

RESEARCH ARTICLE

Machine learning-based segmentation and quantitative histological assessment of bovine liver fibrosis: Comparative performance of three open-source models



Omar Younis Altaey¹ , Shahbaa Khalil AL-Taee² , Saffanah Khuder Mahmood¹  and Ghada Abdulrahman Sultan¹ 

1. Department of Anatomy, College of Veterinary Medicine, University of Mosul, Mosul 41001, Iraq.

2. Department of Pathology and Poultry Diseases, College of Veterinary Medicine, University of Mosul, Mosul 41001, Iraq.

ABSTRACT

Background and Aim: Conventional semiquantitative histopathological scoring of hepatic fibrosis is affected by interobserver variability and provides limited quantitative information on the spatial distribution and progression of collagen deposition. Machine learning (ML)-based image analysis may provide a reproducible approach for objective fibrosis assessment in veterinary pathology. This study compared three open-source ML models, Trainable Weka Segmentation (TWS), Ilastik, and Fiber-ML, for segmentation and quantitative assessment of bovine liver fibrosis and evaluated their diagnostic performance, agreement with METAVIR staging, and computational efficiency.

Materials and Methods: This cross-sectional observational study included liver samples from 35 cattle diagnosed with hepatic fibrosis according to the clinical sign records. Masson's trichrome-stained sections were classified according to METAVIR stages F0–F4 (F0 represents normal liver samples, n=7). A total of 525 images were obtained from perisinusoidal, pericentral, and periportal regions. Images from 25 animals (375 images) constituted the training/validation dataset with five-fold cross-validation, whereas images from 10 animals (150 images) formed an independent testing dataset. A similar dataset was annotated by experts to generate manual ground-truth segmentation, and images from the same animal appeared in both training and testing datasets. Model performance was assessed using collagen proportionate area (CPA), Dice similarity coefficient (DSC), sensitivity, specificity, accuracy, correlation with METAVIR scores, and computational-resource requirements.

Results: Ilastik and Fiber-ML generally achieved greater segmentation agreement than TWS ($p \leq 0.05$), with DSC values reaching 0.91 and 0.92, compared to 0.87 in TWS, depending on the hepatic region. Overall model pixel-level segmentation accuracy ranged from 84.7% to 91.8%, with the highest performance observed in the periportal region. CPA measurements generated by the models showed strong positive correlations with METAVIR stages ($p = 0.97$ – 0.99). Fiber-ML required substantially less computational resources, while Ilastik recorded fewer software errors than TWS, including shorter training time and lower CPU and RAM utilization. All computational benchmarking was conducted under identical hardware and software conditions.

Conclusion: Open-source ML models enabled objective segmentation and quantitative assessment of bovine hepatic fibrosis. Ilastik and Fiber-ML demonstrated strong agreement with expert-derived reference segmentation and fibrosis staging, while Fiber-ML offered additional computational efficiency. These findings support ML-assisted digital histopathology as a future complementary tool for quantifying and spatially characterizing hepatic fibrosis in cattle. Further validation is required before their reproducibility and diagnostic applicability can be established.

Keywords: bovine liver fibrosis, collagen proportionate area, digital histopathology, hepatic fibrosis, machine learning, METAVIR staging, segmentation, veterinary pathology.

INTRODUCTION

The liver plays a central role in maintaining animal health and productivity through its essential functions in metabolism, nutrient homeostasis, and detoxification [1]. Chronic hepatic injury can result in fibrosis, a pathological process characterized by excessive accumulation of extracellular matrix components, particularly fibrillar collagen [2]. In cattle, persistent hepatocellular injury associated with fluke migration, parasitic

Corresponding Author: Omar Younis Altaey and Shahbaa Khalil AL-Taee

E-mail: Omar.younes@uomosul.edu.iq; Shahbaa_khal@uomosul.edu.iq

Received: 05-07-2026, **Accepted:** 24-09-2026, **Published online:** 17-10-2026

Co-authors: SKM: saffanhjeber@uomosul.edu.iq; GAS: Aliabc898@uomosul.edu.iq

How to cite: Altaey OY, AL-Taee SK, Mahmood SK, Sultan GA. Machine learning-based segmentation and quantitative histological assessment of bovine liver fibrosis: Comparative performance of three open-source models. Vet World. 2026;19(10):4641–4656.

Copyright: Altaey *et al.* This article is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

infestation, toxic exposure, or oxidative stress can promote hepatic fibrosis and ultimately contribute to disruption of normal liver architecture, jaundice, cholecystitis, and portal hypertension [3, 4]. Hepatic injury activates Kupffer cells, which release transforming growth factor-beta (TGF- β), a profibrogenic cytokine that promotes the transformation of vitamin A-storing hepatic stellate cells into proliferative myofibroblasts that secrete collagen into the extracellular matrix [3]. Progressive collagen deposition also contributes to the loss of sinusoidal endothelial fenestrations and subsequent tissue hypoxia, which can further activate hepatic stellate cells and perpetuate the fibrogenic response [4].

Conventional assessment of hepatic fibrosis relies primarily on histopathological examination using semiquantitative scoring systems, including the METAVIR and Ishak systems, which are widely used as reference approaches for fibrosis staging [5]. However, semiquantitative assessment is susceptible to interobserver variability and has limited ability to objectively quantify lesion progression and identify the spatial distribution of early or subclinical fibrotic changes [6]. These limitations are particularly relevant because hepatic fibrosis can exhibit heterogeneous patterns of collagen deposition and bridging across different regions of the liver [3]. Digital image analysis provides an alternative quantitative approach by measuring the collagen proportionate area (CPA), defined as the percentage of total tissue area occupied by collagen (collagen area/total tissue area $\times$ 100). CPA provides an objective numerical estimate of collagen burden within the examined tissue and may improve the accuracy and reproducibility of fibrosis assessment compared with semiquantitative scoring alone [7].

Advances in artificial intelligence and machine learning (ML) have further expanded the potential of digital histopathology by converting qualitative visual assessments into quantitative morphometric measurements that can facilitate fibrosis detection, staging, and monitoring [7, 8]. Although ML-based approaches have demonstrated considerable potential for detecting and quantifying hepatic abnormalities in human medicine and laboratory models, their application in veterinary medicine and animal research remains comparatively limited [9]. Machine learning tools have been applied to veterinary imaging for applications such as ultrasonographic detection of liver disease [9], liver tumor classification [10], and prediction of fatty liver disease [11]. Nevertheless, ML-based image segmentation for quantitative assessment of hepatic fibrosis in cattle remains uncommon, and evidence from other animal models may not directly apply to veterinary diagnosis and monitoring in large animals such as cattle and horses [9–11].

Despite advances in digital image analysis and ML-assisted pathology, an important methodological gap remains in the objective assessment of bovine hepatic fibrosis. In particular, there is limited evidence directly comparing open-source ML tools for segmenting and quantifying collagen deposition in histological sections of bovine liver. The relative ability of these tools to reproduce expert-derived fibrosis measurements across anatomically distinct hepatic regions, including perisinusoidal, pericentral, and periportal areas, remains insufficiently characterized. Furthermore, diagnostic performance alone does not determine the practical applicability of an ML tool; computational requirements, processing efficiency, and operational stability are also important considerations for implementation in veterinary pathology laboratories. Therefore, comparative evaluation against manual reference segmentation and an established semiquantitative fibrosis staging system is required to determine whether accessible open-source ML approaches can provide reliable and computationally practical quantitative assessment of bovine hepatic fibrosis.

This study compared three open-source ML models, Trainable Weka Segmentation (TWS), Ilastik, and Fiber-ML, for segmenting and quantitatively assessing fibrosis in Masson's trichrome-stained bovine liver sections. Specifically, the study evaluated their ability to quantify CPA in perisinusoidal, pericentral, and periportal regions and assessed segmentation agreement using the Dice similarity coefficient (DSC). Diagnostic performance was evaluated using sensitivity, specificity, and accuracy against expert-derived manual segmentation, while the relationship between quantitative CPA measurements and semiquantitative METAVIR fibrosis stages was examined to determine agreement with conventional fibrosis assessment. In addition, the models were compared in terms of computational performance, including training time, central processing unit (CPU) utilization, random-access memory (RAM) utilization, and software crashes, to assess their potential practicality for ML-assisted digital histopathological evaluation of bovine hepatic fibrosis.

MATERIALS AND METHODS

Ethical approval

The study was approved by the Institutional Animal Care and Use Committee (IACUC), College of Veterinary Medicine, University of Mosul, Iraq, under approval number UM.VET.2025.108, dated 01-09-2025. The liver

samples were obtained from condemned cattle carcasses during routine slaughterhouse inspection. No animals were euthanized or slaughtered specifically for this study. All laboratory procedures were conducted in accordance with the standard laboratory practices of the College of Veterinary Medicine, University of Mosul.

Study period and location

The study was conducted from October 2025 to February 2026 in the Anatomy and Histotechnology Laboratories, College of Veterinary Medicine, University of Mosul, Mosul, Iraq (36.385313° N, 43.144403° E).

Study design

This cross-sectional observational study evaluated three machine learning (ML)-based image segmentation models: ImageJ Trainable Weka Segmentation (TWS; version 21.0.7) https://github.com/fiji/Trainable_Segmentation, Ilastik (version 1.4.2) <https://www.ilastik.org/>, and Fiber-ML (version 1.0) <https://gitlab.com/balvayda/fiber-ml>, for the segmentation and quantitative assessment of bovine liver fibrosis. The models were compared with expert-derived manual segmentation and evaluated across the perisinusoidal, pericentral, and periportal regions of the liver.

Sample collection and histological processing

Liver samples were collected from 35 cattle (*Bos taurus*) diagnosed with hepatic fibrosis by the Department of Veterinary Sanitary Inspection at the municipal slaughterhouse of Mosul Governorate, Iraq. The sampled livers had been condemned as unfit for human consumption during slaughterhouse inspection. Samples exhibiting autolysis, large abscesses, cystic lesions, or neoplastic lesions were excluded.

Tissue samples were fixed in 10% neutral-buffered formalin (Sigma Aldrich, HT501850, USA) for 48 h before histological processing [12]. After fixation, the tissues were dehydrated sequentially in 70%, 80%, and 95% ethanol (Scharlau, ET00052500, Spain) for 1 h at each concentration, followed by three changes of 100% ethanol for 1 h each. The samples were then cleared in two changes of xylene (Scharlab, XI0052, Spain) for 45 min each and infiltrated with paraffin wax in three changes for 1.5 h each [12]. Paraffin-embedded tissue blocks were sectioned at a thickness of 5 µm using a rotary microtome (BIOBASE BK-2218, China) [12].

Staining

Tissue sections were deparaffinized and rehydrated through graded ethanol concentrations of 100%, 95%, and 70% for 2 min each. The sections were then placed in Bouin's solution (Epredia, 88038, USA) for 1 h at 56–60°C, rinsed under tap water for 15 min, stained with Weigert's iron hematoxylin for 10 min, rinsed, and stained with Biebrich scarlet-acid solution for 10 min. Subsequently, the sections were differentiated in phosphomolybdic-phosphotungstic acid for 10 min, and collagen was stained with aniline blue for 5 min. The slides were rinsed and immersed in 1% acetic acid for 2 min to stabilize the staining, dehydrated in ethanol, cleared in xylene for 2 min, mounted with Distyrene Plasticizer Xylene (DPX), Masson's trichrome stain kit supplied by (Epredia, 22110648, USA), and allowed to dry overnight at room temperature (25°C) [12].

Image acquisition

A total of 175 histological slides were prepared, comprising five slides from each of the 35 animals. Three regions of interest (ROIs), perisinusoidal, pericentral, and periportal regions, were randomly selected for image acquisition to minimize selection bias. A total of 525 images were captured at 400× magnification using an 18-megapixel microscope-mounted camera (OMAX, A35100U3, China). Each animal contributed 15 images (five slides × three ROIs). The images were subsequently categorized according to the METAVIR fibrosis scoring system [13] (Figure 1).

Ground truth design and analysis

Images from 10 animals (n = 150 images) were used for observer evaluation and were excluded from model training and validation. The observers were blinded to animal history, METAVIR scores, and ML model findings. Three certified histopathologists independently annotated the image dataset using Fiji/ImageJ version 1.54 (National Institutes of Health, USA). Histogram matching was applied to normalize the images and minimize staining variability [14]. Binary masks distinguishing fibrous from nonfibrous tissue were generated, while artifacts, tissue folds, and debris were excluded.

Discrepancies among observers were resolved through consensus discussion. Interobserver agreement was assessed using the Dice similarity coefficient (DSC). The resulting DSC values ranged from 0.83 to 0.91, representing excellent agreement [15]. The consensus annotations were subsequently used to establish the reference ground truth for model evaluation (Figure 2).

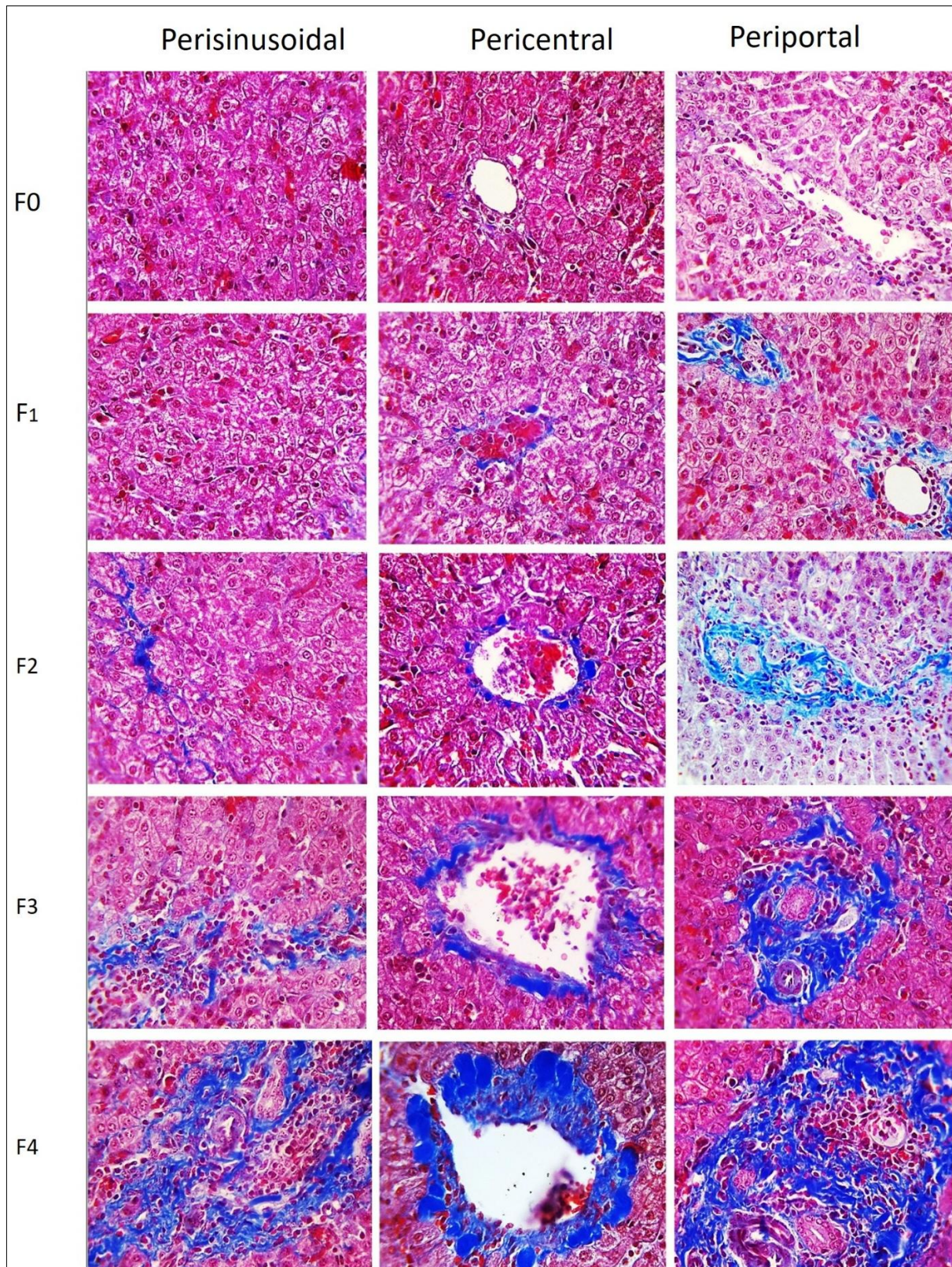


Figure 1: Histological staging of progressive liver fibrosis (F0-F4) in the perisinusoidal, pericentral, and periportal regions. F0: No fibrosis, with normal architecture in the periportal and pericentral regions. F1: Mild fibrosis characterized by thin, unconnected septa in the perisinusoidal, pericentral, and periportal regions. F2: Fibrosis characterized by thick but unconnected septa developing in all regions. F3: Marked fibrosis characterized by numerous septa connecting periportal regions to adjacent portal tracts and central veins (portal-to-central bridging), with central vein wall thickening and collagen deposition in the perisinusoidal region. F4: Advanced fibrosis involving all regions, with thick septa in the periportal regions encircling distorted portal structures, thickened vein walls, and marked perisinusoidal collagen deposition. Sections were stained with Masson's trichrome stain (scale bar = 100 μ m). F0–F4 = METAVIR fibrosis stages, where F0 = no fibrosis, F1 = mild fibrosis, F2 = moderate fibrosis, F3 = advanced fibrosis with bridging septa, and F4 = cirrhosis or advanced architectural distortion. Masson's trichrome staining identifies collagen fibers in blue and surrounding tissue predominantly in red or pink.

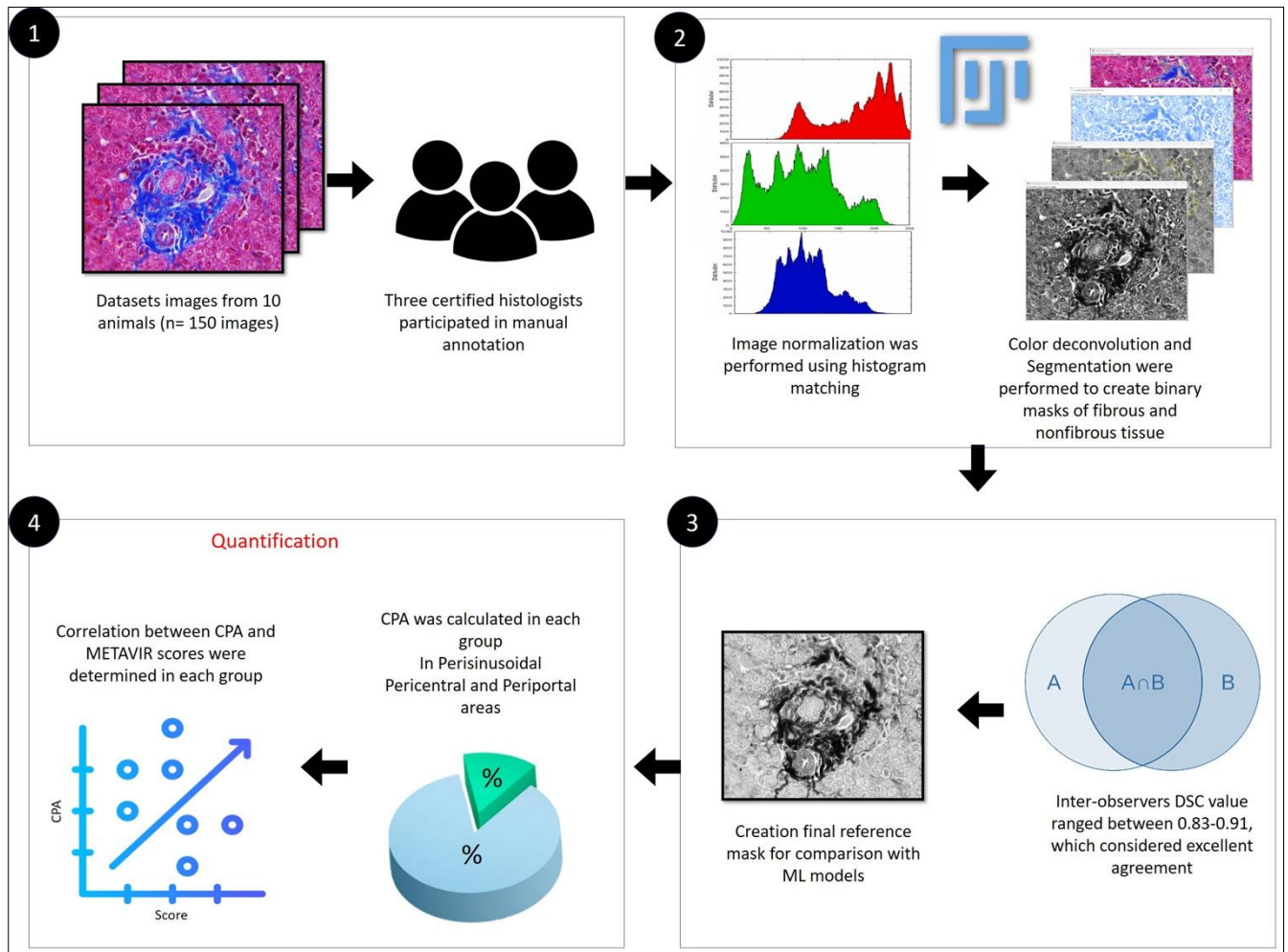


Figure 2: Ground truth reference design comprising: (1) expert annotation and scoring of image groups, with 150 images manually annotated; (2) image normalization and digital processing, including color deconvolution and segmentation-mask generation; (3) assessment of interobserver agreement using DSC to generate the final consensus reference mask; (4) quantification of CPA; and (5) correlation of CPA measurements with different METAVIR scores. DSC = Dice similarity coefficient; CPA = collagen proportionate area; ML = machine learning; METAVIR = histological fibrosis staging system (F0–F4). Ground-truth reference masks were established through consensus annotations by three certified histopathologists. DSC values range from 0 (no overlap) to 1 (perfect overlap).

Machine learning analysis

Data splitting: To prevent data leakage, the image dataset was divided at the animal level into independent training/validation and testing datasets. The training/validation dataset comprised 25 animals (n = 375 images) and was used exclusively for model training and validation. The independent testing dataset comprised 10 animals (n = 150 images) and was withheld from model training and validation for final model evaluation; the testing set is a similar ground truth set used for subsequent comparison. Images from an individual animal were not included in more than one dataset.

Five-fold cross-validation was performed within the training dataset to optimize model hyperparameters, including the number of trees, feature scales, probability thresholds, and other model-specific settings. In each fold, images from 20 cattle (n = 300 images) were used for training and images from five cattle (n = 75 images) were used for validation (Figure 3).

Model validation: All images were resized to 2048 × 1536 pixels in JPEG format and normalized before analysis. The three ML tools used random forest classifiers for pixel-wise segmentation and were configured as follows (Figure 4):

Parameter	TWS [16]	Ilastik [17]	Fiber-ML [18]
Version	ImageJ TWS-based 21.0.7	1.4.2	1.0
Algorithm	Random forest	Random forest	Random forest
Features	Gaussian blur, Sobel filter, Hessian	Gaussian smoothing, Laplacian of	Default feature set

Parameter	TWS [16]	Illastik [17]	Fiber-ML [18]
	eigenvalues, membrane patch features	Gaussian, structure tensor eigenvalues	
Feature scales (σ , pixels)	1.0-16.0	0.70, 1.00, 1.60, 3.50, 5.00, 10.00	Default
Classifier settings	200 trees, maximum depth 0	Probability threshold 0.5	200 trees
Output	Segmentation mask	Probability map thresholded to a segmentation mask	Binary segmentation mask with morphometric parameters, including circularity, solidity, and aspect ratio

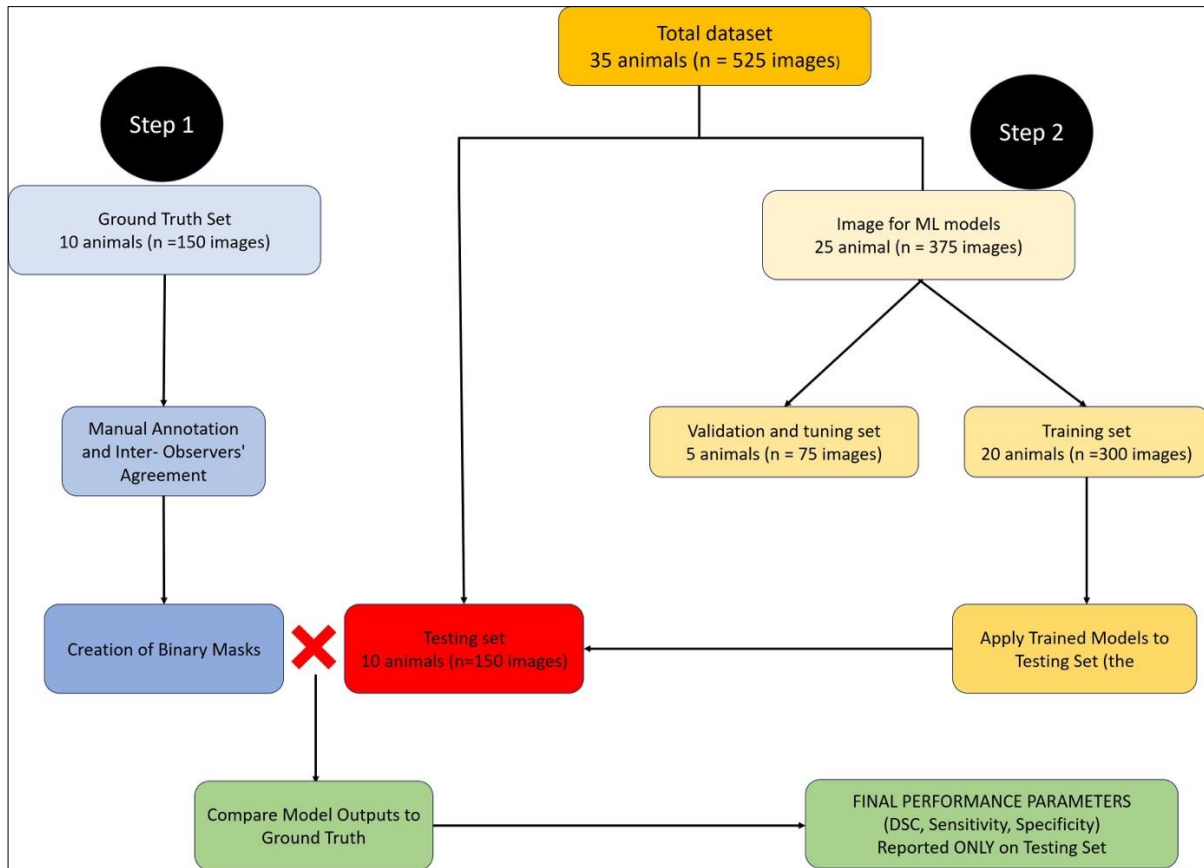


Figure 3: Flowchart illustrating animal-level splitting of the image dataset. The training and validation dataset comprised 375 images from 25 animals and was subjected to five-fold cross-validation. An independent dataset comprising 150 images from 10 animals was withheld for final testing and comparison of ML-generated segmentation masks with manually generated reference masks. ML = machine learning; DSC = Dice similarity coefficient. The dataset comprised 525 histological images from 35 cattle, with 375 images from 25 animals allocated to training and validation and 150 images from 10 animals reserved for independent testing. Five-fold cross-validation was performed within the training/validation dataset, with animal-level separation maintained between datasets.

Quantification: Color deconvolution was performed to separate collagen staining from background color vectors [14]. CPA was calculated for each METAVIR score within the three ROIs from each segmentation mask [19] using the following formula:

$$\text{CPA (\%)} = (\text{total detected collagen pixels} / \text{total tissue area pixels}) \times 100$$

Segmentation accuracy analysis: Segmentation agreement between each ML-generated mask and the reference ground-truth mask was quantified using DSC. The Fiji TNTPNFP_score plugin was used to measure spatial overlap between the automated and manual segmentation masks. DSC was calculated as follows [20]:

$$\text{DSC} = 2 \times |A \cap B| / (|A| + |B|)$$

where A represents the automated segmentation, B represents the manual segmentation, $|A \cap B|$ represents their intersection, and $|A|$ and $|B|$ represent the total areas of the automated and manual segmentations, respectively. DSC ranges from 0, indicating no spatial overlap, to 1, indicating perfect overlap. The DSC was calculated for each image generated using the three ML tools.

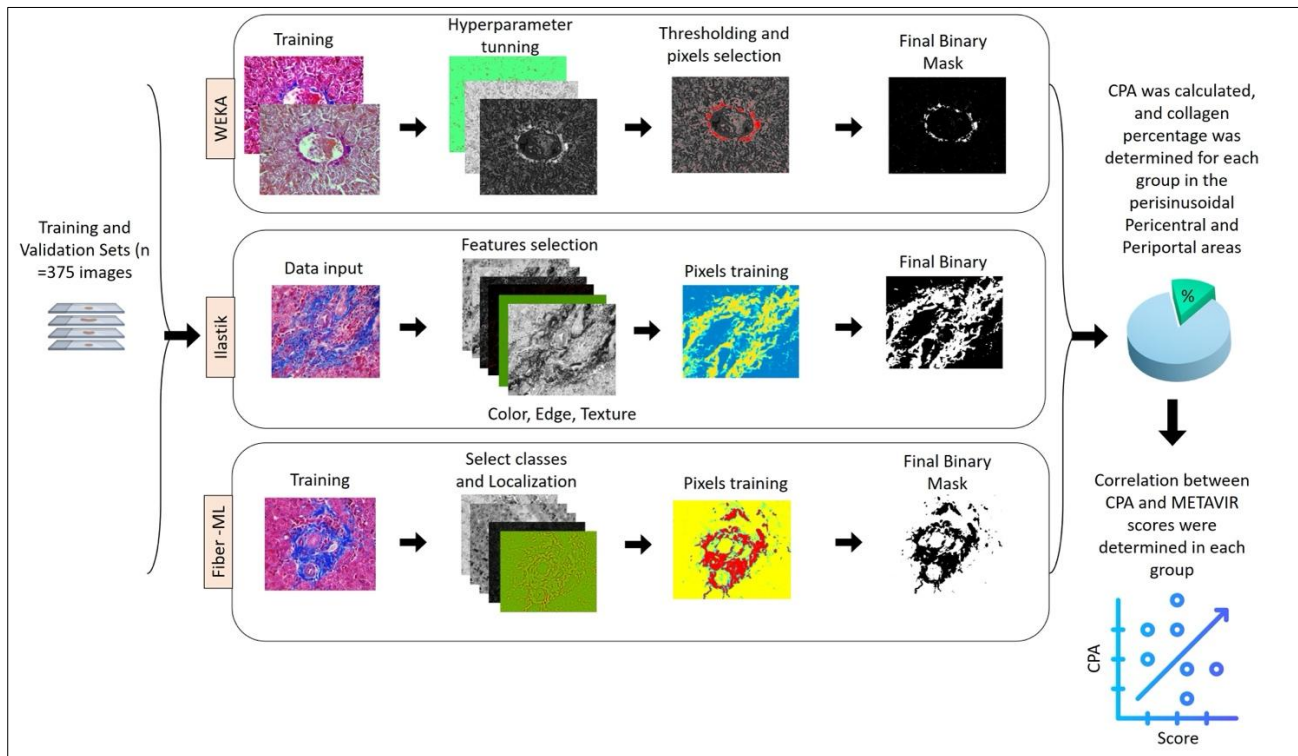


Figure 4: Comparative ML pipeline for liver fibrosis segmentation. The TWS, Ilastik, and Fiber-ML models included: (1) pixel training; (2) segmentation processing based on edge, color, texture, pixel intensity, and orientation features; (3) segmentation-mask generation; (4) thresholding and patching of the datasets; (5) quantification of CPA in the perisinusoidal, pericentral, and periportal regions; and (6) correlation of CPA measurements with different METAVIR scores. ML = machine learning; TWS = Trainable Weka Segmentation; CPA = collagen proportionate area; METAVIR = histological fibrosis staging system (F0–F4). The three open-source segmentation models were applied to histological images to generate binary collagen masks and quantify collagen deposition in the perisinusoidal, pericentral, and periportal hepatic regions.

Performance analysis: Diagnostic performance was evaluated by pixel-wise comparison of ML-generated segmentation masks with the corresponding manual reference annotations. Sensitivity, specificity, false-positive rate, false-negative rate, and accuracy were calculated using standard classification metrics [21].

Metric	Formula	Interpretation
Sensitivity	$TP/(TP + FN)$	Proportion of actual fibrous areas correctly identified
Specificity	$TN/(TN + FP)$	Proportion of actual nonfibrous areas correctly identified
False-positive rate	$FP/(FP + TN)$	Proportion of nonfibrous areas incorrectly classified as fibrous
False-negative rate	$FN/(FN + TP)$	Proportion of fibrous areas incorrectly classified as nonfibrous
Accuracy	$(TP + TN)/total$	Overall correctness of model segmentation

TP = true positive; TN = true negative; FP = false positive; FN = false negative.

Computational resources analysis: Training time (ms), central processing unit (CPU) utilization (%), random-access memory (RAM) utilization (%), and the number of software crashes were recorded under standardized computational conditions using a single workstation equipped with an Intel Core i5-1334U processor at 2.30 GHz, 32 GB DDR4 RAM, a 512 GB solid-state drive, and Windows 11 Pro 64-bit. Computational-resource assessment was performed using the training dataset of 375 images, each measuring 2048 × 1536 pixels in JPEG format.

Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics version 28 (IBM Corp., Armonk, NY, USA). Data normality was assessed using the Shapiro-Wilk test, and homogeneity of variances was evaluated using Levene's test. CPA (%), DSC, sensitivity (%), and specificity (%) were compared among the three ML tools using repeated-measures one-way analysis of variance (ANOVA), accounting for within-animal comparisons because the models were evaluated using the animal experimental units. When significant differences were detected, Tukey's honestly significant difference post hoc test was used for pairwise comparisons [22].

Associations between CPA (%) and METAVIR scores were evaluated using Spearman's rank correlation coefficient (ρ). All statistical tests were two-tailed, and statistical significance was set at $p < 0.05$. Data are presented as the mean ± standard error of the mean (SEM) [22].

RESULTS

Postmortem and microscopic findings

Slaughterhouse inspection records indicated that the examined cattle had histories of emaciation, anorexia, and diarrhea, accompanied by icterus and diffuse body edema. The animals originated from different locations within Mosul Governorate and were presented for slaughter. All examined cases were condemned and considered unfit for human consumption by the slaughterhouse inspection committee.

Gross examination revealed firm, shrunken livers with nodular and granular surfaces. The hepatic parenchyma was dark reddish-brown, with greenish patches observed in some specimens. Microscopic examination revealed varying degrees of fibrous tissue deposition, and the lesions were categorized according to the METAVIR scoring system.

At F0, no fibrosis was observed in the examined regions. The periportal spaces exhibited normal architecture without collagen deposition; the perisinusoidal spaces showed no excess collagen, and the central veins showed no fibrous thickening or scarring. At F1, mild collagen deposition was observed around the central veins and periportal regions. Collagen deposition was confined predominantly to the walls of the portal tracts, causing their expansion, while the perisinusoidal spaces remained uninvolved. At F2, slender, finger-like fibrous septa developed in the periportal regions; these septa were few and remained unconnected, with occasional deposition of collagen fibers in the perisinusoidal spaces. Pericentral zones showed minimal collagen deposition without marked bridging. At F3, marked fibrosis was observed, with numerous septa extending from periportal regions to adjacent portal tracts. The pericentral regions exhibited septa extending toward the central veins, resulting in portal-to-central bridging, together with early fibrous thickening of the central vein walls. Mild-to-moderate collagen deposition was also observed in the perisinusoidal regions. At F4, severe fibrosis involved the periportal, pericentral, and perisinusoidal regions. Thick periportal fibrous septa encircled regenerative nodules and markedly distorted the portal tract architecture. Prominent perisinusoidal collagen deposition was evident, while the pericentral veins were thickened and frequently incorporated into the fibrous bands separating the nodules (Figure 5).

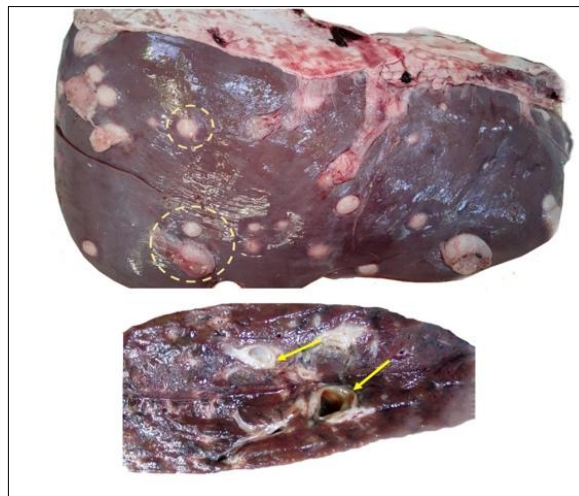


Figure 5: Gross appearance of the liver showing a firm, shrunken organ with a nodular and granular surface (dashed circles). The parenchyma shows dark reddish-brown discoloration with interspersed greenish patches and thickening of the bile duct walls (arrows).

CPA findings

In the perisinusoidal region, no significant differences were observed between manual segmentation and the ML models at F0 and F1 ($p = 0.14$ and $p = 0.23$, respectively). At F0, CPA values were $0.71 \pm 0.13\%$ for manual segmentation, $0.58 \pm 0.13\%$ for TWS, $0.64 \pm 0.12\%$ for Ilastik, and $0.69 \pm 0.18\%$ for Fiber-ML. At F1, the corresponding values were $2.21 \pm 0.15\%$, $1.83 \pm 0.15\%$, $1.97 \pm 0.13\%$, and $2.02 \pm 0.12\%$, respectively. At F2, Ilastik ($5.07 \pm 0.37\%$) and Fiber-ML ($5.12 \pm 0.37\%$) yielded higher CPA values than manual segmentation ($3.65 \pm 0.14\%$) and TWS ($3.22 \pm 0.17\%$) ($p = 0.018$). At F3, all three ML models yielded lower CPA values than manual segmentation (manual, $12.59 \pm 0.72\%$; TWS, $10.91 \pm 0.72\%$; Ilastik, $8.30 \pm 0.22\%$; and Fiber-ML, $8.35 \pm 0.22\%$) ($p = 0.37$). At F4, CPA values from Ilastik ($35.01 \pm 0.38\%$) and Fiber-ML ($35.06 \pm 0.38\%$) were close to the manual value ($36.52 \pm 0.32\%$), whereas TWS yielded a lower value ($26.45 \pm 0.23\%$) ($p = 0.011$) (Figure 6).

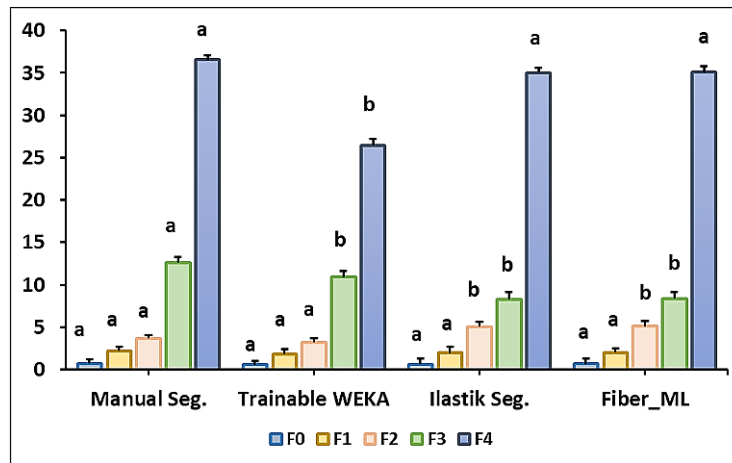


Figure 6: CPA percentages obtained using manual segmentation, TWS, Ilastik, and Fiber-ML across different METAVIR scores in the perisinusoidal region. CPA = collagen proportionate area; TWS = Trainable Weka Segmentation; F0–F4 = METAVIR fibrosis stages, ranging from no fibrosis (F0) to advanced fibrosis (F4). Values represent mean $\pm$ standard error of the mean. Different lowercase letters (a, b) above bars within the same fibrosis stage indicate statistically significant differences among segmentation methods within each METAVIR score ($p < 0.05$).

In the pericentral region, no significant differences were observed at F0 and F1 ($p = 0.12$ and $p = 0.10$, respectively). At F0, CPA values were $1.23 \pm 0.13\%$ for manual segmentation, $0.83 \pm 0.13\%$ for TWS, $1.03 \pm 0.13\%$ for Ilastik, and $1.23 \pm 0.12\%$ for Fiber-ML. At F1, the corresponding values were $2.46 \pm 0.20\%$, $1.94 \pm 0.06\%$, $2.68 \pm 0.27\%$, and $2.88 \pm 0.27\%$, respectively. At F2, the Fiber-ML value ($4.87 \pm 0.11\%$) was close to the manual value ($5.15 \pm 0.21\%$), while TWS and Ilastik yielded $3.33 \pm 0.31\%$ and $3.61 \pm 0.11\%$, respectively ($p = 0.024$). At F3, CPA values were $12.79 \pm 0.59\%$ for manual segmentation, $11.93 \pm 0.43\%$ for TWS, $11.29 \pm 0.21\%$ for Ilastik, and $11.49 \pm 0.21\%$ for Fiber-ML ($p = 0.17, 0.15, 0.17, 0.16$). At F4, Ilastik ($26.19 \pm 0.38\%$) and Fiber-ML ($26.35 \pm 0.41\%$) produced values closer to manual segmentation ($28.70 \pm 0.28\%$) than TWS ($21.43 \pm 0.60\%$) ($p = 0.019$) (Figure 7).

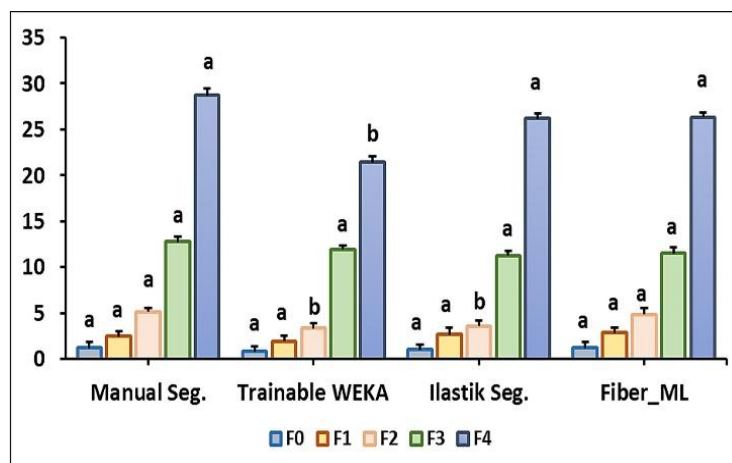


Figure 7: CPA percentages obtained using manual segmentation, TWS, Ilastik, and Fiber-ML across different METAVIR scores in the pericentral region. CPA = collagen proportionate area; TWS = Trainable Weka Segmentation; F0–F4 = METAVIR fibrosis stages, ranging from no fibrosis (F0) to advanced fibrosis (F4). Values represent mean $\pm$ standard error of the mean. Different lowercase letters (a, b) above bars within the same fibrosis stage indicate statistically significant differences among segmentation methods within each METAVIR score ($p < 0.05$).

In the periportal region, manual segmentation and the ML models yielded comparable CPA values at F0 (manual, $1.56 \pm 0.05\%$; TWS, $1.16 \pm 0.05\%$; Ilastik, $1.36 \pm 0.05\%$; and Fiber-ML, $1.46 \pm 0.05\%$) and F1 (manual, $2.97 \pm 0.13\%$; TWS, $2.09 \pm 0.09\%$; Ilastik, $2.27 \pm 0.17\%$; and Fiber-ML, $2.37 \pm 0.17\%$). At F2, the values were $6.64 \pm 0.03\%$ for manual segmentation, $5.10 \pm 0.23\%$ for TWS, $6.06 \pm 0.05\%$ for Ilastik, and $6.16 \pm 0.07\%$ for Fiber-ML. At F3, the corresponding values were $15.69 \pm 0.42\%$, $13.72 \pm 0.21\%$, $15.01 \pm 0.32\%$, and $15.11 \pm 0.32\%$, respectively. Ilastik and Fiber-ML produced values comparable to manual segmentation at F2 and F3 ($p = 0.26$ and $p = 0.31$, respectively). At F4, CPA values were $38.10 \pm 0.70\%$ for manual segmentation, $27.37 \pm 0.95\%$ for TWS, $34.41 \pm 0.23\%$ for Ilastik, and $33.54 \pm 0.23\%$ for Fiber-ML (Figure 8).

CPA was strongly positively correlated with METAVIR score across the examined ROIs for manual

segmentation and all three ML models ($p = 0.97-0.99$). Simple linear regression slopes ranged from $\beta = 6.09-8.14$ in the perisinusoidal region, $\beta = 5.12-6.52$ in the pericentral region, and $\beta = 6.40-8.57$ in the periportal region ($p \leq 0.001$) (Figure 9).

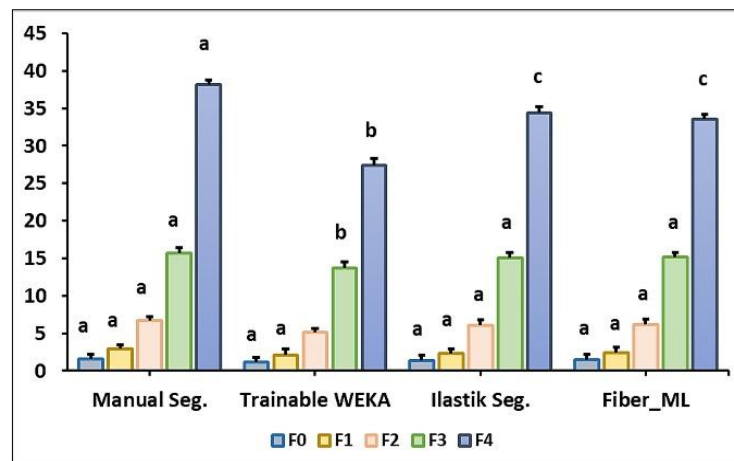


Figure 8: CPA percentages obtained using manual segmentation, TWS, Ilastik, and Fiber-ML across different METAVIR scores in the periportal region. Different letters among values within the same score indicate significant differences. CPA = collagen proportionate area; TWS = Trainable Weka Segmentation; F0–F4 = METAVIR fibrosis stages, ranging from no fibrosis (F0) to advanced fibrosis (F4). Values represent mean $\pm$ standard error of the mean. Different lowercase letters (a, b, c) above bars within the same fibrosis stage indicate statistically significant differences among segmentation methods within each METAVIR score ($p < 0.05$).

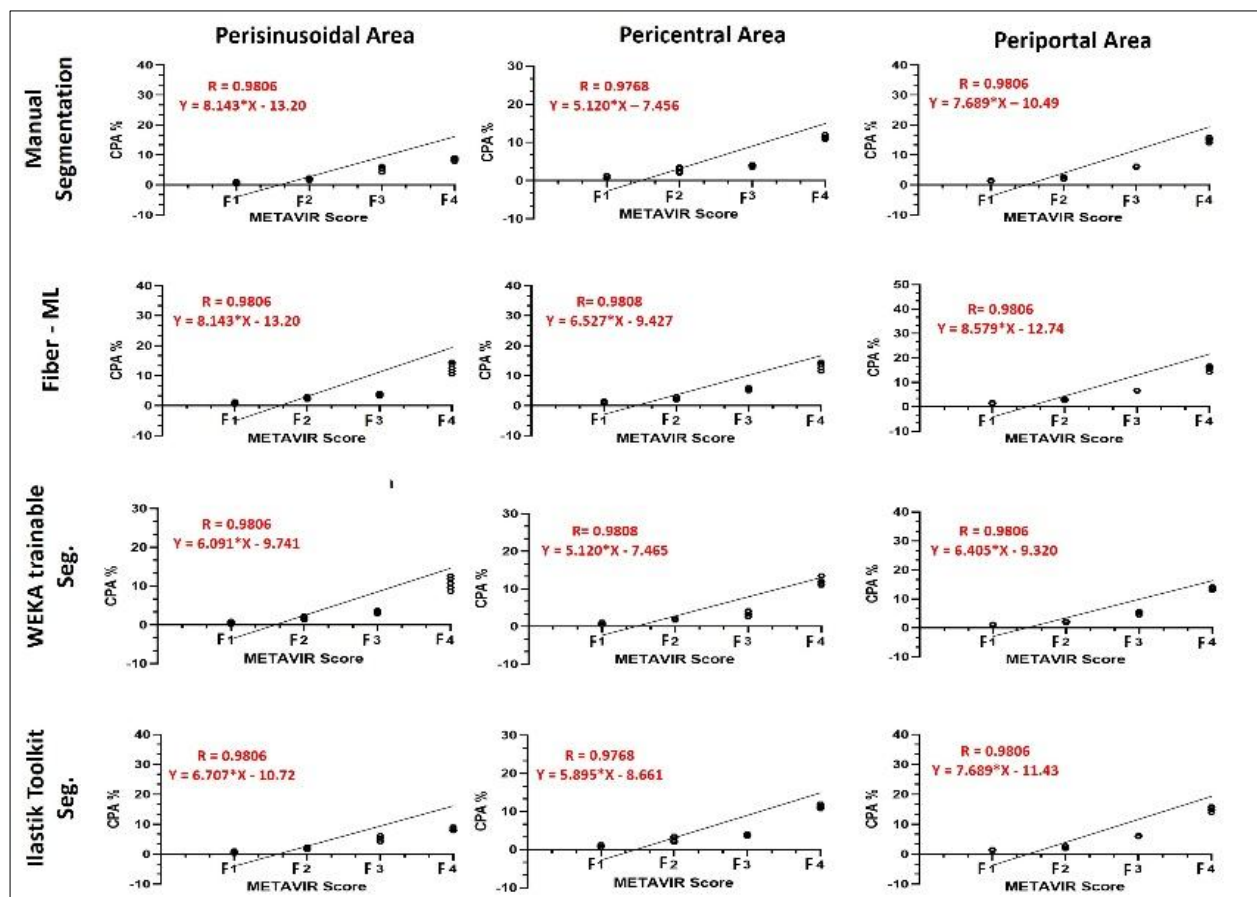


Figure 9: Correlations between CPA values and METAVIR scores in the perisinusoidal, pericentral, and periportal regions. Manual segmentation and the ML models showed strong positive correlations between CPA and METAVIR score ($p = 0.97-0.99$). CPA = collagen proportionate area; ML = machine learning; TWS = Trainable Weka Segmentation; METAVIR = histological fibrosis staging system (F0–F4); ρ = Spearman's rank correlation coefficient; R = correlation coefficient as displayed in the regression panels. Regression equations describe the relationship between CPA (%) and fibrosis stage. The horizontal axis represents METAVIR fibrosis stages, and the vertical axis represents CPA (%). Statistical significance was set at $p < 0.05$.

Segmentation and performance findings

Segmentation agreement was assessed using DSC across the three hepatic regions. In the perisinusoidal region, Ilastik and Fiber-ML yielded DSC values of 0.91 and 0.90, respectively, compared with 0.87 for TWS ($p = 0.021$ and $p = 0.023$). In the pericentral region, DSC values were 0.89 for Ilastik and 0.90 for Fiber-ML, compared with 0.84 for TWS ($p = 0.037$ and $p = 0.021$). In the periportal region, DSC values ranged from 0.90 to 0.92, with no significant differences among the three models (Table 1).

Table 1: DSC values across ML models in the perisinusoidal, pericentral, and periportal ROIs

Segmentation method	Perisinusoidal area, mean $\pm$ SEM	Pericentral area, mean $\pm$ SEM	Periportal area, mean $\pm$ SEM
TWS	0.87 $\pm$ 0.01 ^a	0.84 $\pm$ 0.01 ^a	0.90 $\pm$ 0.04 ^a
Ilastik	0.91 $\pm$ 0.02 ^b	0.89 $\pm$ 0.03 ^b	0.91 $\pm$ 0.02 ^a
Fiber-ML	0.90 $\pm$ 0.02 ^b	0.90 $\pm$ 0.02 ^b	0.92 $\pm$ 0.02 ^a

DSC = Dice similarity coefficient; ML = machine learning; ROI = region of interest; TWS = Trainable Weka Segmentation; SEM = standard error of the mean. Values are presented as mean $\pm$ SEM. DSC ranges from 0 (no spatial overlap) to 1 (perfect spatial overlap) between automated and manually annotated segmentation masks. Different superscript letters (^a, ^b) within the same column indicate statistically significant differences among segmentation methods ($p < 0.05$). Values sharing the same superscript letter are not significantly different.

In the perisinusoidal region, sensitivity did not differ significantly among the models and ranged from 89.8% to 90.8%. Specificity was higher for Ilastik and Fiber-ML than for TWS ($p = 0.019$ and $p = 0.21$). The false-negative rate ranged from 8.6% to 10.6%, whereas the false-positive rate ranged from 7.1% to 14.7%. In the pericentral region, sensitivity and specificity were higher for Ilastik and Fiber-ML than for TWS ($p = 0.043$ and $p = 0.041$). The false-negative rate ranged from 10.8% to 12.7%, and the false-positive rate ranged from 9.1% to 17.8%.

In the periportal region, sensitivity ranged from 89.9% to 91.3% and specificity from 90.5% to 92.3%, with no significant differences among the models. False-negative and false-positive rates ranged from 8.8% to 10.0% and 7.6% to 9.5%, respectively. Overall accuracy across the three models and hepatic regions ranged from 84.7% to 91.8%, with the highest numerical accuracy observed for Fiber-ML in the periportal region (91.8%) (Table 2).

Table 2: Performance evaluation of ML models in the perisinusoidal, pericentral, and periportal ROIs.

Region	Method	Sensitivity, mean $\pm$ SEM (%)	Specificity, mean $\pm$ SEM (%)	False-negative rate, mean $\pm$ SEM (%)	False-positive rate, mean $\pm$ SEM (%)	Accuracy, mean $\pm$ SEM (%)
Perisinusoidal	TWS	90.8 $\pm$ 0.34 ^a	86.1 $\pm$ 0.47 ^a	8.6 $\pm$ 0.25 ^a	14.7 $\pm$ 0.40 ^a	88.4 $\pm$ 0.31 ^a
	Ilastik	89.8 $\pm$ 0.29 ^a	92.6 $\pm$ 0.26 ^b	10.6 $\pm$ 0.23 ^a	7.1 $\pm$ 0.20 ^b	91.2 $\pm$ 0.21 ^a
	Fiber-ML	90.1 $\pm$ 0.31 ^a	91.4 $\pm$ 0.29 ^b	10.1 $\pm$ 0.24 ^a	8.5 $\pm$ 0.24 ^b	90.7 $\pm$ 0.23 ^a
Pericentral	TWS	86.6 $\pm$ 0.45 ^a	83.0 $\pm$ 0.49 ^a	12.7 $\pm$ 0.34 ^a	17.8 $\pm$ 0.45 ^a	84.7 $\pm$ 0.37 ^a
	Ilastik	88.9 $\pm$ 0.33 ^{ab}	89.0 $\pm$ 0.37 ^b	11.1 $\pm$ 0.29 ^a	11.0 $\pm$ 0.33 ^b	89.0 $\pm$ 0.26 ^b
	Fiber-ML	89.3 $\pm$ 0.32 ^b	90.7 $\pm$ 0.31 ^b	10.8 $\pm$ 0.24 ^a	9.1 $\pm$ 0.26 ^b	90.0 $\pm$ 0.22 ^b
Periportal	TWS	89.9 $\pm$ 0.33 ^a	90.5 $\pm$ 0.32 ^a	10.0 $\pm$ 0.26 ^a	9.5 $\pm$ 0.28 ^a	90.2 $\pm$ 0.24 ^a
	Ilastik	91.1 $\pm$ 0.29 ^a	91.3 $\pm$ 0.30 ^a	8.9 $\pm$ 0.21 ^a	8.7 $\pm$ 0.23 ^a	91.2 $\pm$ 0.20 ^a
	Fiber-ML	91.3 $\pm$ 0.28 ^a	92.3 $\pm$ 0.26 ^a	8.8 $\pm$ 0.20 ^a	7.6 $\pm$ 0.21 ^a	91.8 $\pm$ 0.19 ^a

ML = machine learning; ROI = region of interest; TWS = Trainable Weka Segmentation; SEM = standard error of the mean. Values are presented as mean $\pm$ SEM (%). Sensitivity represents the proportion of correctly identified fibrous pixels, whereas specificity represents the proportion of correctly identified nonfibrous pixels. The false-negative and false-positive rates indicate the proportions of incorrectly classified fibrous and nonfibrous pixels, respectively. Accuracy represents the overall proportion of correctly classified pixels. Different superscript letters (^a, ^b) within the same hepatic region and performance parameter indicate statistically significant differences among segmentation methods ($p < 0.05$). Values sharing at least one superscript letter are not significantly different.

Computational analysis findings

Computational performance differed among the three ML models. TWS required the longest training time at 42,000 ms, compared with 7,700 ms for Ilastik and 3,000 ms for Fiber-ML. TWS also showed the highest CPU and RAM utilization, at 92.0% and 31.20%, respectively. In comparison, CPU and RAM utilization were 24.10% and 18.70% for Ilastik, and 19.70% and 17.60% for Fiber-ML.

Across 50 software runs, TWS recorded five errors or freezes (10%), Ilastik recorded two (4%), and Fiber-ML recorded three (6%). These assessments were performed in a Windows 11 Pro environment using an Intel Core i5-1334U processor operating at 2.30 GHz with 32 GB DDR4 RAM (Table 3).

DISCUSSION

Quantitative agreement between CPA and METAVIR fibrosis staging

The present study showed close agreement between manual and ML-based CPA measurements, although

the degree of agreement varied by fibrosis stage and hepatic region. No significant differences between manual segmentation and the ML models were observed at F0 and F1 in the perisinusoidal, pericentral, and periportal regions. At F2, agreement with manual segmentation varied among the models and regions, whereas at the more advanced F3 and F4 stages, Ilastik and Fiber-ML generally produced CPA estimates closer to manual segmentation, particularly in the pericentral and periportal regions. Importantly, CPA measurements obtained by manual segmentation and all three ML models showed strong positive correlations with METAVIR scores ($p = 0.97-0.99$), indicating that increasing histopathological fibrosis stage was accompanied by a corresponding increase in quantitatively measured collagen area.

Table 3: Training time and computational resource consumption of the ML models.

Segmentation method	Training time (ms) in 2048× 1536 px. images	CPU consumption (%)	RAM consumption (%)	Number of errors (50 software runs)
TWS	42,000	92.00	31.20	5 (10%)
Ilastik	7,700	24.10	18.70	2 (4%)
Fiber-ML	3,000	19.70	17.60	3 (6%)

ML = machine learning; TWS = Trainable Weka Segmentation; CPU = central processing unit; RAM = random-access memory; ms = milliseconds; px = pixels. Training time represents the duration required for model training, whereas CPU and RAM consumption represent computational-resource utilization during processing. The number of errors indicates software errors recorded across 50 runs for each model. Computational performance was evaluated under identical hardware and software conditions using 375 histological images, each measuring 2048 × 1536 pixels.

These findings are consistent with previous studies demonstrating the potential of ML-assisted approaches for quantitative assessment of hepatic fibrosis. The ML-based qFibrosis model showed a strong relationship with conventional fibrosis scoring ($R^2 = 0.867$) [8]. Similarly, a Python-based ML package, Liver-Quant, showed a strong positive correlation ($p = 0.92$) with pathological scores in Masson's trichrome-stained liver sections [23]. An artificial intelligence-based tool for nonalcoholic fatty liver disease also demonstrated correlations ranging from 0.60 to 0.86 between automated CPA measurements and pathologist-assigned fibrosis scores [24]. Automated CPA measurement can provide a continuous quantitative assessment of fibrosis rather than relying exclusively on categorical histopathological stages [25]. This approach may therefore complement conventional scoring by providing a numerical measure of collagen burden. Nevertheless, automated measurements may be influenced by staining variability, tissue-processing artifacts, and image characteristics, emphasizing the importance of appropriate image normalization, segmentation, and validation [25].

Regional variation in fibrosis quantification

The variation in CPA measurements among the perisinusoidal, pericentral, and periportal regions is consistent with the heterogeneous distribution of collagen within fibrotic liver tissue [25]. This regional heterogeneity is particularly relevant to automated image analysis because differences in collagen density, tissue architecture, staining intensity, and background structures can influence segmentation performance. In the present study, ML-based CPA measurements showed close agreement with manual measurements at early fibrosis stages (F0–F1), whereas differences among models became more apparent as fibrosis progressed. At advanced stages, Ilastik and Fiber-ML generally yielded measurements closer to manual segmentation than TWS in several regional comparisons.

These findings suggest that the performance of an automated segmentation model may depend not only on overall fibrosis severity but also on the anatomical distribution and morphological complexity of collagen deposition. Therefore, regional assessment of fibrosis may provide additional information beyond a single whole-section measurement and should be considered when ML-based approaches are developed for quantitative veterinary histopathology.

Segmentation performance of the ML models

The Dice similarity coefficient (DSC) demonstrated a high degree of overlap between ML-generated segmentation masks and expert-derived reference masks. Ilastik and Fiber-ML achieved DSC values of 0.90-0.91 in the perisinusoidal region and 0.89-0.90 in the pericentral region, whereas TWS achieved 0.87 and 0.84, respectively. In the periportal region, all three models showed high segmentation agreement, with DSC values ranging from 0.90 to 0.92. These findings indicate that all three approaches were capable of reproducing manual fibrosis segmentation, although performance varied according to the model and hepatic region.

Previous studies have also demonstrated the utility of Ilastik for tissue segmentation. Its pixel-based classification approach enables discrimination of tissue structures using image features related to texture, intensity, and edges [26]. Ilastik has also demonstrated high segmentation performance in previous image-

analysis applications [27]. TWS similarly provides supervised pixel classification but may require greater manual training and refinement to achieve consistent segmentation of morphologically complex structures [28]. Fiber-ML has been developed for quantitative analysis of fibrous tissue and has demonstrated the ability to perform pixel-based classification of histological fibrosis [18]. The values around 0.75–0.80 are commonly classified as "good agreement," reflecting substantial overlap with the ground truth. A DSC of 0.80–0.85 is typically considered "almost perfect" or "excellent" [21].

Differences among the models may also reflect their respective feature extraction and classification workflows. Fiber-ML was developed specifically for the analysis of fibrosis stains, including Sirius Red and Masson's trichrome, whereas Ilastik incorporates interactive pixel classification using multiple image features and iterative user-guided training [17, 18]. By contrast, differences in feature selection and classification behavior in TWS may increase susceptibility to staining artifacts or background structures under some image conditions [18]. These methodological differences may partly explain the regional variations in segmentation agreement observed in the present study.

Diagnostic performance

All three ML models demonstrated high diagnostic performance for identifying fibrous and nonfibrous tissue. In the perisinusoidal region, sensitivity ranged from 89.8% to 90.8%, while Ilastik and Fiber-ML showed higher specificity than TWS. In the pericentral region, Ilastik and Fiber-ML demonstrated higher sensitivity and specificity than TWS. In the periportal region, all three models performed comparably, with sensitivity ranging from 89.9% to 91.3% and specificity from 90.5% to 92.3%. Overall accuracy ranged from 84.7% to 91.8%, demonstrating that the three models could distinguish fibrous from nonfibrous tissue with relatively high accuracy.

The diagnostic performance observed in this study is broadly consistent with previous reports describing the effectiveness of random forest (RF)-based approaches in liver image analysis. An RF algorithm used to segment and classify malignant and benign liver tumors achieved 91.7% sensitivity and 92.1% accuracy when evaluated alongside XGBoost and multilayer perceptron models [29]. Another RF-based approach for predicting liver fibrosis achieved 96.8% accuracy, 95.7% sensitivity, and 97.8% specificity [30]. However, direct comparison with the present results should be interpreted cautiously because these studies involved different clinical applications, datasets, imaging modalities, and analytical endpoints. Published diagnostic-performance data directly comparing TWS, Ilastik, and Fiber-ML for bovine hepatic fibrosis remain limited.

Collectively, the present findings indicate that automated segmentation can provide reproducible quantitative measurements of collagen distribution across different hepatic regions. Such approaches could complement conventional histopathological assessment by reducing dependence on subjective visual estimation and providing continuous quantitative measurements of fibrosis. However, these models should be considered adjunctive analytical tools rather than replacements for histopathological interpretation until their performance is validated in larger, independent datasets.

Computational performance

Substantial differences in computational requirements were observed among the three models. TWS required 42,000 ms of training time and showed CPU and RAM utilization of 92.0% and 31.2%, respectively. In comparison, Ilastik required 7,700 ms, with CPU and RAM utilization of 24.1% and 18.7%, respectively, whereas Fiber-ML required 3,000 ms and showed CPU and RAM utilization of 19.7% and 17.6%, respectively. Across 50 software runs, errors or freezes occurred in 10% of TWS runs, 4% of Ilastik runs, and 6% of Fiber-ML runs.

The greater computational demand observed with TWS may be associated with generating and processing multiple pixel-level image features and filters during classifier training [14]. In contrast, the lower training time and resource utilization observed with Fiber-ML and Ilastik indicate that these platforms required fewer computational resources under the experimental conditions used in this study. Computational efficiency is relevant to the practical implementation of ML-assisted histopathology, particularly in laboratories where high-performance computing resources may not be readily available. Nevertheless, these findings are specific to the hardware configuration, image characteristics, software versions, classifier settings, and benchmarking procedure used in the present study and should not be generalized to other computational environments without further evaluation.

Study strengths and limitations

A strength of this study was the direct comparison of three open-source ML tools using the same histological image dataset and expert-derived reference segmentation. The animal-level separation of training/validation and

testing datasets reduced the risk of data leakage, while evaluation across perisinusoidal, pericentral, and periportal regions allowed assessment of model performance across different patterns of collagen distribution. The study also used multiple complementary performance measures, including CPA, DSC, sensitivity, specificity, accuracy, and computational resource utilization.

Several limitations should be considered when interpreting the findings. The study included a relatively small number of animals from a single geographic area and used samples obtained through a single slaughterhouse-based setting. The analysis was restricted to Masson's trichrome-stained sections, and the models were not externally validated using images generated at other laboratories, scanners, microscopes, staining protocols, or cattle populations. Differences in staining intensity and histological processing may influence automated segmentation performance. Furthermore, although the models demonstrated strong agreement with manual segmentation and METAVIR scores, the findings do not establish their clinical performance across the broader spectrum of naturally occurring bovine liver diseases.

Future perspectives

Future studies should evaluate these ML approaches using larger, multicenter datasets encompassing different cattle populations, hepatic diseases, staining batches, imaging systems, and laboratory conditions. External validation is particularly important for determining model generalizability and robustness. Comparisons with additional fibrosis stains and whole-slide imaging could further establish the reproducibility of automated CPA quantification. Deep-learning approaches, including convolutional neural networks, may also provide opportunities for automated feature extraction and fibrosis classification with less dependence on manually selected image features. Comparative evaluation of such approaches against the open-source ML models assessed in this study could help determine their relative advantages for veterinary digital pathology.

CONCLUSION

This study showed that TWS, Ilastik, and Fiber-ML can be used for automated segmentation and quantitative assessment of fibrosis in Masson's trichrome-stained bovine liver sections. CPA measurements generated by the three ML models showed strong positive correlations with METAVIR fibrosis scores ($p = 0.97-0.99$), supporting their ability to reflect increasing collagen deposition across fibrosis stages. Segmentation performance varied according to the hepatic region and model, with Ilastik and Fiber-ML generally showing greater agreement with manual segmentation than TWS, particularly in the perisinusoidal and pericentral regions. Overall diagnostic accuracy ranged from 84.7% to 91.8%, with the highest DSC values in the periportal region (0.90-0.92). Computational assessment also demonstrated substantial differences among the models, with Fiber-ML requiring the shortest training time and lowest CPU and RAM utilization, whereas Ilastik recorded the lowest frequency of software errors.

Automated CPA quantification provides an objective and continuous measure of collagen deposition that can complement conventional semiquantitative histopathological assessment of hepatic fibrosis. The ability of the evaluated ML models to quantify fibrosis across perisinusoidal, pericentral, and periportal regions may facilitate more standardized assessment of the heterogeneous distribution of collagen in bovine liver tissue. Their open-source availability also provides an opportunity for implementation in veterinary pathology laboratories where access to specialized commercial digital pathology platforms may be limited.

TWS, Ilastik, and Fiber-ML demonstrated the potential for automated quantitative assessment of bovine hepatic fibrosis, with strong correlations between ML-derived CPA measurements and METAVIR fibrosis scores. Ilastik and Fiber-ML generally showed closer segmentation agreement with expert-derived reference masks, while Fiber-ML required fewer computational resources and Ilastik showed the lowest error frequency under the conditions evaluated. These findings support the potential use of open-source ML-assisted image analysis as a complementary tool for objective quantification of bovine liver fibrosis. Larger-scale external validation is required before these approaches can be routinely incorporated into veterinary diagnostic pathology workflows.

DATA AVAILABILITY

The data supporting the findings are available within the manuscript

GENERATIVE AI DECLARATION

The authors affirm that the generative artificial intelligence (AI) systems have only been employed in the

linguistic enhancement of this manuscript. The AI systems were neither used in the data analysis nor in the interpretation of the results or in any other way of producing scientific content.

AUTHORS' CONTRIBUTIONS

OYA: Conceptualization, formal analysis, and investigation. SKA: Data curation, investigation, and validation. SKM: Formal analysis, methodology, validation, and visualization. GAS: Writing – original draft. OYA, SKA, SKM, and GAS: Methodology, project administration, and writing – review and editing. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors thank the staff of the Department of Anatomy, College of Veterinary Medicine, University of Mosul, Iraq, for their support and encouragement during this study. The authors received no funding for this study.

COMPETING INTERESTS

The authors declare that they have no competing interests.

PUBLISHER'S NOTE

Veterinary World remains neutral with regard to jurisdictional claims in the published institutional affiliations.

REFERENCES

1. Erdal R, Birsoy K, Unlu G. Amino acid metabolism in liver mitochondria: from homeostasis to disease. *Metabolites*. 2025;15(7):446.
2. Lee JH, Yang MH, Han G, Jung WH, Bae MA, Han JS, *et al*. Artificial intelligence based quantitative analysis of hepatic fibrosis in carbon tetrachloride-induced mouse model of metabolic dysfunction-associated steatohepatitis. *Toxicol Res*. 2025:1–16.
3. Du X, Hua R, He X, Hou W, Li S, Yang A, *et al*. *Echinococcus granulosus* ubiquitin-conjugating enzymes (E2D2 and E2N) promote the formation of liver fibrosis in TGFβ1-induced LX-2 cells. *Parasit Vectors*. 2024;17(1):190.
4. Gao C, Chen G, Jia H, Zhu H, Cai Y, Yang D, *et al*. Beyond hepatic stellate cell heterogeneity: resolving fibrosis, restoring regeneration. *Int J Mol Med*. 2026;14;57(3):62.
5. Chowdhury AB, Mehta KJ. Liver biopsy for assessment of chronic liver diseases: a synopsis. *Clin Exp Med*. 2023;23(2):273–285.
6. Yu Y, Wang J, Ng CW, Ma Y, Mo S, Fong EL, *et al*. Deep learning enables automated scoring of liver fibrosis stages. *Sci Rep*. 2018;8(1):16016.
7. Ramot Y, Deshpande A, Morello V, Michieli P, Shlomov T, Nyska A. Microscope-based automated quantification of liver fibrosis in mice using a deep learning algorithm. *Toxicol Pathol*. 2021;49(5):1126–1133.
8. Neuschwander-Tetri BA, Akbary K, Carpenter DH, Nouredin M, Alkhouri N. The emerging role of second harmonic generation/two photon excitation for precision digital analysis of liver fibrosis in MASH clinical trials. *J Hepatol*. 2025;83(3):790–799.
9. Cheng G, Dai M, Xiao T, Fu T, Han H, Wang Y, *et al*. Quantitative evaluation of liver fibrosis based on ultrasound radio frequency signals: an animal experimental study. *Comput Methods Programs Biomed*. 2021;199:105875.
10. Trivedi NK, Tiwari RG, Anand A, Gautam V, Witarasyah D, Misra A. Application of machine learning for diagnosis of liver cancer. In: 2022 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS). IEEE; 2022. p. 1–5.
11. Pei X, Deng Q, Liu Z, Yan X, Sun W. Machine learning algorithms for predicting fatty liver disease. *Ann Nutr Metab*. 2021;77(1):38–45.
12. Carson FL, Cappellano C. Chapter 2: Processing. In: Carson FL, Cappellano C. *Histotechnology: a self-instructional text*. 5th ed. Chicago: American Society for Clinical Pathology Press; 2020. p. 149–179.
13. Pirasteh A, Periyasamy S, Meudt J, Liu Y, Lee LM, Schachtschneider KM, *et al*. Staging liver fibrosis by fibroblast activation protein inhibitor PET in a human-sized swine model. *J Nucl Med*. 2022;63(12):1956–61.
14. Tosta TA, de Faria PR, Neves LA, do Nascimento MZ. Computational normalization of H&E-stained histological

images: progress, challenges and future potential. *Artif Intell Med.* 2019;95:118–32.

15. Kim S, Lee J, Chung JH. Histological assessment and interobserver agreement in major pathologic response for non–small cell lung cancer with neoadjuvant therapy. *Cancer Res Treat.* 2025;57(2):401–11.
16. Yamanishi C, Parigoris E, Takayama S. Kinetic analysis of label-free microscale collagen gel contraction using machine learning-aided image analysis. *Front Bioeng Biotechnol.* 2020;8:582602.
17. Berg S, Kutra D, Kroeger T, Straehle CN, Kausler BX, Haubold C, *et al.* Ilastik: interactive machine learning for (bio) image analysis. *Nat Methods.* 2019;16(12):1226–32.
18. Facchin C, Certain A, Yoganathan T, Delacroix C, Garcia AA, Gaillard F, *et al.* FIBER-ML, an open-source supervised machine learning tool for quantification of fibrosis in tissue sections. *Am J Pathol.* 2022;192(5):783–793.
19. Papatheodoridi M, Hall AR, Rodríguez-Perálvarez M, Pieri G, Germani G, Gale J, *et al.* Histological sub-classification of cirrhosis using collagen proportionate area in patients with chronic hepatitis C. *Liver Int.* 2021;41(7):1608–1613.
20. Belmans F, Seldeslachts L, Vanhoffelen E, Tielemans B, Vos W, Maes F, *et al.* Automated quantification of lung pathology on micro-CT in diverse disease models using deep learning. *EBioMedicine.* 2025;1(119):1–12.
21. Ostmeier S, Axelrod B, Isensee F, Bertels J, Mlynash M, Christensen S, *et al.* USE-Evaluator: performance metrics for medical image segmentation models supervised by uncertain, small or empty reference annotations in neuroimaging. *Med Image Anal.* 2023;90:102927.
22. Yang HJ, Lee HS. Common statistical methods used in medical research. *Kosin Med J.* 2025;40(1):21–30.
23. Farzi M, McGenity C, Cratchley A, Leplat L, Bankhead P, Wright A, *et al.* Liver-Quant: feature-based image analysis toolkit for automatic quantification of metabolic dysfunction-associated steatotic liver disease. *Comput Biol Med.* 2025;190:110049.
24. Gawrieh S, Sethunath D, Cummings OW, Kleiner DE, Vuppalanchi R, Chalasani N, *et al.* Automated quantification and architectural pattern detection of hepatic fibrosis in NAFLD. *Ann Diagn Pathol.* 2020;47:151518.
25. Soon G, Wee A. Updates in the quantitative assessment of liver fibrosis for nonalcoholic fatty liver disease: histological perspective. *Clin Mol Hepatol.* 2020;27(1):44.
26. Pavone AM, Giannone AG, Cabibi D, D’Aprile S, Denaro S, Salvaggio G, *et al.* Digital pathology: a comprehensive review of open-source histological segmentation software. *BioMedInformatics.* 2024;4(1):173–196.
27. Yates SC, Groeneboom NE, Coello C, Lichtenthaler SF, Kuhn PH, Demuth HU, *et al.* QUINT: workflow for quantification and spatial analysis of features in histological images from rodent brain. *Front Neuroinform.* 2019;13:75.
28. Jentzsch M, Albiez V, Kardamakis TC, Speck T. Analysis of the peel structure of different *Citrus* spp. via light microscopy, SEM and μ CT with manual and automatic segmentation. *Soft Matter.* 2024;20(12):2804–2811.
29. Boppana SH, Thota M, Maddineni G, Komati SK, Dang AK, Aakash F, *et al.* S2839 multimodal machine learning integration of clinical biomarkers and CT radiomics for precise liver lesion classification. *Am J Gastroenterol.* 2025;120(10S2):S610.
30. Hou M, Zhu Y, Zhou H, Zhou S, Zhang J, Zhang Y, *et al.* Innovative machine learning approach in liver fibrosis and disease severity evaluation in MAFLD patients using MRI fat content analysis. *Clin Exp Med.* 2025;25(1):275.
