sup}} = \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_a / \tau)} \quad (9)$$

Substituting, the complete expression for the CW-SCL becomes the following:

$$\mathcal{L}^{\text{CW-sup}} = -\frac{1}{N} \sum_{i=1}^N \frac{w_{y_i}}{\bar{w} |P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in A(i)} \exp(z_i \cdot z_a / \tau)} \quad (10)$$

The class weights are computed inversely proportional to the class frequencies to handle imbalance, defined as follows:

$$w_c = \frac{1/n_c}{1/C \sum_{k=1}^C 1/n_k} \quad (11)$$

where n_c is the number of samples in class c , and C is the total number of classes. This normalization ensures that the mean of all class weights is 1.

LivSCP includes pretraining the encoder with CW-SCL and then training the projection and classification heads on the embeddings produced by the frozen pretrained encoder with SL.

3. Results

3.1. Quantitative Evaluation

Table 2 depicts the comparative performance of the proposed method (*LivSCP*) with SL at different quantization levels across various metrics. As observed from Table 2, there is an absolute improvement of 5.38% in accuracy and recall, 3.88% in precision, 5.62% in F1-score, and 1.6% in mAUROC with the proposed method *LivSCP* as compared to SL in FP32 precision. The ROC Curves for models trained through *LivSCP* and SL for FP32, FP16, and INT8 precision are shown in Figure 4 and Figure 5, respectively.

From here, figures in red and green indicate the lowest and highest values of the evaluation metrics, respectively.

Table 2. Comparative performance of the proposed method (*LivSCP*) with the baseline (SL) at different quantization levels for the ViT-based network illustrated in Figure 1.

Method	Accuracy	Precision	Recall	F1 Score	mAUROC
SL-FP32	0.9272	0.9422	0.9272	0.9248	0.9812
<i>LivSCP</i> -FP32 [†]	0.9810	0.9810	0.9810	0.9810	0.9972
SL-FP16	0.9272	0.9422	0.9272	0.9298	0.9811
<i>LivSCP</i> -FP16	0.9810	0.9810	0.9810	0.9810	0.9972
SL-INT8	0.9287	0.9432	0.9287	0.9312	0.9807
<i>LivSCP</i> -INT8	0.9810	0.9810	0.9810	0.9810	0.9964

[†] Proposed. See Appendix A for the details on quantization.

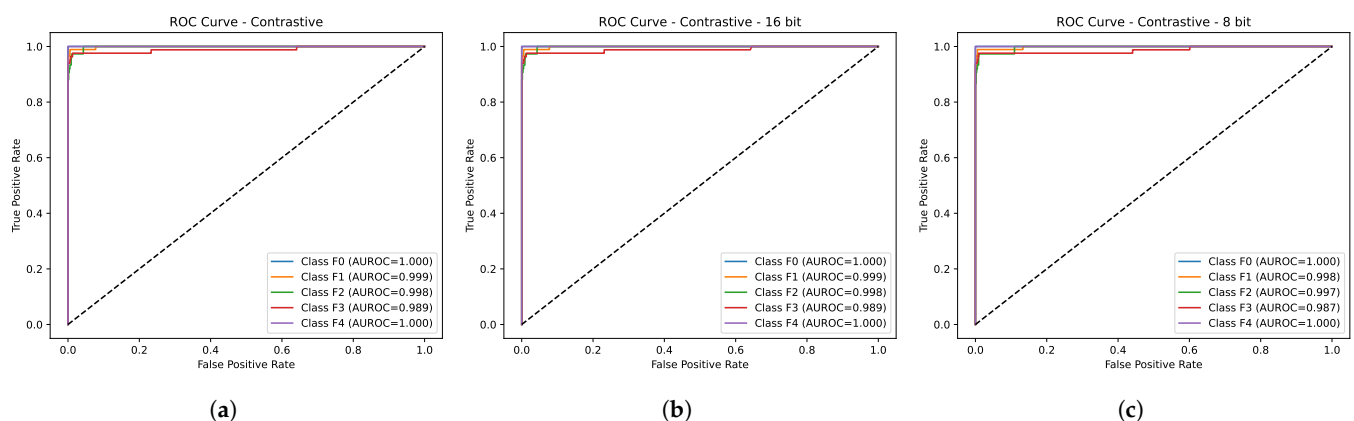


Figure 4. ROC curves for models trained through *LivSCP*. (a) FP32. (b) FP16. (c) INT8.

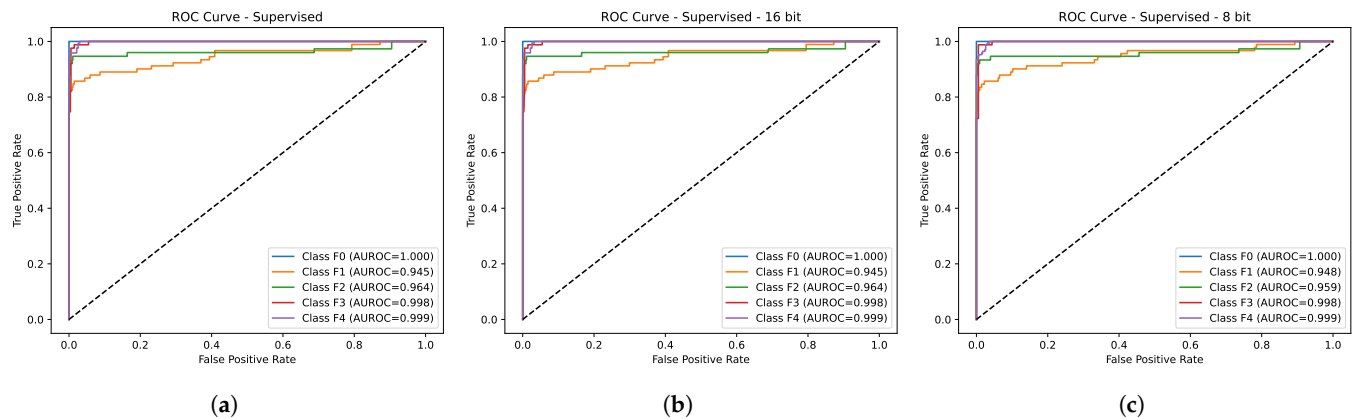


Figure 5. ROC curves for models trained through SL. (a) FP32. (b) FP16. (c) INT8.

3.2. Implementation Details

We use PyTorch (version 2.6.0) to implement, train, and test the models. NVIDIA P100 GPUs are used for accelerated training and inference. Since older GPUs do not support INT8 inference, CPUs are used for inference in Table 2. All convolutional layers are initialized using Kaiming Initialization [40], while the linear layers are initialized using Xavier Initialization [41]. The scale and shift parameters in the BatchNorm layers are initialized with ones and zeros, respectively. AdamW optimizer [42] is used while training all the models. The classes MulticlassAccuracy, MulticlassPrecision, MulticlassRecall, MulticlassF1Score, MulticlassAUROC, and MulticlassROC from torchmetrics are used to compute the performance metrics. Our source code, trained models, and results are publicly available at <https://github.com/adityabhongade/LivSCP/> (accessed on 6 December 2025).

For pretraining the encoder in LivSCP, we use temperature $\tau = 0.07$, learning rate $\eta = 3 \times 10^{-4}$, and L2 weight decay $\lambda = 10^{-4}$. The encoder is (pre)trained for a maximum of 50 epochs with a custom implementation for early stopping. These values were obtained via a small grid-search method due to computational constraints. GridSearchCV class was used, as provided in the model_selection module from the sklearn library.

4. Discussion

4.1. Contribution

This work primarily aims to improve the slow-growing performance of liver fibrosis classification methods. We make the following contributions to the application with this work:

- **Improved performance:** As seen above, our method, *LivSCP*, improves classification performance without any changes to the network architecture of ViT. It outperforms all existing methods and our baseline (ViT with SL) consistently across all evaluation metrics.
- **Solution to Low Data and Computation:** Our method is successful in low-data and -computation settings. We have demonstrated that with a dataset of 6323 images in total (~ 5058 images for training), and an NVIDIA P100 GPU, we found an absolute performance boost of up to 5.38% in accuracy with respect to our baseline and up to 14.93% with an existing method [22].
- **Rigorous performance analysis:** We conduct a rigorous quantitative analysis of the classification performance of existing methods, our baseline, and the proposed method at different precision (quantization) levels.

4.2. Comparison with Baseline

We consider ViT trained with SL as our baseline. We simultaneously implement this baseline along with our method with the same training settings.

In Table 2, a comparison of the classification performance of the proposed method and baseline is illustrated at different precision levels. It can be seen that INT8 performance is generally better than that of the FP32 and FP16, especially in the case of SL. This is a classic case of quantization acting as regularization [43]. We can observe that LivSCP consistently outperforms SL across all precision levels and evaluation metrics. There is an improvement of an absolute 5.38% in accuracy. This is a remarkable improvement since the network architectures, learning rate, initial weights, and optimizer are all the same.

4.3. Comparison with Existing Methods

A comparison of diagnostic performance for liver fibrosis classification using the proposed method with existing methods on the basis of accuracy and AUROC is given in Table 3. Lee et al. [13] reported an accuracy of 83.5 and 76.4 % on internal and external datasets of four classes of liver fibrosis with an AUROC of 0.901 and 0.857 for F4 stage classification. Joo et al. [22] reported a mean accuracy of 84.37% and Park et al. [25] reported a mean accuracy and mean AUROC of 0.95 for liver disease classification.

Table 3. Comparison of diagnostic performance of different methods for liver fibrosis classification.

Method	Accuracy (in %)	AUROC
Lee et al. [13]	83.5 & 76.4	0.901 & 0.857 (F4)
Joo et al. [22]	84.37	—
Park et al. [25]	94	0.95
Proposed	98.10	0.9972

The comparison of classification accuracy obtained using the proposed network with five DL networks from [22] is given in Table 4. ResNet and EfficientNet showed strong performance with an accuracy of 85.92% and 85.17%, respectively. However, the proposed method reported an accuracy of 98.10%, indicating significant improvement in classification performance and demonstrating the effectiveness of the proposed method over existing DL methods.

Table 4. Comparison of accuracy values obtained using the proposed method with five DL networks from [22].

Network	Accuracy (in %)
VGGNet	83.17
ResNet	85.92
DenseNet	84.17
EfficientNet	85.17
ViT (SL)	83.42
Proposed	98.10

Table 5 shows the AUROC values obtained for liver fibrosis classification across stages (F0–F4) using our proposed method and five DL networks from [25]. The proposed method demonstrates superior performance across all fibrosis stages with AUROC values of 1.000 for F0 and F4, 0.999 for F1, 0.998 for F2, and 0.989 for F3, indicating that it achieved better discrimination across all fibrosis stages.

Table 5. Comparison of AUROC values obtained using the proposed method with five DL networks from [25].

Network	F0	F1	F2	F3	F4
VGGNet	0.96	0.96	0.98	0.94	0.96
ResNet	0.96	0.96	0.97	0.93	0.94
DenseNet	0.95	0.96	0.95	0.94	0.96
EfficientNet	0.96	0.96	0.97	0.94	0.96
ViT (SL)	0.97	0.94	0.96	0.94	0.96
Proposed	1	0.999	0.9983	0.98915	1

4.4. Implications

There are several technical and non-technical (social and human) implications of using LivSCP in clinical applications.

- **Explainability:** DL models used for automated disease classification are no less than unexplained black boxes. LivSCP, being one, will have similar characteristics, and we might question its interpretability. This creates immense scope for further research in developing solutions that are inherently explainable and do not need external methods to explain them, such as SHAP [44], GradCAM [45], etc.
- **Case-specificity:** LivSCP has been developed to address scenarios where both data and computational resources are limited. However, its performance may vary and may even improve when more data and greater computing capacity are available. Therefore, our method should not be generalized to broader studies without considering these constraints.
- **Practical Errors:** LivSCP has no solution to reduce the effect of imaging errors and noise, inconsistencies, etc. This creates another opportunity to develop methods that are more error-immune.

5. Conclusions and Future Scope

5.1. Conclusions

Through our experiments, we found that our method, *LivSCP*, delivers superior performance compared to existing methods and our baseline. The performance gains are consistent across all metrics and precision levels. The proposed method achieves a state-of-the-art performance, with an accuracy, precision, recall, and F1-score of 98.10% each, and an AUROC of 0.9972. A major advantage of LivSCP is that it does not require any modification to the network architecture. Our method is particularly well-suited for scenarios with limited labeled data and computational resources. Therefore, we propose a sophisticated approach for training liver fibrosis classification models that achieves high performance even with limited data and computational resources. The proposed method has been compared against various existing techniques and demonstrates strong potential for clinical applications.

5.2. Future Scope

The objectives discussed here may be explored in the future to improve the performance, reliability, and adaptability of liver fibrosis classification methods. We believe that with the availability of multiple data modalities for liver fibrosis classification, developing multimodal methods will become significantly easier. This would help improve classification performance as well as adaptability.

In the future, efforts should also focus on developing methods that can handle imaging errors and inconsistencies more effectively, since these factors strongly influence the efficacy of such systems. Additionally, methods that are inherently explainable should

be prioritized. In medical applications, interpretability plays a crucial role in making any system trustworthy. Such inherently explainable systems will also help us gain deeper insights into the diseases themselves.

Author Contributions: Conceptualization, A.B. and Y.D.; methodology, A.B.; software, A.B.; validation, Y.D., A.B. and P.F.; formal analysis, A.B.; investigation, A.B.; resources, P.F.; data curation, P.F.; writing—original draft preparation, A.B.; writing—review and editing, Y.D. and A.B.; visualization, A.B.; supervision, Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study did not involve direct research on human participants, human tissues, or the collection of new human data. Instead, we used a publicly available and anonymized dataset from Kaggle. The dataset can be accessed at <https://kaggle.com/datasets/vibhingupta028/liver-histopathology-fibrosis-ultrasound-images> (accessed on 6 December 2025). As the data are open-access and fully anonymized, no ethical approval or Institutional Review Board (IRB) clearance was required for this research.

Informed Consent Statement: Informed consent was waived because this study did not involve direct research on human participants, human tissues, or the collection of new human data. Instead, we utilized a publicly available and anonymized dataset from Kaggle.

Data Availability Statement: The data is available publicly at <https://kaggle.com/datasets/vibhingupta028/liver-histopathology-fibrosis-ultrasound-images> (accessed on 6 December 2025). The source code and trained models are public at <https://github.com/adityabhongade/LivSCP/> (accessed on 6 December 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NAFLD	Non-alcoholic Fatty Liver Disease
ARLD	Alcohol Related Liver Disease
DL	Deep Learning
CNN	Convolutional Neural Network
HCC	Hepatocellular Carcinoma
US	Ultrasound
SWE	Shear Wave Elastography
SL	Supervised Learning
UL	Unsupervised Learning
RL	Reinforcement Learning
SSL	Self-Supervised Learning
SCP	Supervised Contrastive Pretraining
SFT	Supervised Fine-tuning
MLP	Multilayer Perceptron
ViT	Vision Transformer
GAP	Global Average Pooling
mAUROC	Mean Area Under the Receiver Operating Characteristic Curve
CW-SCL	Class Weighted Supervised Contrastive Loss

Appendix A

Quantizing Models

We quantize the models to FP16 and INT8 after training or finetuning. For FP16, `half()` method is used directly. In the case of INT8, `quantize_dynamic()` method from `torch.ao.quantization` is used on linear, convolutional, layer normalization, and batch

normalization layers [46]. In dynamic quantization, the weights are quantized in advance, while the activations are dynamically quantized during inference.

References

- Asrani, S.K.; Devarbhavi, H.; Eaton, J.; Kamath, P.S. Burden of liver diseases in the world. *J. Hepatol.* **2019**, *70*, 151–171. [\[CrossRef\]](#)
- Teng, M.L.; Ng, C.H.; Huang, D.Q.; Chan, K.E.; Tan, D.J.; Lim, W.H.; Yang, J.D.; Tan, E.; Muthiah, M.D. Global incidence and prevalence of nonalcoholic fatty liver disease. *Clin. Mol. Hepatol.* **2022**, *29*, S32. [\[CrossRef\]](#)
- Niu, X.; Zhu, L.; Xu, Y.; Zhang, M.; Hao, Y.; Ma, L.; Li, Y.; Xing, H. Global prevalence, incidence, and outcomes of alcohol related liver diseases: A systematic review and meta-analysis. *BMC Public Health* **2023**, *23*, 859.
- Amini-Salehi, E.; Letafatkar, N.; Norouzi, N.; Joukar, F.; Habibi, A.; Javid, M.; Sattari, N.; Khorasani, M.; Farahmand, A.; Tavakoli, S.; et al. Global prevalence of nonalcoholic fatty liver disease: An updated review Meta-Analysis comprising a population of 78 million from 38 countries. *Arch. Med. Res.* **2024**, *55*, 103043. [\[CrossRef\]](#) [\[PubMed\]](#)
- Elhence, A.; Bansal, B.; Gupta, H.; Anand, A.; Singh, T.P.; Goel, A.; Shalimar. Prevalence of non-alcoholic fatty liver disease in India: A systematic review and meta-analysis. *J. Clin. Exp. Hepatol.* **2022**, *12*, 818–829. [\[PubMed\]](#)
- Anton, M.C.; Shanthi, B.; Sridevi, C. Prevalence of non-alcoholic fatty liver disease in urban adult population in a tertiary care center, Chennai. *Indian J. Community Med.* **2023**, *48*, 601–604. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ufere, N.N.; Satapathy, N.; Philpotts, L.; Lai, J.C.; Serper, M. Financial burden in adults with chronic liver disease: A scoping review. *Liver Transplant.* **2022**, *28*, 1920–1935. [\[CrossRef\]](#)
- Zhang, Y.; Luo, M.; Ming, Y. Global burden of cirrhosis and other chronic liver diseases caused by specific etiologies from 1990 to 2021. *BMC Gastroenterol.* **2025**, *25*, 641. [\[CrossRef\]](#)
- Bravo, A.A.; Sheth, S.G.; Chopra, S. Liver biopsy. *N. Engl. J. Med.* **2001**, *344*, 495–500. [\[CrossRef\]](#)
- Rockey, D.C.; Caldwell, S.H.; Goodman, Z.D.; Nelson, R.C.; Smith, A.D. Liver biopsy. *Hepatology* **2009**, *49*, 1017–1044. [\[CrossRef\]](#)
- Schuppan, D.; Afdhal, N.H. Liver cirrhosis. *Lancet* **2008**, *371*, 838–851. [\[CrossRef\]](#) [\[PubMed\]](#)
- Castera, L.; Forns, X.; Alberti, A. Non-invasive evaluation of liver fibrosis using transient elastography. *J. Hepatol.* **2008**, *48*, 835–847. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lee, J.H.; Joo, I.; Kang, T.W.; Paik, Y.H.; Sinn, D.H.; Ha, S.Y.; Kim, K.; Choi, C.; Lee, G.; Yi, J.; et al. Deep learning with ultrasonography: Automated classification of liver fibrosis using a deep convolutional neural network. *Eur. Radiol.* **2020**, *30*, 1264–1273. [\[CrossRef\]](#) [\[PubMed\]](#)
- Anteby, R.; Klang, E.; Horesh, N.; Nachmany, I.; Shimon, O.; Barash, Y.; Kopylov, U.; Soffer, S. Deep learning for noninvasive liver fibrosis classification: A systematic review. *Liver Int.* **2021**, *41*, 2269–2278. [\[CrossRef\]](#)
- Decharatanachart, P.; Chaiteerakij, R.; Tiyyarattanachai, T.; Treeprasertsuk, S. Application of artificial intelligence in chronic liver diseases: A systematic review and meta-analysis. *BMC Gastroenterol.* **2021**, *21*, 10. [\[CrossRef\]](#)
- Kagadis, G.C.; Drazinos, P.; Gatos, I.; Tsantis, S.; Papadimitroulas, P.; Spiliopoulos, S.; Karnabatidis, D.; Theotokas, I.; Zoumpoulis, P.; Hazle, J.D. Deep learning networks on chronic liver disease assessment with fine-tuning of shear wave elastography image sequences. *Phys. Med. Biol.* **2020**, *65*, 215027. [\[CrossRef\]](#)
- Sun, C.; Xu, A.; Liu, D.; Xiong, Z.; Zhao, F.; Ding, W. Deep learning-based classification of liver cancer histopathology images using only global labels. *IEEE J. Biomed. Health Inform.* **2019**, *24*, 1643–1651. [\[CrossRef\]](#)
- Che, H.; Brown, L.G.; Foran, D.J.; Noshier, J.L.; Hacıhaliloglu, I. Liver disease classification from ultrasound using multi-scale CNN. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 1537–1548. [\[CrossRef\]](#)
- Lin, Y.S.; Huang, P.H.; Chen, Y.Y. Deep learning-based hepatocellular carcinoma histopathology image classification: Accuracy versus training dataset size. *IEEE Access* **2021**, *9*, 33144–33157. [\[CrossRef\]](#)
- Al-Hasani, M.; Sultan, L.R.; Sagreya, H.; Cary, T.W.; Karmacharya, M.B.; Sehgal, C.M. Ultrasound radiomics for the detection of early-stage liver fibrosis. *Diagnostics* **2022**, *12*, 2737. [\[CrossRef\]](#)
- Duan, Y.Y.; Qin, J.; Qiu, W.Q.; Li, S.Y.; Li, C.; Liu, A.S.; Chen, X.; Zhang, C.X. Performance of a generative adversarial network using ultrasound images to stage liver fibrosis and predict cirrhosis based on a deep-learning radiomics nomogram. *Clin. Radiol.* **2022**, *77*, e723–e731. [\[CrossRef\]](#) [\[PubMed\]](#)
- Joo, Y.; Park, H.C.; Lee, O.J.; Yoon, C.; Choi, M.H.; Choi, C. Classification of liver fibrosis from heterogeneous ultrasound image. *IEEE Access* **2023**, *11*, 9920–9930. [\[CrossRef\]](#)
- Antony Asir Daniel, V.; Jeha, J. An optimal modified faster region CNN model for diagnosis of liver diseases from ultrasound images. *IETE J. Res.* **2024**, *70*, 3572–3589. [\[CrossRef\]](#)
- Ai, H.; Huang, Y.; Tai, D.I.; Tsui, P.H.; Zhou, Z. 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* **2024**, *24*, 5513. [\[CrossRef\]](#)

25. Park, H.C.; Joo, Y.; Lee, O.J.; Lee, K.; Song, T.K.; Choi, C.; Choi, M.H.; Yoon, C. Automated classification of liver fibrosis stages using ultrasound imaging. *BMC Med. Imaging* **2024**, *24*, 36. [CrossRef]
26. Le, M.H.N.; Kha, H.Q.; Tran, N.M.; Nguyen, P.K.; Huynh, H.H.; Huynh, P.K.; Lam, H.; Le, N.Q.K. Radiomics in liver research: A paradigm shift in disease detection and staging. *Eur. J. Radiol. Artif. Intell.* **2025**, *2*, 100016. [CrossRef]
27. Xia, F.; Wei, W.; Wang, J.; Wang, Y.; Wang, K.; Zhang, C.; Zhu, Q. Ultrasound radiomics-based logistic regression model for fibrotic NASH. *BMC Gastroenterol.* **2025**, *25*, 66. [CrossRef]
28. Balestrieri, R.; Lecun, Y. How Learning by Reconstruction Produces Uninformative Features For Perception. In Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria, 21–27 July 2024; Volume 235, pp. 2566–2585.
29. Gupta, V. Liver Histopathology (Fibrosis) Ultrasound Images. Kaggle. 2024. Available online: <https://www.kaggle.com/datasets/vibhingupta028/liver-histopathology-fibrosis-ultrasound-images> (accessed on 1 January 2025).
30. Bicubic Interpolation. Wikipedia Article. Available online: https://en.wikipedia.org/wiki/Bicubic_interpolation (accessed on 6 December 2025).
31. Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised Contrastive Learning. In *Proceedings of the Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: New York, NY, USA, 2020; Volume 33, pp. 18661–18673.
32. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q. RandAugment: Practical Automated Data Augmentation with a Reduced Search Space. In *Proceedings of the Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: New York, NY, USA, 2020; Volume 33, pp. 18613–18624.
33. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 4 May 2021.
34. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.U.; Polosukhin, I. Attention is All you Need. In *Proceedings of the Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: New York, NY, USA, 2017; Volume 30.
35. Wightman, R. Vision Transformer (ViT) Base Patch16 224 - AugReg + ImageNet-21k + Fine-Tuned on ImageNet-1k. Model Version: Augreg2_in21k_ft_in1k. Available online: https://huggingface.co/timm/vit_base_patch16_224.augreg2_in21k_ft_in1k (accessed on 3 March 2025).
36. Wang, Z.; Liu, L.; Duan, Y.; Kong, Y.; Tao, D. Continual Learning with Lifelong Vision Transformer. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 171–181. [CrossRef]
37. Xue, M.; Zhang, H.; Song, J.; Song, M. Meta-attention for ViT-backed Continual Learning. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 150–159. [CrossRef]
38. Li, D.; Cao, G.; Xu, Y.; Cheng, Z.; Niu, Y. Technical Report for ICCV 2021 Challenge SSLAD-Track3B: Transformers Are Better Continual Learners. *arXiv* **2022**, arXiv:2201.04924.
39. Hendrycks, D.; Gimpel, K. Gaussian Error Linear Units (GELUs). *arXiv* **2023**, arXiv:1606.08415.
40. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1026–1034. [CrossRef]
41. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; Volume 9, pp. 249–256.
42. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
43. AskariHemmat, M.; Hemmat, R.A.; Hoffman, A.; Lazarevich, I.; Saboori, E.; Mastropietro, O.; Sah, S.; Savaria, Y.; David, J.P. QReg: On Regularization Effects of Quantization. *arXiv* **2022**, arXiv:2206.12372. [CrossRef]
44. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 4–9 December 2017; NIPS'17, pp. 4768–4777.
45. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* **2020**, *128*, 336–359. [CrossRef]
46. PyTorch Contributors. torch.ao.quantization.quantize_dynamic. PyTorch Documentation. 2025. Available online: https://docs.pytorch.org/docs/stable/generated/torch.ao.quantization.quantize_dynamic.html (accessed on 2 October 2025).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.