Deep Learning-Based Image Analysis Algorithm for Classification and Quantification of Multiple Histopathological Lesions of the Rat Liver

○Taishi Shimazaki[1], Kyotaka Muta[1], Naohito Yamada[1], Yuzo Yasui[1], Ameya Deshpande[2], Anindya Hajra[2], Tijo Thomas[2] and Toshiyuki Shoda[1]

1: Toxicology Research Laboratories, Yokohama Research Center, Central Pharmaceutical Research Institute, Japan Tobacco Inc.

2: AIRA Matrix Private Limited

～Introduction～

- Artificial intelligence (AI)-based image analysis is increasingly being used for preclinical safety-assessment studies in the pharmaceutical industry. In this study, we present an AI-based solution for preclinical toxicology studies.
- We trained a set of algorithms to learn and quantify seven types of typical histopathological findings (including vacuolation, bile duct hyperplasia, and single-cell necrosis) in whole slide images (WSIs) of the livers of young Sprague Dawley (SD) rats by using a U-Net-based deep learning network[1-3]. The trained algorithms were validated using 255 liver WSIs to detect, classify, and quantify the seven findings in the liver.

～Materials & Methods～

- Generation of WSIs
  - 406 HE-stained glass slides of liver specimens from 8 week-old male SD rats, which were treated with several compounds in toxicity studies, were scanned using a NanoZoomer S360 (Hamamatsu Photonics K.K., Japan) at 20x magnification and converted into WSIs.

- Datasets (Table 1)
  - The total of 406 dataset was divided into a development dataset and a validation dataset. 92 WSIs from the development dataset were used to train the deep learning models for the seven lesions. The models were tested and progressively finetuned based on two rounds of feedback from pathologists on two different test datasets comprising 41 and 18 WSIs. (Table 1 summarizes the data distributions.)

Table 1: Number of WSIs used for training and validation of the algorithm

Findings	Training	1'st Test	2'nd Test	Validation
Vacuolation (spontaneous) of hepatocytes	8	4	18	205
Vacuolation (drug-induced) of hepatocytes	10	5	18	255
Bile duct hyperplasia	13	9	18	255
Single cell necrosis of hepatocytes	13	6	18	255
Microgranuloma	15	8	18	255
Extramedullary hematopoiesis	8	4	18	255
Hepatocellular hypertrophy	10	5	18	255
WSIs with no histopathological findings	15	-	-	
Total number of WSIs	92	41	18	255*

Required Number of WSI — Development Dataset (Training, 1'st Test, 2'nd Test); Validation Dataset (Validation)

*: Vacuolation (spontaneous) was validated with 205 WSIs.

- Workflow of algorithm development (Fig. 1)
  - Ground truth annotations for the seven lesions were generated by data-marking experts under the guidance of pathologists, who further verified the annotated data after marking.
  - The annotated WSIs from the training dataset were used to train the models based on a customized U-Net architecture[1]. The models were then tested and were gradually altered and improved to ensure that the algorithm and pathologists achieved agreement.
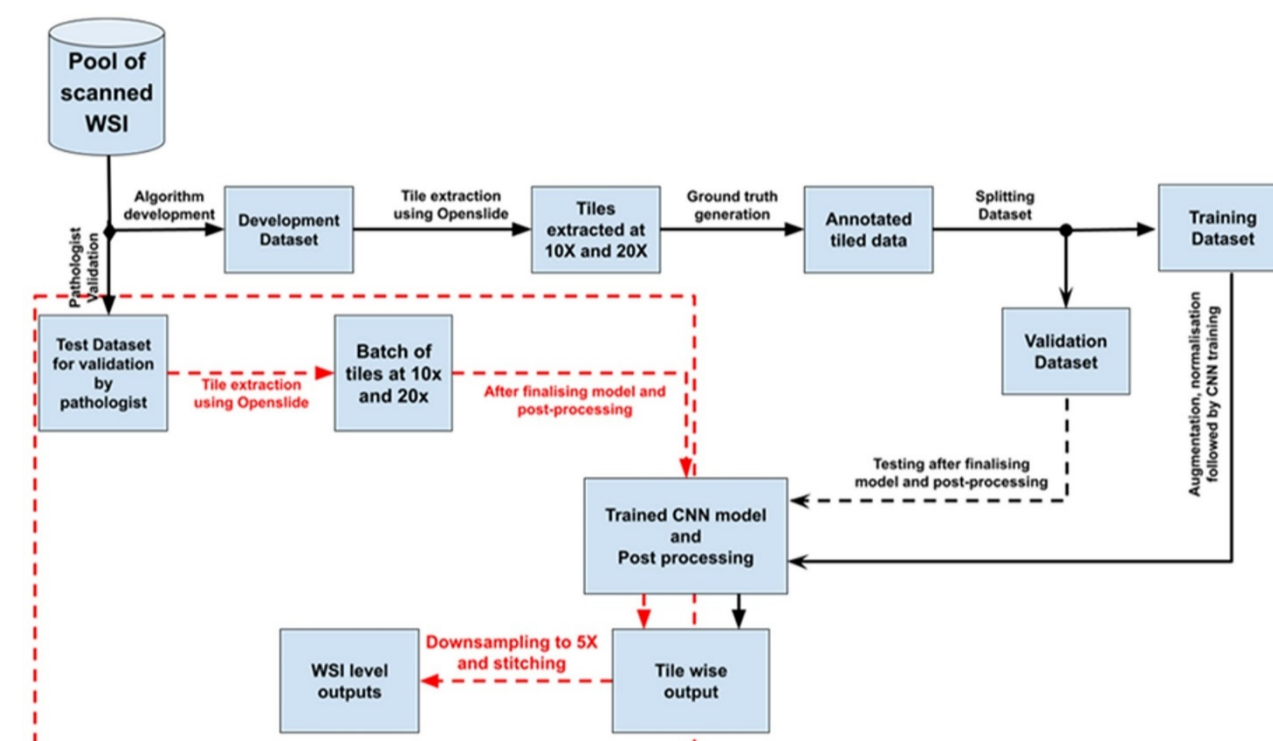

Fig.1: Workflow of algorithm development
Black dotted lines: to connect the process steps after the model development is completed.
Red dotted lines: to highlight the process steps connected with a validation by the pathologists.

- Validation of the algorithm
  - From the analysis of the 255 WSIs of the liver (validation dataset) by the trained algorithm, 2 categories of information were gathered. The first shows annotated image results with diagnosis discovered by the algorithm (Fig. 2), whereas the second includes a quantification for each of the findings.
  - First, pathologists double-checked the annotated data to ensure that the true lesion locations were marked. Then, histopathological data ("no findings (−)" or "findings (+)") diagnosed by the pathologists were concatenated with the quantitative values obtained from the algorithm for each specimen.
  - The most reliable thresholds were calculated for each finding based on a receiver operating characteristic (ROC) curve using JMP software (version 13.0.0, SAS Institute, Inc., USA). The best threshold value was calculated by maximizing Youden's index (Recall + Specificity − 1) in the ROC curve. The discriminative performance was evaluated based on the area under the ROC curve (AUC-ROC).
  - Based on the threshold value from the ROC curve, binary diagnostic results by the pathologists were classified into four classes: true positive, false positive, false negative, or true negative for each finding. The statistical parameters, including the F1-score, were calculated (Table 2).
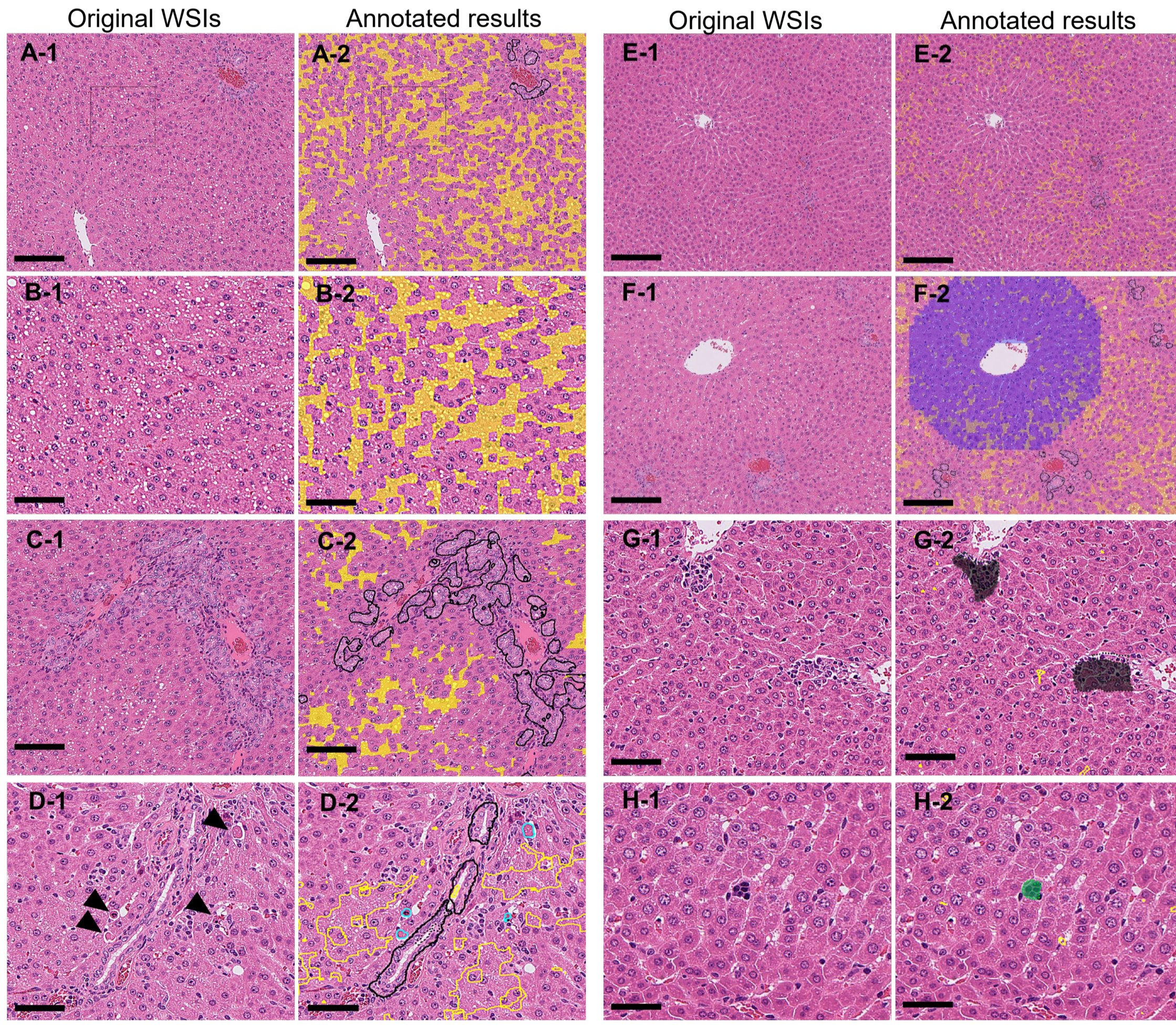
～Result～

Fig. 2: Annotated results on the WSIs after image analysis by the algorithm.
Original WSIs: Images before image analysis by the algorithm.
Annotated results: Images with colored annotation of findings and diagnosis, after image analysis by the algorithm.
A-1: Vacuolation (drug-induced) at the periportal area to the midzonal and normal bile ducts within the Glisson's sheath.
A-2: The vacuolated area was annotated (filled) with yellow. (Bar = 200 μm).
B-1: Higher magnification of the dashed area of A-1.
B-2: Higher magnification of the dashed area of A-2. The vacuolated area was annotated with yellow. (Bar = 100 μm)
C-1: Bile duct hyperplasia (drug-induced) at the periportal area.
C-2: The lesional areas (bile duct hyperplasia and vacuolation) were annotated with black and yellow, respectively. (Bar = 100 μm)
D-1: Single cell necrosis (arrowhead) and slightly vacuolated hepatocytes at the periportal area (drug-induced).
D-2: Lesional areas (single cell necrosis and vacuolation) were annotated with light blue and yellow, respectively. (Bar = 100 μm)
E-1: Non-treated liver (control animal of F-1).
E-2: The areas of vacuolation (spontaneous) and bile ducts were annotated with yellow and black, respectively. (Bar = 200 μm)
F-1: Hepatocellular hypertrophy (drug-induced) at the central area.
F-2: The area of hypertrophy was annotated in blue. (Bar = 50 μm)
G-1: Microgranuloma (spontaneous) near central veins.
G-2: Microgranuloma was annotated with gray. (Bar = 100 μm)
H-1: An erythroblastic island (spontaneous) in the sinusoids.
H-2: Extramedullary hematopoiesis was annotated as green. (Bar = 50 μm)
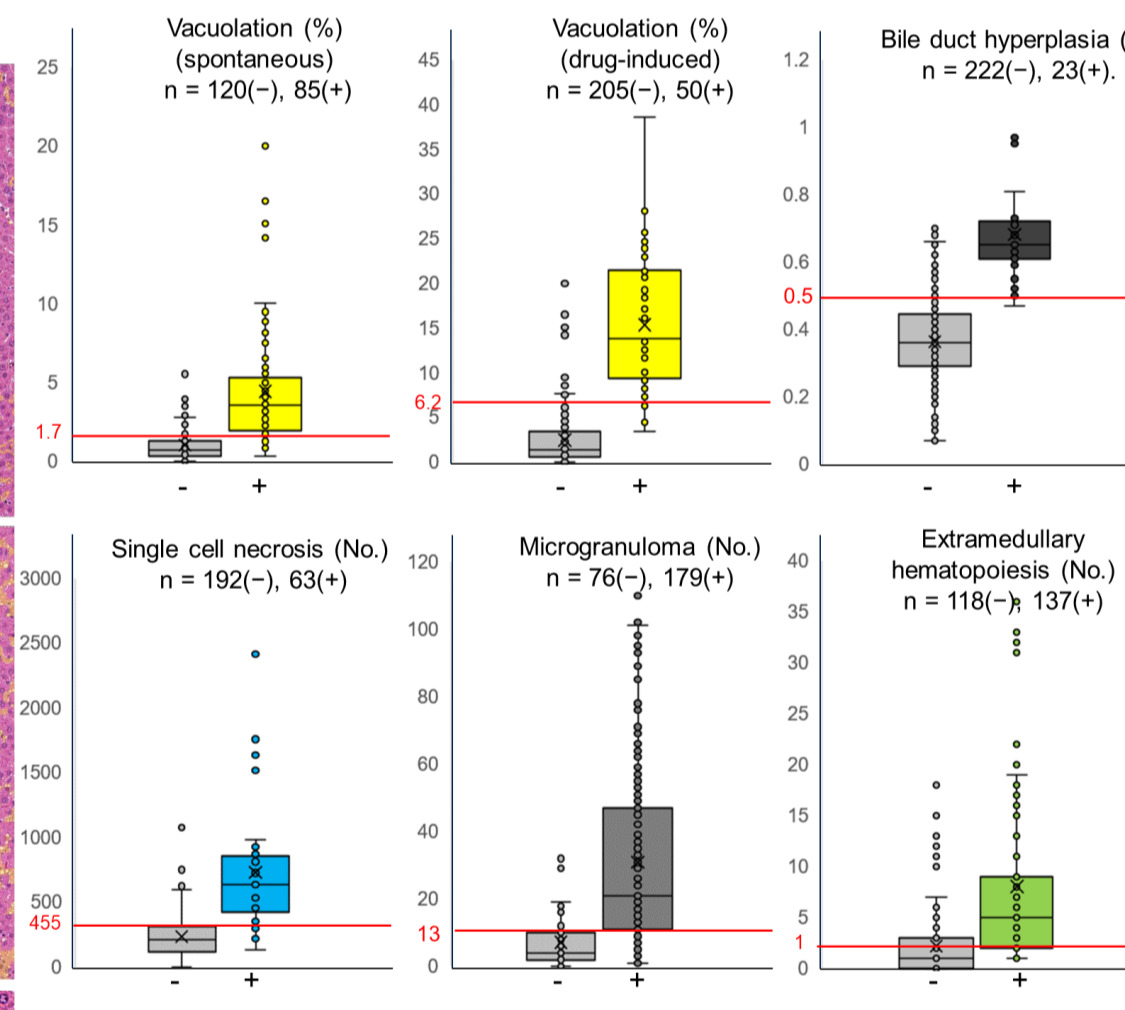

Fig. 3: Comparison of the quantitative values between binary classifications by the pathologists.
The horizontal axis shows binary classification judged as "no findings (−)" or "findings (+)" by pathologists, and the vertical axis shows the quantitative values calculated by the algorithm. Here, 〔%〕 indicates the ratio and 〔No.〕 indicates the number of annotated areas of findings in the WSI. The red line crosses the vertical axis and its numerical value indicates the threshold value of the finding calculated from the ROC curve. In the "no findings (−)" group, plotted samples above the threshold value indicate false positives, and plotted samples below the threshold value indicate true negatives. By contrast, in the "finding (+)" group, plotted samples above the threshold value indicate true positives, and plotted samples below the threshold value indicate false negatives.

Fig. 3 shows that, as for the five findings other than extramedullary hematopoiesis, the thresholds bisected the body of the box, indicating that approximately 75% of the total sample could be classified as a true positive or true negative. However, for extramedullary hematopoiesis, the thresholds intersect the body of the box in both groups.

Table 2: Statistical parameters derived as indices for performance of lesion detection for each finding

Findings	AUC-ROC	Threshold	Recall	Specificity	Precision	Balanced accuracy	F1 score
Vacuolation (spontaneous)	0.91	1.72*	0.84	0.86	0.81	0.85	0.82
Vacuolation (drug-induced)	0.97	6.23*	0.96	0.92	0.75	0.94	0.84
Bile duct hyperplasia	0.97	0.5*	0.96	0.87	0.42	0.91	0.59
Single cell necrosis of hepatocyte	0.93	455**	0.84	0.93	0.80	0.89	0.82
Microgranuloma	0.84	13**	0.67	0.88	0.93	0.78	0.78
Extramedullary hematopoiesis	0.74	1**	0.98	0.41	0.66	0.69	0.79
Hepatocellular hypertrophy	NA	NA	0.68	0.86	0.30	0.77	0.42

*: Percentage of lesional area in a WSI
**: Number of foci in a WSI
NA (Not applicable): No quantitative values because only qualitative data (normal or abnormal) are generated by the algorithm.

Table 2 shows that six findings, except for hepatocellular hypertrophy, indicated a high AUC on the ROC curve and the F1-score, which is a comprehensive evaluation index of accuracy and comprehensiveness based on the numbers of true positive, true negative, false positive and false negative.

$$Recall = \frac{True\ Positives}{(True\ Positives + False\ Negatives)}$$

$$Specificity = \frac{True\ Negatives}{(True\ Negatives + False\ Positives)}$$

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)}$$

$$Balanced\ Accuracy = \frac{(Recall + Precision)}{2}$$

$$F1\ Score = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

～Discussion & Conclusions～

- The algorithms showed consistently good performance in detecting histopathological findings. Approximately 75% of all specimens could be classified as true positive or true negative. In general, findings with clear boundaries with the surrounding normal structures, such as vacuolation and single-cell necrosis, were accurately detected with high statistical scores.
- The results of quantitative analysis and classification of the diagnosis based on the threshold values between "no findings" and "findings" correlated well with diagnoses made by professional pathologists. However, the scores for findings with ambiguous boundaries, such as hepatocellular hypertrophy, were poor.
- These results suggest that deep learning-based algorithms can detect, classify, and quantify multiple findings simultaneously on rat liver WSIs with high accuracy. Thus, it can be a useful supportive tool for a histopathological evaluation, especially for primary screening in rat toxicity studies.

References
1. Ronneberger O, Fischer P and Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer. 2015.
2. Abraham N and Khan NM. A novel focal Tversky loss function with improved attention U-Net for lesion segmentation. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). Venice. Italy. 683-687. 2019.
3. Kingma D and Ba J. (2014). Adam: A Method for Stochastic Optimization. International Conference on Learning Representations. In: 3rd International Conference for Learning Representations. San Diego. 2015.