

<https://doi.org/10.1038/s41746-025-01870-1>

Automating liver biopsy segmentation with a robust, open-source tool for pathology research: the HOTSPoT model



Giorgio Cazzaniga¹, Vincenzo L'Imperio¹, Emanuela Bonoldi², María-Carlota Londoño³, João Madaleno⁴, Augusta Cipriano⁵, Tom J. G. Gevers⁶, Iryna V. Samarska⁷, Özgür M. Koc^{6,8}, Alejandra Villamil⁹, Maria Florencia Sanchez⁹, Vincenza Calvaruso¹⁰, Alberto Quattrocchi¹¹, Daniela Cabibi¹¹, Bastian Engel¹², Federica Malinverno¹³, Elisa Merelli¹⁴, Laura Cristofori¹³, Marco Carbone^{14,15}, Fabio Pagni¹, Pietro Invernizzi^{13,14} & Alessio Gerussi^{13,14} ✉

Artificial intelligence applications in liver pathology remain limited, with existing tools either narrowly focused or lacking external validation. This study introduces HOTSPoT, an open-source, validated transformer-based model for automated segmentation of portal tracts in H&E-stained liver biopsy whole slide images. A multi-institutional dataset of 223 cases was used, with annotations by expert hepatopathologists. HOTSPoT achieved high performance with mean Dice scores of 0.92 (train/val) and 0.91 (test), and mean IoUs of 0.86, 0.85, and 0.84, respectively, showing minimal domain shift. Automated portal tract quantification showed strong concordance with manual assessments (κ up to 0.90), and portal area correlated with fibrosis stage ($r = 0.87$, $p < 0.001$). The model is available as a TorchScript file with a modified WSInfer library, enabling efficient WSI-level inference and integration with QuPath for advanced pathology analysis.

The transition from traditional microscopy to Whole Slide Imaging (WSI) has revolutionized pathology, enabling the quantitative study of pathological tissues through advanced computer vision techniques applied to high-resolution digital slides^{1,2}. This technological leap has laid the foundation for integrating artificial intelligence (AI) into liver pathology, with the promise of standardizing tissue analyses and enhancing research through precise assessment of morphological elements and improved integration with ancillary methods such as *in situ* -omics³. Despite the increasing application of AI in liver pathology (e.g., fibrosis and Metabolic dysfunction-Associated Steatotic Liver Disease -MASLD), rare inflammatory diseases with significant clinical burden remain largely underexplored^{4,5}. Histological evaluation plays a crucial role in the diagnostic workflow of these conditions and requires detailed evaluation of inflammation, considering its localization and spread across specific compartments⁶.

Although some commercial applications exist for dividing liver tissue into its main components, their datasets and intended uses are often restricted to specific diseases and usually lack an external validation⁷. Moreover, there is a lack of reliable open-source options that would enable pathologists and researchers to perform automated zonation of liver tissue⁸. This capability could unlock various applications, not limited to studying the localization of inflammatory infiltrates, but also extending to tasks such as data coregistration across advanced methodologies through automated

annotation and targeted region selection for studying specific compartments, thereby saving computational resources⁹. As well as other promising computer vision techniques, such as weakly supervised classification methods, as demonstrated in our previous research, these strategies offer significant potential for advancing tissue analysis¹⁰.

This study aims to bridge this gap by developing an AI-based approach for automating the segmentation and semantic recognition of portal and lobular compartments in liver biopsies.

Results

Cases

A total of 223 cases were collected for training the model, encompassing a diverse range of conditions. The largest group consisted of cases classified as AIH (85 cases, 38%), followed by primary biliary cholangitis (PBC, 54 cases, 24%), normal or near normal liver parenchyma (25 cases, 11%), drug-induced liver injury (DILI) (16 cases, 7%), MASLD (15 cases, 7%), graft-versus-host disease (GVHD) (10 cases, 5%), primary sclerosing cholangitis (PSC, 10 cases, 5%), infectious disease (5 cases, 2%), and iron overload disorder (3 cases, 1%). The sex distribution within the study was nearly balanced, with females constituting 59.64% (132 cases) and males 40.36% (90 cases) of the total. The mean age of participants was 50.2 years, ranging from a minimum of 8 years to a maximum of 85 years (median 50). The test

A full list of affiliations appears at the end of the paper. ✉ e-mail: alessio.gerussi@unimib.it

set comprised 16 AIH cases and eight PBC cases from five institutions external to the training and validation datasets, along with five GVHD cases from one internal institution. After tile extraction, the dataset comprised 54,338 patch pairs at 1 $\mu\text{m}/\text{px}$ magnification. Following an 80/20 split, 43,470 patches were allocated for training and 10,868 for validation. At 2 $\mu\text{m}/\text{px}$ magnification, the dataset contained a total of 28,197 patches, which were similarly divided into 22,558 for training and 5639 for validation. The test set included 8789 patches at 1 $\mu\text{m}/\text{px}$ magnification and 4611 patches at 2 $\mu\text{m}/\text{px}$ magnification. In the training and validation sets, the dataset composition was ~17% background, 71% hepatic lobule regions, and 12% portal areas. In the test set, the corresponding proportions were ~28% background, 61% hepatic lobule regions, and 11% portal areas.

Model development

The model was trained over a total of 100 epochs, with early stopping triggered at epoch 49 based on the validation loss, preventing further unnecessary training and reducing computational costs (Fig. 1). The overall training time was 10.6 h, and the average time per epoch was 13.1 min. The resulting model, called “Hematoxylin & Eosin-based Open-access Tool for Segmentation of Portal Tracts” (HOTSPoT), achieved a mean Dice coefficient of 0.92, 0.92 and 0.91 and a mean IoU of 0.86, 0.85 and 0.84, respectively, on the training, validation and test set. Detailed performance metrics are provided in Table 1, which presents the mean and per-class Dice coefficient and IoU values. Segmentation performance across the test set is further elaborated in Table 2. The average inference time for WSI, considering the test set was 1.3 s, while full test set inference time is 38 s, indicating scalability to larger datasets. The average time for inference and deployment on QuPath was 20 s (range 5 s–1.2 m). Variations in time were primarily influenced by the preprocessing steps of the tissue finder, which analyzed the entire slide to handle cases with dark or noisy backgrounds effectively. The QuPath implementation allows the graphic visualization of the automatic segmentation process, as shown in Fig. 2.

Upon completion of the training, the model was exported as a fully traced and weight-optimized TorchScript file for efficient deployment and inference. HOTSPoT was implemented in Python using the PyTorch 2.3 framework. The source code together with the torchscript full traced model, is available in the GitHub repository at the following URL: <https://github.com/Gizmopath/HOTSPoT>.

A modified version of the existing WSInfer library, designed for the deployment of WSI-level models in TorchScript format, was implemented and used for slide-level inference¹¹. This library was built and used in previous projects and it is easily editable to fit any segmentation project¹². The inference is supported at the WSI level, with model predictions being converted into GeoJSON files, an open standard for representing geographical features, that can be imported into QuPath. Within the ReverseQuPath folder, the source code required to run inference at WSI level using a pixel-based model is provided at the following URL: <https://github.com/Vsc0/nutshell>.

HOTSPoT clinical deployment: biopsy adequacy assessment

The clinical validation of the model included a total of 35 liver biopsy cases. The clinical diagnoses comprised autoimmune hepatitis (AIH, $n = 4$), drug-induced liver injury (DILI, $n = 8$), metabolic dysfunction-associated steatohepatitis (MASH, $n = 6$), primary biliary cholangitis (PBC, $n = 2$), ductal plate malformations ($n = 3$), primary sclerosing cholangitis (PSC, $n = 3$), and vascular/sinusoidal diseases ($n = 6$). Among these, 26 patients presented with a Ludwig fibrosis score of less than 3 at the time of diagnosis; inference on whole-slide images yielded a mean of 16.96 ± 6.08 portal tracts per case, showing a very high level of agreement with both observers ($\kappa = 0.90$ with original diagnosis and $\kappa = 0.80$ with the second observer).

The mean total biopsy area across the 35 cases was $19.56 \pm 7.88 \text{ mm}^2$, while the mean manually assessed length was 1.94 cm ($r = 0.92$, $p < 0.001$). Portal area percentages quantified by HOTSPoT on the total tissue area (mean $0.13 \pm 0.14\%$) differed significantly across the five Ludwig fibrosis stages (0–4), as shown in Fig. 3, with a correlation with the histological

fibrosis of 0.87 ($p < 0.001$), suggesting that HOTSPoT quantification may serve as a useful surrogate for immediate “macroscopic” fibrosis assessment without special stains.

Discussion

The segmentation of portal areas in the liver has become an area of increasing interest in AI-based research, due to its potential to automate tasks that require complex manual annotations¹³. One of the key goals of this study was to develop a model capable of generalizing across a wide range of liver biopsy samples, obtained from different institutions and scanned with various devices¹⁴. HOTSPoT’s generalizability is demonstrated by its performance on test sets from multiple centers, which used distinct scanners, magnifications, and pre-analytical protocols, suggesting that its architecture is resilient also to geographical diversity of the dataset. The scalability of HOTSPoT can be partially attributed to its transformer-based architecture, which is known for its ability to handle large, complex datasets and retain contextual information across various regions of interest. Traditional convolutional neural networks (CNNs) often face limitations when processing high-resolution images or when capturing long-range dependencies across distant regions in an image¹⁵. Additionally, in this experiment, data augmentation techniques were conducted under human supervision to ensure that the generated images remained realistic and reflective of typical pathological variability, avoiding the introduction of artificial artifacts. By simulating the diverse conditions encountered in real-world pathology practice, these techniques enhanced the model’s ability to generalize across different tissue appearances, staining quality, and slide artifacts, improving its performance on unseen biopsy samples while reducing the risk of overfitting^{16,17}. HOTSPoT’s scalability is further demonstrated by its small size (approximately 15 MB) and its ability to perform inference almost instantaneously, even on large datasets¹⁸. Additionally, its seamless integration with QuPath enables general pathologists to incorporate advanced AI tools into their workflow, facilitating research and enhancing the potential for broader adoption and contribution to scientific practice¹⁹. In particular, the model empirically demonstrated superior performance on clean, thin, artifact-free sections, as reflected by differing metrics across the various validation centers. For example, Fig. 4a (portal tract) and b (centrilobular vein) shows a nearly perfect segmentation, whereas the presence of artifacts (such as pen markings, dust particles, and slide imperfections) led to misclassification of non-tissue regions as tissue (Fig. 4c). The test set, which includes multiple cases of both chronic and acute hepatitis, serves as a rigorous stress test for a portal-space recognition algorithm: disease-induced alterations can be profound, and centrilobular regions may closely resemble portal spaces. The outstanding performance metrics attest to the algorithm’s robustness, although it can occasionally “capture” small, highly inflamed lobular areas in error (Fig. 4d).

Several experiments already exist in the literature, with various technologies, algorithms, and datasets developed to address this challenge, providing valuable insights for comparison in recent studies. A study published in the Journal of Pathology Informatics in January 2022 focused on the automated detection of portal spaces and central veins in liver tissue images of rats, achieving an F1 score of 0.81 compared to expert manual annotations. However, when applied to steatotic liver tissue, the model’s performance dropped to an F1 score of 0.59, demonstrating the difficulty in generalizing the model to unseen pathological tissues during training²⁰. A study by Hanyi Yu et al. in 2022 proposed the segmentation of portal areas in human liver biopsies. The introduced model, named MUSA-UNet, is a variant of the traditional UNet architecture. The authors used a dataset of histological images from 53 transplant patients stained with H&E and Masson’s trichrome, achieving precision of 0.94, recall of 0.85, F1 score of 0.89, and accuracy of 0.89 on the validation set²¹. Hung-Wen Tsai et al. tackled the segmentation of periportal regions infiltrated by lymphocytes, an important aspect for diagnosing and classifying chronic hepatitis. Their advanced framework integrates automatic segmentation based on a variant of DeepLabv3, achieving IoU values of 0.905 for portal spaces and 0.744 for bile ducts. In the second step, lymphocyte infiltration was detected by

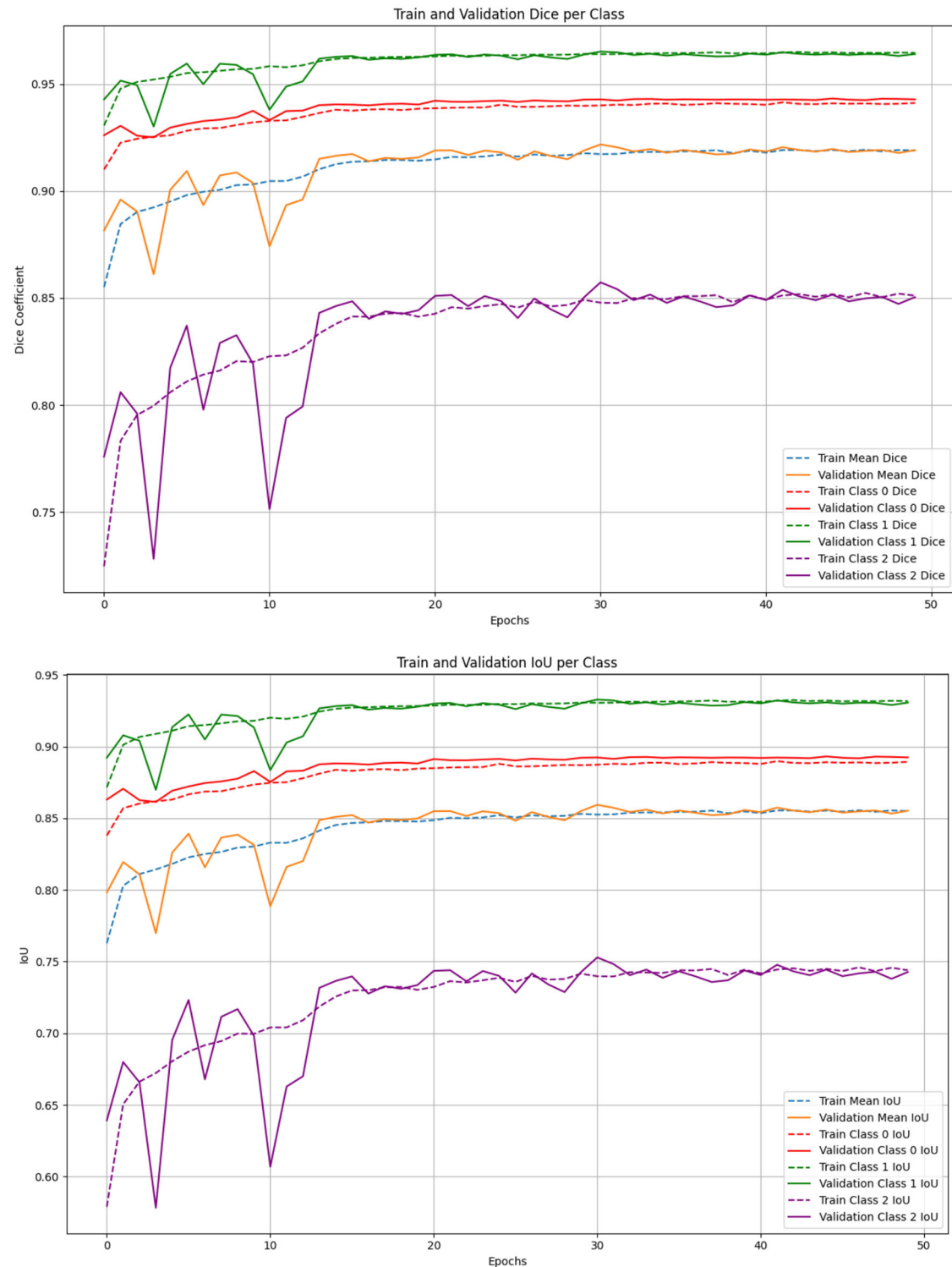


Fig. 1 | Training process of the HOTSPOT model. The training process illustrates the model's IoU and dice coefficient improvement over the epochs, with steady improvement in both training and validation performance. IoU Intersection over Union.

analyzing the protrusions at portal boundaries, identifying suspicious infiltration areas²². In 2024, Caner Ercan et al. introduced AI(H), an AI-based model using convolutional neural networks (CNNs) to analyze liver biopsies for autoimmune hepatitis (AIH). The training dataset consisted of 123 pre-treatment liver biopsies from confirmed AIH cases, analyzed on a proprietary platform. The AI(H) model achieved 99.4% accuracy in tissue detection, 88.0% for hepatic microanatomy, 83.9% for necroinflammation features, 81.7% for bile duct damage, and 79.2% for portal inflammation²³.

Despite significant progress, these studies have certain limitations. Most did not release open-source models, which would have facilitated broader adoption within the scientific community. Additionally, many datasets were derived from a single clinical center, potentially limiting the models' applicability and generalizability²⁴. External validation remains a critical step in ensuring the robustness and reliability of AI models for histological analysis. While high accuracy can often be achieved on training datasets, evaluating model

Table 1 | Model performance on training, validation and test sets overall and divided by compartment of interest

	Training set (2 μm/ px)	Validation set (2 μm/px)	Test set (2 μm/ px)	Training set (1 μm/ px)	Validation set (1 μm/px)	Test set (1 μm/ px)
Mean						
Dice	0.92	0.91	0.90	0.92	0.92	0.91
IoU	0.86	0.85	0.84	0.86	0.85	0.84
Background						
Dice	0.98	0.98	0.96	0.94	0.94	0.93
IoU	0.97	0.96	0.93	0.89	0.89	0.87
Lobule						
Dice	0.97	0.95	0.94	0.96	0.96	0.93
IoU	0.91	0.90	0.91	0.93	0.93	0.91
Portal areas						
Dice	0.83	0.81	0.79	0.85	0.84	0.83
IoU	0.71	0.69	0.67	0.74	0.74	0.73

IoU Intersection over Union

Table 2 | Summary of test set performance metrics across different institutions and diseases

	Tiles	Disease	Inference Time	Inf/Vis Time QuPath	Mean Dice	Mean IoU
Barcelona	1186	AIH (5)	7 s	4.4 m	0.92	0.86
Buenos Aires	950	AIH (5)	5 s	0.6 m	0.91	0.85
Coimbra	1187	AIH (5)	6 s	0.5 m	0.93	0.88
Monza	1744	GVHD (5)	9 s	1.2 m	0.92	0.84
Maastricht	1107	AIH (2) PBC (3)	6 s	1.7 m	0.91	0.84
Palermo	1605	PBC (5)	5 s	1.4 m	0.89	0.82
Full test set	8789	AIH(16) PBC(8) GVHD(5)	38 s	9.8 m	0.91	0.84

The table includes the number of tiles processed, disease types, inference times, inference + visualization times in QuPath, mean dice coefficient, and mean IoU. The number of cases for each disease is indicated in parentheses.
AIH Autoimmune Hepatitis, GVHD Graft-Versus-Host-Disease, IoU Intersection over Union, PBC Primary Biliary Cholangitis

performance on diverse, independent data from different institutions is essential to enhance generalizability and reduce the risk of overfitting^{25,26}. Furthermore, the development and adoption of open-source solutions in pathology is increasingly important for advancing research and automated diagnostics. Open access to software and algorithms promotes collaboration, reduces barriers associated with proprietary systems, and fosters innovation and transparency within the field²⁷.

As a first immediately clinically implementable application, the high concordance metrics achieved demonstrate that our model can both precisely quantify two-dimensional tissue area (mm²) and automatically count the number of portal tracts—complementing traditional manual adequacy assessments with a more comprehensive appraisal of specimen adequacy²⁸. Furthermore, by producing a fully segmented biopsy -delineating portal tracts, interface (one hepatocyte diameter from the tract), and lobular regions- the tool lays the groundwork for future pathology workflows in which segmented images could be used to guide more precise measurements of inflammatory infiltrates, fibrosis staging, and other morphometric features, ultimately supporting more accurate and reproducible diagnoses.

However, HOTSPoT holds significant potential for advancing liver pathology in several key applications (Fig. 5). It can extract morphometric features and localize cells or clusters within liver tissue, enabling a multi-dimensional analysis that includes morphometric data, being particularly promising for inflammatory and autoimmune conditions²⁹. Additionally, it optimizes computational resources by automating liver region segmentation, enabling targeted analysis of specific compartments, such as lobules (e.g., steatosis quantification) or portal areas (e.g., portal tract study), in line with an ecologically sustainable use of AI in pathology^{30,31}. HOTSPoT’s advanced segmentation capabilities support the integration of spatial-omics data, such as proteomics or transcriptomics, with histological tissue context. By providing precise spatial annotations, it facilitates mapping molecular markers to specific liver regions, offering valuable insights into disease mechanisms and responses to treatment³².

Foundation models, such as large pretrained Vision Transformers (ViT) or models like Swin Transformer, are increasingly being explored in medical imaging. These models are typically pretrained on massive datasets like ImageNet and then fine-tuned for specific tasks. While foundation models show great promise in areas like classification, their applicability in pathology is still somewhat limited by their focus on neoplastic conditions^{33,34}. Pathological studies, particularly in liver diseases, require specialized models that can handle the complexities of tissue segmentation, especially in non-neoplastic contexts, where inflammatory and autoimmune conditions are not sufficiently represented in the large, pretrained datasets typically used in foundation models. This challenge calls for the development of models that can be fine-tuned for non-neoplastic liver diseases, as well as the creation of specialized datasets for these diseases.

Fig. 2 | Graphic visualization of the automatic segmentation process. Inference via the constructed model and tile level segmentation were automatically imported as geoJSON objects into QuPath for image visualization and analysis. On the left, biopsy of a patient with cirrhotic liver in AIH. On the right, biopsy of a non-cirrhotic patient with quiescent PBC.

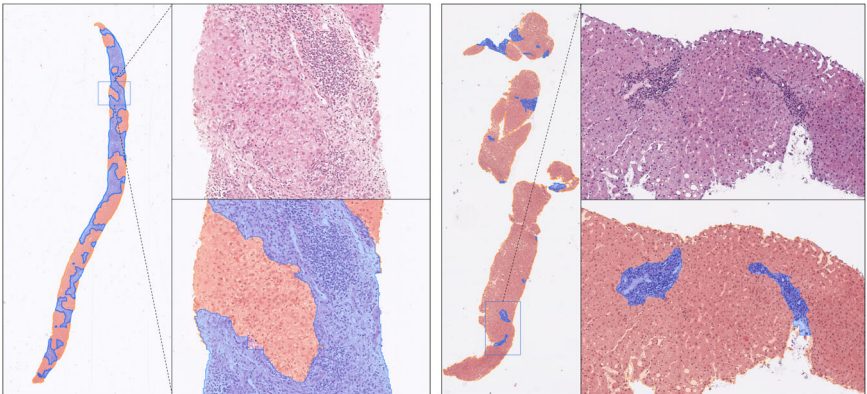


Fig. 3 | Portal area quantification by HOTSPoT correlates with Ludwig fibrosis stages. Box plots showing the percentage of biopsy area occupied by the portal region relative to the total tissue area, stratified by histological fibrosis class. A strong positive correlation between portal area percentage and fibrosis class was observed (Pearson's $r = 0.87$, $p < 0.001$).

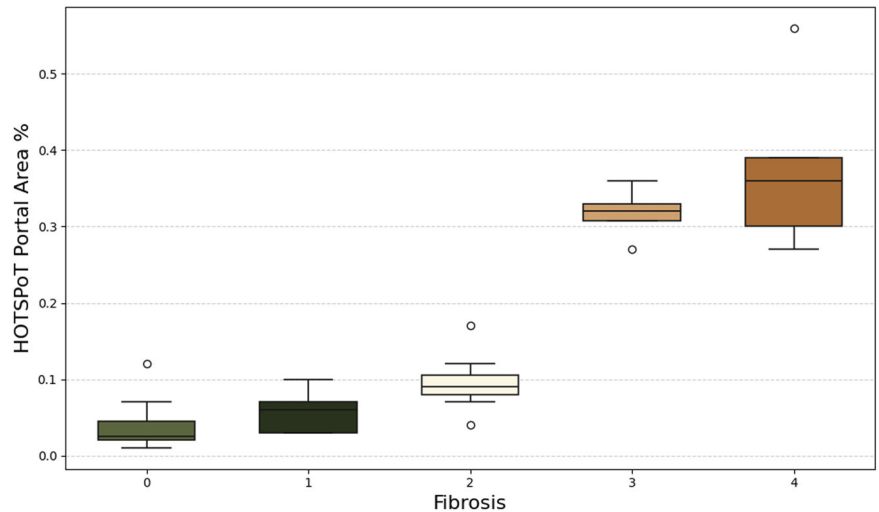
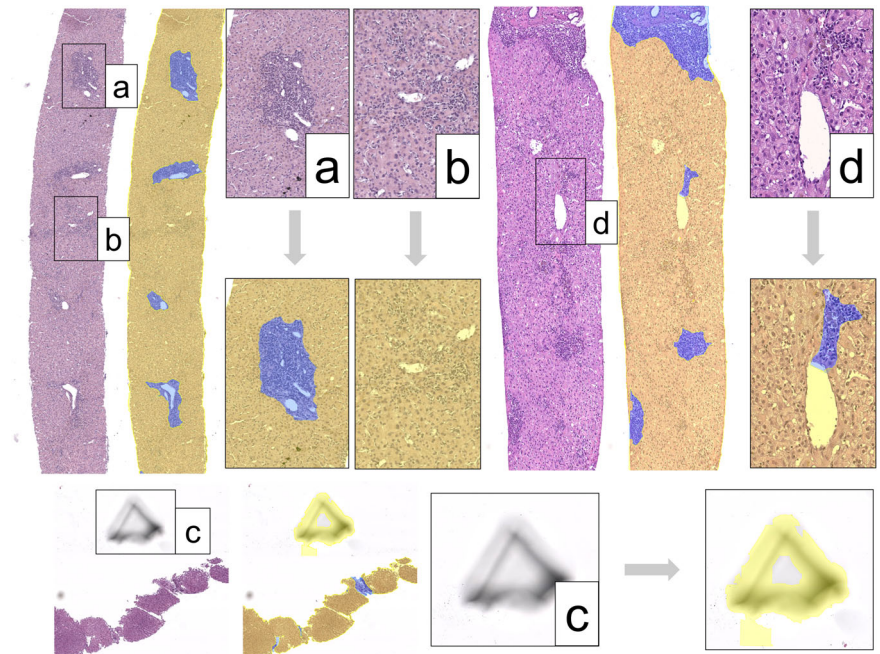


Fig. 4 | Variability of the algorithm across different settings. Segmentation performance of the model under varying conditions. Panel (a) shows accurate identification of a portal tract in a clean, artifact-free section, while (b) illustrates successful segmentation of a centrilobular vein. In contrast, (c) highlights misclassification caused by artifacts such as pen markings and debris. Panel (d) demonstrates an instance where the algorithm erroneously captures inflamed lobular areas resembling portal spaces in the context of hepatitis.



While the annotation of portal tracts is considered a low-variability task due to their clear histological definition, the use of a single expert may inherently limit the assessment of inter-observer consistency. Although no evidence of significant variability is reported in the literature for this specific feature, future studies including multiple annotators could further confirm the robustness and generalizability of the annotations. Variations in scanner types, magnifications, and staining protocols were considered during model development; however, potential biases introduced by site-specific differences in sample preparation and digitization may still impact performance in real-world clinical settings.

This study presents a promising approach for automating the segmentation and semantic recognition of liver tissue in WSI using a transformer-based model. HOTSPoT demonstrated high performance across a variety of scanning protocols and pathological conditions, showcasing its potential for standardizing liver tissue analysis and supporting research in rare liver diseases. Furthermore, the trained model is available for free, enabling local use and fostering broader accessibility within the research community.

Methods

Cases

Hematoxylin and eosin (H&E) stained liver biopsies WSIs were retrieved from three different institutions (Fig. 6): IRCCS Fondazione San Gerardo dei Tintori (Monza, Italy); Institute of Pathology, Hannover Medical School (Hannover, Germany) and ASST Grande Ospedale Metropolitano Niguarda (Milan, Italy), and anonymized. These slides were digitized using four different scanners at different magnifications. Demographic (sex and age) and key pathological data (final diagnosis, Ishak fibrosis stage) were collected for each case. Additionally, 20 cases from four external institutions: Hospital Clínic de Barcelona (Barcelona, Spain), Hospital Italiano de Buenos Aires (Buenos Aires, Argentina), Hospitais da Universidade de Coimbra (Coimbra, Portugal), Maastricht University Medical Centre (Maastricht, Netherlands) and Policlinico di Palermo (Palermo, Italy) and 5 cases from one institution internal to the training/validation set were used as a test set, each scanned with a different institution-device combination from those used in the training/validation set. Technical

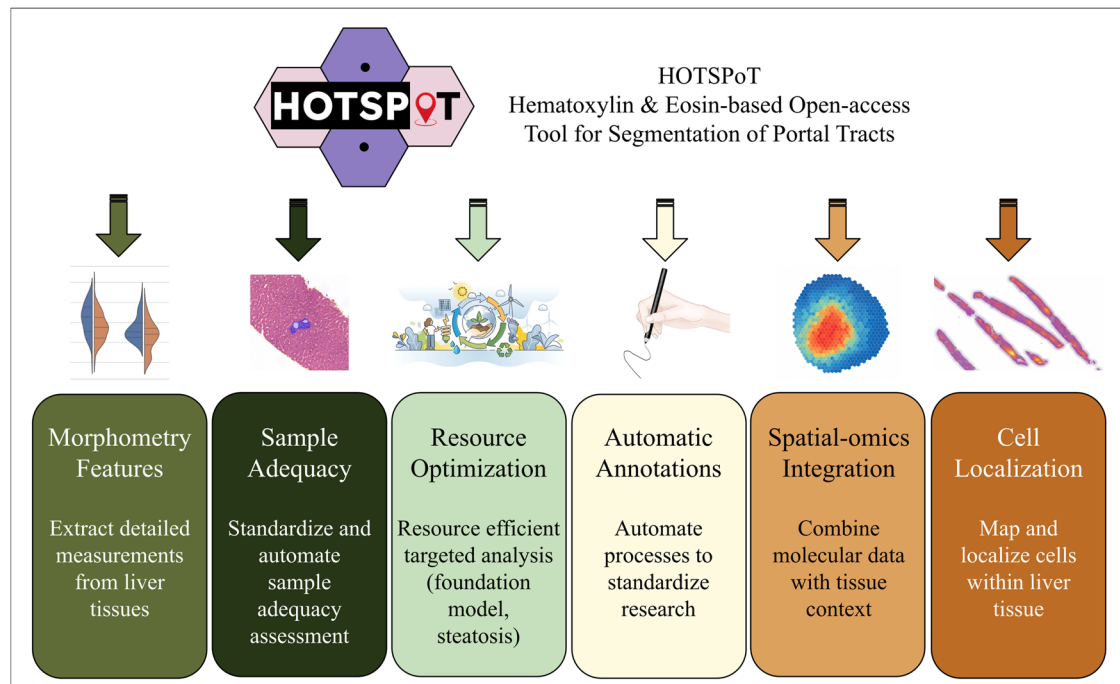


Fig. 5 | Clinical and research applications of the HOTSPoT liver biopsy segmentation model. HOTSPoT enables advanced liver pathology analysis by extracting morphometric features, localizing cell clusters, and supporting targeted

compartment analysis. Its segmentation capabilities facilitate efficient resource use and integration of spatial-omics data with histological context, offering insights into disease mechanisms and treatment responses.

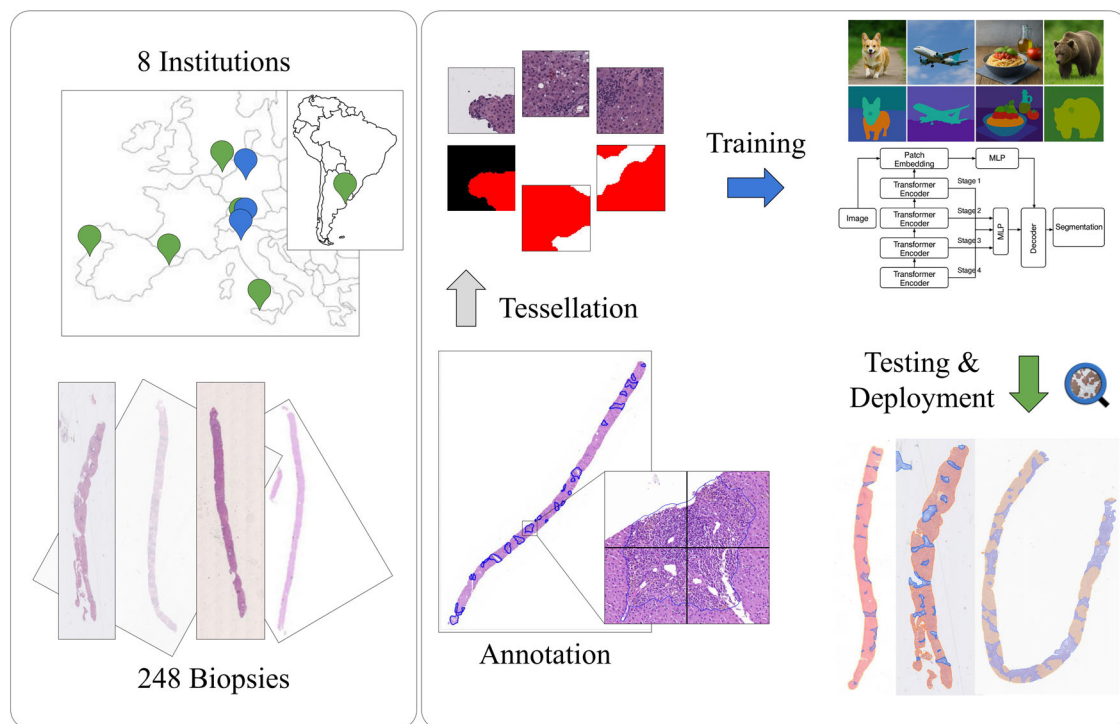


Fig. 6 | Study design. H&E-stained liver biopsy WSIs were collected from multiple expert liver pathology centers across Europe and South America, digitized using diverse scanners and magnifications. The dataset includes training/validation cases

from three institutions and an independent test set from five external centers, ensuring variability and robustness across acquisition settings.

and scanning characteristics of the whole dataset are reported in Table 3. All participating institutions are recognized as expert referral centers for liver diseases and (in part) members of the European Reference Network for rare liver diseases.

Approval was obtained from the local ethics committees of each institution. The coordinating center is University of Milano-Bicocca which coordinates the PBC Database (NCT05151809) and AIH Database (NCT06078098), approved by CET Lombardia 3.

Table 3 | Summary of the sample size and technical characteristics of H&E slides from the centers involved in the study

Institution	Scanner	TV/Test	N°	Format	Magnific.	µm/px
Monza	Aperio	TV	71	.svs	20X	0.4940
Monza	Hamamatsu	TV	30	.ndpi	20X	0.4416
Monza	3DHISTECH	TV	64	.mrxs	67X	0.1725
Hannover	Aperio	TV	64	.svs	40X	0.2513
Milan	Epredia	TV	43	.mrxs	20X	0.2425
Barcelona	Hamamatsu	Test	5	.ndpi	40X	0.2200
Buenos Aires	Micro Visioneer	Test	4	.tif	10X	0.5860
Coimbra	Roche	Test	5	.tiff	40X	0.2500
Maastricht	3DHISTECH	Test	5	.mrxs	20X	0.2425
Monza	Leica	Test	5	.svs	40X	0.2621
Palermo	Philips	Test	5	.tiff	40X	0.2500

TV Train/Validation Set, Test Test Set.

Annotations

Tissue on the biopsy was identified using QuPath’s custom tissue finder with a thresholding approach³⁵. Portal tracts were annotated within the same software based on a standard definition (presence of at least two out of three portal triad structures, including at least one portal artery or interlobular duct)³⁶ and semantically delineated from the lobular compartment by a hepatopathologist (GC). As the resolution of each WSI is in the order of tens of thousands of pixels and the region of interest (ROI) size varies among the different images, a tiling procedure was applied on the ROIs to obtain smaller images—of a fixed size—suitable as input for the transformer model. Tiles of 256 × 256 pixels (no overlap) were extracted from tissue areas at a magnification of 1 µm/px (approximately 10× magnification). These tiles (original images) were extracted along with their corresponding masks, in which distinct colors were used to label background, liver lobule tissue, and portal areas. The same extraction procedure was also performed at a magnification of 2 µm/px. The percentage of pixels belonging to each of the three classes (see below) was computed in the training/validation and test sets to provide an estimate of class distribution across the dataset.

Model development

This experiment focused on training a deep learning model for multi-class semantic segmentation using the SegFormer architecture, fine-tuned for segmenting images into three distinct classes: background, liver lobule, and portal areas. The model utilized a pretrained SegFormer-b0 structure (nvidia/segformer-b2-finetuned-ade-512-512), which was adapted specifically for this task using images and their corresponding masks³⁷. Data augmentation, including rotations, flips, elastic transformations, grid distortions, and color jittering, was applied to images and masks, with strength tailored by a pathologist to ensure realism and reflect laboratory variability without introducing artifacts. Early stopping based on validation loss prevented overfitting, and performance was evaluated at each epoch using metrics such as intersection over union (IoU) and Dice coefficient both overall and per class. Dice coefficient measures the similarity between the predicted and ground truth segmentations by evaluating how much they overlap relative to their total size at patch level, while IoU evaluates the ratio between the overlapping area (intersection) and the total combined area (union) of the prediction and the ground truth. The inference at the WSI level was carried out using a modified version of the WSInfer library, specifically adapted for deploying WSI-level models in TorchScript format. This library is fully customizable to accommodate any segmentation workflow. No overlapping between tiles was used during tile-based inference; instead, model predictions are output as GeoJSON files, an open standard for representing geographic features, which can then be seamlessly imported into QuPath. Training and inference were performed on a

machine equipped with 16 GB of RAM and an NVIDIA GeForce GTX 1070 GPU with Max-Q Design.

HOTSPoT clinical deployment: biopsy adequacy assessment

The resulting model (HOTSPoT), was applied to evaluate the adequacy of 35 consecutive cases, including 25 internal cases from the IRCCS Fondazione San Gerardo dei Tintori (Monza) and 10 consultation cases referred to the same unit, without restriction by underlying disease. WSI-level inference was performed on each case to identify portal regions, which were automatically labeled and counted, providing a continuous estimate of the total number of portal tracts where feasible, specifically in cases with Ludwig stage <3. For all 35 cases, the total biopsy area on the slide (in mm²) was extracted, as well as the cumulative portal tract area and its proportion relative to the entire biopsy. The same portal-tract count was also extracted from the original pathology report and was independently assessed by a second observer (AQ), allowing comparison of both interobserver agreement and the performance of the model (HOTSPoT). Biopsy length (in cm) was manually measured for each case. Agreement between the model and human observers in counting continuous portal tracts was assessed using Weighted Cohen’s Kappa. To evaluate the reliability of HOTSPoT as a surrogate for immediate macroscopic assessment of histologic fibrosis, Pearson correlation coefficients were calculated between HOTSPoT’s biopsy area quantification and manual length measurements, as well as between HOTSPoT’s portal area percentage and histological fibrosis scores.

Data availability

The liver biopsy whole slide images used in this study are not currently publicly available due to data sharing restrictions. All underlying data, including the relevant spreadsheets necessary to interpret and verify the findings of this study, are available in the following GitHub repository: <https://github.com/Gizmopath/HOTSPoT>.

Code availability

The source code together with the torchscript full traced model, is available in the GitHub repository at the following URL: <https://github.com/Gizmopath/HOTSPoT>. A modified version of the existing WSInfer library, designed for the WSI-level deployment of models in TorchScript format is provided at the following URL: <https://github.com/Vsc0/nutshell>.

Received: 13 March 2025; Accepted: 7 July 2025;
Published online: 18 July 2025

References

1. Kiran, N. et al. Digital pathology: transforming diagnosis in the digital age. *Cureus* **15**, e44620 (2023).

2. Hölscher, D. L. et al. Next-generation morphometry for pathomics-data mining in histopathology. *Nat. Commun.* **14**, 470 (2023).
3. Allaume, P. et al. Artificial intelligence-based opportunities in liver pathology—a systematic review. *Diagnostics* **13**, 1799 (2023).
4. Ratziu, V. et al. Artificial intelligence-assisted digital pathology for non-alcoholic steatohepatitis: current status and future directions. *J. Hepatol.* **80**, 335–351 (2024).
5. Miyaaki, H. et al. Liver fibrosis analysis using digital pathology. *Med. Mol. Morphol.* **57**, 161–166 (2024).
6. Lohse, A. W. et al. Consensus recommendations for histological criteria of autoimmune hepatitis from the International AIH Pathology Group: results of a workshop on AIH histology hosted by the European Reference Network on Hepatological Diseases and the European Society of Pathology: results of a workshop on AIH histology hosted by the European Reference Network on Hepatological Diseases and the European Society of Pathology. *Liver Int.* **42**, 1058–1069 (2022).
7. Matthews, G. A., McGenity, C., Bansal, D. & Treanor, D. Public evidence on AI products for digital pathology. *Pathology*, **7**, 300 (2024).
8. Marée, R. Open practices and resources for collaborative digital pathology. *Front. Med.* **6**, 255 (2019).
9. Wang, S., Yang, D. M., Rong, R., Zhan, X. & Xiao, G. Pathology image analysis using segmentation deep learning algorithms. *Am. J. Pathol.* **189**, 1686–1698 (2019).
10. Gerussi, A. et al. Deep learning helps discriminate between autoimmune hepatitis and primary biliary cholangitis. *JHEP Rep.* **7**, 101198 (2025).
11. Kaczmarzyk, J. R. et al. Open and reusable deep learning for pathology with WSInfer and QuPath. *NPJ Precis. Oncol.* **8**, 9 (2024).
12. L'Imperio, V. et al. Machine learning streamlines the morphometric characterization and multiclass segmentation of nuclei in different follicular thyroid Lesions: Everything in a NUTSHELL. *Mod. Pathol.* **37**, 100608 (2024).
13. Cazzaniga, G. et al. Improving the annotation process in computational pathology: A pilot study with manual and semi-automated approaches on consumer and medical grade devices. *J. Imaging Inf. Med.* <https://doi.org/10.1007/s10278-024-01248-x> (2024).
14. Asadi-Aghbolaghi, M. et al. Learning generalizable AI models for multi-center histopathology image classification. *NPJ Precis. Oncol.* **8**, 151 (2024).
15. Bai, Y., Mei, J., Yuille, A. & Xie, C. Are Transformers more robust than CNNs? Preprint at <https://arxiv.org/abs/2111.05464> (2021).
16. Faryna, K., van der Laak, J. & Litjens, G. Automatic data augmentation to improve generalization of deep learning in H&E stained histopathology. *Comput. Biol. Med.* **170**, 108018 (2024).
17. Marini, N. et al. Data-driven color augmentation for H&E stained images in computational pathology. *J. Pathol. Inform.* **14**, 100183 (2023).
18. Reis-Filho, J. S. & Kather, J. N. Overcoming the challenges to implementation of artificial intelligence in pathology. *J. Natl. Cancer Inst.* **115**, 608–612 (2023).
19. McGenity, C. et al. Survey of liver pathologists to assess attitudes towards digital pathology and artificial intelligence. *J. Clin. Pathol.* **77**, 27–33 (2023).
20. Budelmann, D. et al. Automated detection of portal fields and central veins in whole-slide images of liver tissue. *J. Pathol. Inform.* **13**, 100001 (2022).
21. Yu, H. et al. Artificial intelligence based liver portal tract region identification and quantification with transplant biopsy whole-slide images. *Comput. Biol. Med.* **150**, 106089 (2022).
22. Tsai, H.-W. et al. Lymphocyte-infiltrated periportal region detection with structurally-refined deep portal segmentation and heterogeneous infiltration features. *IEEE Open J. Eng. Med. Biol.* **5**, 261–270 (2024).
23. Ercan, C. et al. A deep-learning-based model for assessment of autoimmune hepatitis from histology: AI(H). *Virchows Arch.* **485**, 1095–1105 (2024).
24. Stacke, K., Eilertsen, G., Unger, J. & Lundstrom, C. Measuring domain shift for deep learning in histopathology. *IEEE J. Biomed. Health Inform.* **25**, 325–336 (2021).
25. Cabitza, F. et al. The importance of being external. methodological insights for the external validation of machine learning models in medicine. *Comput. Methods Prog. Biomed.* **208**, 106288 (2021).
26. McGenity, C. et al. Artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy. *NPJ Digit. Med.* **7**, 114 (2024).
27. Fehr, J., Citro, B., Malpani, R., Lippert, C. & Madai, V. I. A trustworthy AI reality-check: the lack of transparency of artificial intelligence products in healthcare. *Front. Digit. Health* **6**, 1267290 (2024).
28. Tublin, M. E. et al. Prospective study of the impact of liver biopsy core size on specimen adequacy and procedural complications. *AJR Am. J. Roentgenol.* **210**, 1–6 (2017).
29. Socha, P. et al. Quantitative digital pathology enables automated and quantitative assessment of inflammatory activity in patients with autoimmune hepatitis. *J. Pathol. Inform.* **15**, 100372 (2024).
30. Choudhury, D. et al. Developing a low-cost, open-source, locally manufactured workstation and computational pipeline for automated histopathology evaluation using deep learning. *EBioMedicine* **107**, 105276 (2024).
31. Lan, Y.-C. et al. Ecologically sustainable benchmarking of AI models for histopathology. *NPJ Digit. Med.* **7**, 378 (2024).
32. Pang, J.-M. B. et al. Spatial transcriptomics and the anatomical pathologist: molecular meets morphology. *Histopathology* **84**, 577–586 (2024).
33. Wang, X. et al. A pathology foundation model for cancer diagnosis and prognosis prediction. *Nature* **634**, 970–978 (2024).
34. Vorontsov, E. et al. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nat. Med.* **30**, 2924–2935 (2024).
35. Bankhead, P. et al. QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).
36. Crawford, A. R., Lin, X. Z. & Crawford, J. M. The normal adult human liver biopsy: a quantitative reference standard. *Hepatology* **28**, 323–331 (1998).
37. Xie, E. et al. SegFormer: Simple and efficient design for semantic segmentation with Transformers. Preprint at <https://arxiv.org/abs/2105.15203> (2021).

Acknowledgements

FP includes the project in the Italian Ministry of the University MUR Dipartimenti di Eccellenza 2023-2027 (l. 232/2016, art. 1, commi 314 - 337). A.G., M.C. and P.I. declare that this study was supported by the Italian MUR PNRR PE06 “HEAL ITALIA – Health Extended ALliance for Innovative Therapies, Advanced Lab-research and Integrated Approaches of Precision Medicine” - Spoke 4 - Precision Diagnostics, Italian MUR PRIN 2022 PNRR P2022H7JYZ “Non-invasive biological and molecular characterization of autoimmune liver diseases and variant syndromes” and Bando Giovani “Early Career Award” FRRB - CUP H53C24000630002 “AIVAR - Artificial Intelligence-Enhanced Diagnosis of PBC-AIH Variant Syndrome. B.E. was supported by the PRACTIS – Clinician Scientist Program of Hannover Medical School, funded by the German Research Foundation (DFG, ME 3696/3).

Author contributions

G.C. and A.G. designed the study. G.C. and E.M. digitized the slides for IRCCS Fondazione San Gerardo dei Tintori, created the dataset, and structured the artificial intelligence project. E.B. provided the training set material from ASST Grande Ospedale Metropolitano Niguarda, Milan. B.E. provided the training set material from the Institute of Pathology, Hannover Medical School. M.C.L., J.M., A.C., T.G., I.S., O.M.K., A.V., M.F.S., V.C., and D.C. contributed test set materials from their respective institutions. A.Q. provided portal tracts count for pathological validation of the model. V.L. and F.P. offered specialized insights into the application of digital pathology. F.P., A.G., M.C., F.M., L.C. and P.I. provided administrative support and facilitated funding acquisition. All authors were involved in writing the paper and had final approval of the submitted and published versions.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Alessio Gerussi.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2025

¹Department of Medicine and Surgery, Pathology, Fondazione IRCCS San Gerardo dei Tintori, University of Milano-Bicocca, Monza, Italy. ²U.O.C. Anatomia Patologica e Citogenetica, Department of Hematology, Oncology and Molecular Medicine, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy. ³Liver Unit, Hospital Clínic Barcelona, Fundació de Recerca Clínic Barcelona-Institut d'Investigacions Biomèdiques August Pi i Sunyer (FRCB-IDIBAPS), Universitat de Barcelona, Centro de investigación biomédica en red Enfermedades Hepáticas y Digestivas (CIBEREHD), European Reference Network on Hepatological Diseases (ERN RARE-LIVER), Barcelona, Spain. ⁴Liver Disease Unit, Department of Internal Medicine, Unidade Local de Saúde de Coimbra - Hospitais da Universidade de Coimbra, Coimbra, Portugal; European Reference Network on Hepatological Diseases (ERN RARE-LIVER), Coimbra, Portugal. ⁵Pathology Department, Unidade Local de Saúde de Coimbra - Hospitais da Universidade de Coimbra, Coimbra, Portugal; European Reference Network on Hepatological Diseases (ERN RARE-LIVER), Coimbra, Portugal. ⁶Department of Gastroenterology and Hepatology, Maastricht University Medical Center, Maastricht, The Netherlands. ⁷Department of Pathology, Maastricht University Medical Center, Maastricht, The Netherlands. ⁸Department of Gastroenterology, Geriatrics, Internal and Intensive Care Medicine (Co-MIK), Zuyderland Medical Centre, Heerlen-Sittard-Geleen, The Netherlands. ⁹Hospital Italiano de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina. ¹⁰Section of Gastroenterology and Hepatology, PROMISE, University of Palermo, Palermo, Italy. ¹¹Dipartimento Di Promozione Della Salute, Materno Infantile, Medicina Interna e Specialistica Di Eccellenza (PROMISE), Palermo, Italy. ¹²Department of Gastroenterology, Hepatology, Infectious Diseases and Endocrinology, Hannover Medical School, Hannover, Germany; European Reference Network on Hepatological Diseases, Hannover, Germany. ¹³Division of Gastroenterology, Center for Autoimmune Liver Diseases, European Reference Network on Hepatological Diseases (ERN RARE-LIVER), Fondazione IRCCS San Gerardo dei Tintori, Monza, Italy. ¹⁴Department of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy. ¹⁵Liver Unit, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy.

✉ e-mail: alessio.gerussi@unimib.it