Histological Scoring of Nonalcoholic Fatty Liver Disease using Deep Learning

Agnete Overgaard[1], Christian Eschen[2], Denise Oró[1], Casper BG Salinas[1], Michael Feigh[1], Sanne S Veidal[1]
[1]Gubra, Hørsholm, Denmark
[2]Visiopharm A/S, Hørsholm, Denmark

Corresponding author: ssv@gubra.dk

INTRODUCTION

Liver biopsy is the gold standard to diagnose Non-Alcoholic Fatty Liver Disease (NAFLD) and the more progressive form non-alcoholic steatohepatitis (NASH), and biopsy evaluation is the primary outcome in clinical trials for NAFLD treatment. The diagnosis is confirmed using the histopathological NAFLD Activity Score (NAS), which grades the severity of steatosis, hepatocellular ballooning degeneration, and inflammation. In addition, fibrosis is evaluated using a separate validated staging system. Histopathological disease scoring systems are, however, subjective and prone to inter- and intra-observer variation. We, therefore, applied a deep learning image analysis strategy to obtain a more accurate and objective method for staging NAS and fibrosis in mouse models of NAFLD.

AIM

- To develop deep learning strategies for assessment of NAFLD activity score and fibrosis stage.
- To compare histopathological staging of NAFLD activity and fibrosis by pathologists with the automated algorithms created based on deep learning image analysis.

MATERIALS

Training data used for the segmentation models

Number of annotations	Portal triads	Central veins	Hepatocytes with lipids	Hepatocytes without lipids	Inflammatory nuclei
H&E	38	33	2607	3246	9131
PSR	158	204	-	-	-

For training the segmentation models, annotations representing varying staining intensity and different treatment-related morphological changes (from e.g. liraglutide, OCA, elafibranor), were used.

PSR: Picro Sirius Red. H&E: Hematoxylin & Eosin

Biopsies used for the fibrosis classifier

Strain	N	Diet	Weeks on diet
C57BL/6J	46	CHOW	28-41
C57BL/6J	589	AMLN	7-55
C57BL/6J	315	GAN	28-35
Lep$^{ob/ob}$	303	AMLN	20-24
Lep$^{ob/ob}$	9	CHOW	20-24

For the random forest classifier trained to predict fibrosis stage, we used biopsies from two mice strains with varying exposure to two different NASH diets or chow.

Lep$^{ob/ob}$: leptin deficient mouse strain . AMLN diet: NASH diet composed of 40% high fat diet (mostly Primex), 20% fructose and 2% cholesterol. GAN diet: NASH diet composed 40% high fat diet (mostly palm oil), 20% fructose and 2% cholesterol.

METHODS

Segmentation: Hepatocyte nuclei with lipid droplets, hepatocyte nuclei without lipid droplets, inflammatory nuclei and ballooning hepatocyte nuclei were annotated in H&E stained slides. Portal triads and central veins were annotated in both H&E and PSR stained slides.

Segmentation of the portal triad and central vein was carried out using the convolutional neural network *Deeplabv3+*. The rest of the objects were segmented using the convolutional neural network *U-Net*.

In PSR stained samples, fibrosis was detected in portal triads and the sinusoidal space using the linear Bayesian method, and using the polynomial local linear filter.

The image analysis, including training of the neural networks, was carried out using Visiopharm's Histopathology Image Analysis software.

Histopathological assessment: The Kleiner score was used to assess liver samples for NAFLD activity score and fibrosis stage. The score includes:

- Steatosis (0-3)
- Lobular inflammation (0-3)
- Ballooning degeneration (0-2)
- Fibrosis stage (F1-F4)

RESULTS

Fibrosis strategy

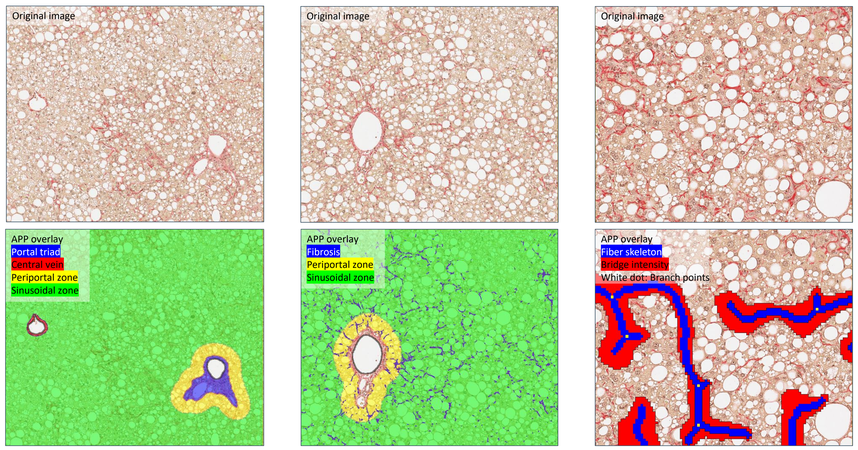
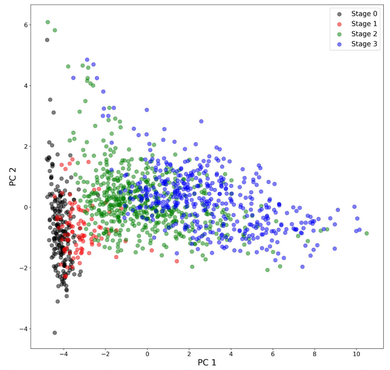


Figure 1 | *Left panel* | Portal triads and central veins are detected using deep learning, and post-processing creates a periportal zone of 100 µm. *Middle panel* | Fibrosis is detected using the linear Bayesian image analysis method in the periportal and sinusoidal zones, and different measures of fiber fragment size and shape, are used to predict bridging. *Right panel* | Bridging is also detected using the Threshold image analysis method based on a polynomial local linear filter feature.

XGBoost classifier predicts manual fibrosis score



		PREDICTED SCORE					
		0	1	2	3	4	Total
TRUE SCORE	0	36	3	0	0	0	39
	1	4	16	1	0	0	21
	2	0	3	86	18	0	107
	3	0	0	8	78	0	86
	4	0	0	0	0	0	0
Total		40	22	95	96	0	253

Cohen's Kappa	
Observed agreement	0.854
Kappa	0.785
Weighted Kappa	0.866

Figure 2 | *Left panel* | Principle Component (PC) analysis detection shows separation of fibrosis stages based on image analysis outputs. *Right panel* | The agreement and Cohen's Kappa values show good prediction of the manual scores. The classifier was trained on 1009 mouse liver biopsies (=80%; random samples) across all fibrosis stages and was validated on 253 samples

PSR image analysis output

Output	Description
Area fraction	Area of bridge, Periportal PSR, or sinusoidal PSR divided by total area
Connectivity	Connectivity is calculated from the size distribution of all detected fragments (PSR or bridge) after any optional post-processing steps.
Major Axis Length	The length of the major axis of an ellipse fitted to the PSR or bridge fragment.
Eccentricity	Measurement of the circularity degree of the object. The eccentricity will approximately equal 1 for at line and 0 for a circle.
Heatmap	Measurement of the intensity distribution of PSR high density regions and find hot spots.

NAFLD Activity Score strategy

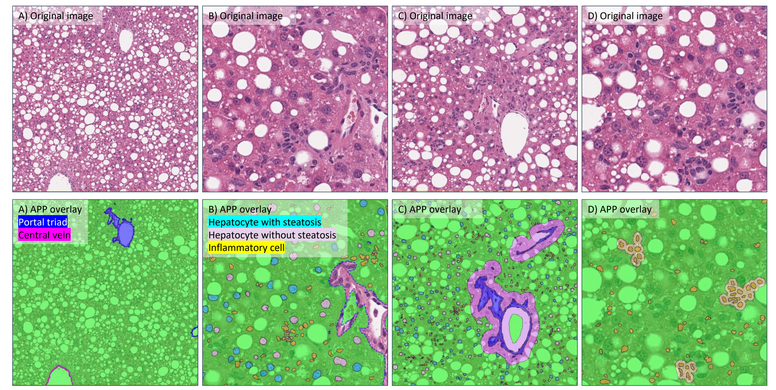


Figure 3 | *A*| Portal triads and central veins are detected using deep learning (10X). | *B*| Deep learning detects nuclei of hepatocytes with steatosis, hepatocytes without steatosis, and inflammatory cells (20X). | *C*| Post-processing excludes periportal inflammation. | *D*| Post-processing converts clusters of ≥4 inflammatory cells into foci. Scores will be calculated based on simple threshold.

Validation of H&E segmentation performance

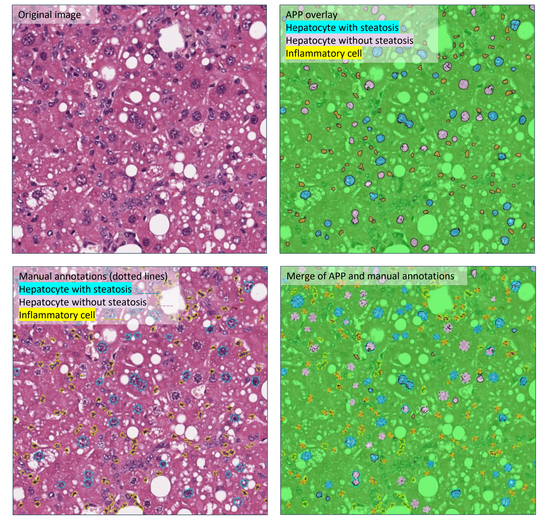


Figure 4 | Liver biopsies representing different staining intensities and disease stages was used for validation, and segmentation performance was calculated:

$$Dice = \frac{2 \times TP}{(TP + FP) + (TP + FN)}$$

	Dice coefficient
Hepatocytes without lipids	0.81
Hepatocytes with lipids	0.84
Inflammatory nuclei	0.81

CONCLUSION

- We demonstrate a deep-learning based approach to obtain NAFLD activity score and fibrosis stage in translational obese mouse models of diet-induced NASH.
- A deep-learning approach for pattern recognition allows rapid and reproducible quantification of histological NASH parameters. Future work will focus on optimizing the apps for clinical application.