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

Taishi Shimazaki[1], Kyotaka Muta[1], Naohito Yamada[1], Yuzo Yasui[1], Toshiyuki Shoda[1], Adarsh Kumar[2], Apeksha Jain[2], Dev Kumar Das[2] and Uttara Joshi[2]

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

2AIRA Matrix, Mumbai, Maharashtra, India

COI Disclosure Information :
We declare no conflicts of interest associated with this poster.

～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 a Deep Learning (DL) -based method for classification and quantification of multiple histopathological lesions in rodent liver and kidney.

The trained algorithms were validated using 255 liver Whole Slide Images (WSIs) to detect, classify, and quantify the seven findings in the liver. A modified form of the U-Net DL model[1-3] was trained using data from WSIs of 92 liver sections and 90 kidney sections. The trained model was used for identifying and quantifying 7 types of histopathological findings in both liver (vacuolation, bile duct hyperplasia, single-cell necrosis, microgranuloma, EMH and hypertrophy) and kidney (vacuolation, basophilia/degeneration/regeneration tubule, dilatation, hyaline cast, mineralization, mononuclear cell infiltration and cyst). The algorithm was validated by comparing the results with pathologists' findings on 255 liver sections 285 kidney sections.

～Materials & Methods～

■ **Generation of WSIs**
- 406 and 418 HE-stained glass slides of liver and kidney specimens, respectively, 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 for liver and 418 for kidney dataset was divided into a development dataset and a validation dataset. 92 WSIs for liver and 90 WSIs for kidney from the development dataset were used to train the deep learning models for each of 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 for liver ,and 28 and 15 WSIs for kidney. (Table 1 summarizes the data distributions.)

■ **Workflow of algorithm development** (Fig. 1)
- Ground truth annotations for all the 14 lesions were created by data-marking experts under the guidance of pathologists, who further verified the annotated data after marking.
- The annotated images were used for training 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 reached an agreement.

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

Findings	Required Number of WSI			
	Development Dataset			Validation Dataset
	Training	1st Test	2nd Test	Validation
Liver				
Vacuolation (spontaneous) of hepatocyte	8	4	18	205
Vacuolation (drug-induced) of hepatocyte	10	5	18	255
Bile duct hyperplasia	13	9	18	255
Single cell necrosis of hepatocytes	13	6	18	255
Microgramuloma	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*
Kidney				
Vacuolation	8	4	15	285
Basophilic tubule/degeneration	12	7	15	285
Dilatation	12	2	15	285
Mineralization	9	6	15	285
Cyst	10	2	15	285
Hyaline cast	16	4	15	285
Mononuclear cell infiltration	10	3	15	285
WSIs with no histopathological findings	13	-	-	
Total number of WSIs	90	28	15	285

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

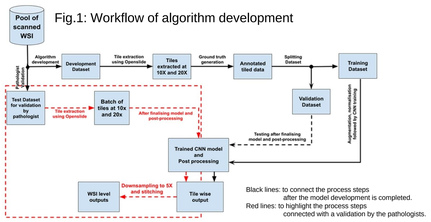

Fig.1: Workflow of algorithm development

■ **Validation of the algorithm**
- From the analysis of the 255 WSIs of the liver and 285 WSIs of the kidney (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).

～Discussion & Conclusions～

- The algorithms showed consistently good performance across all the finding from both kidney and liver section. Approximately 75% of the validation data is accurately classified by the algorithm . In general, lesions that are well defined having contrasting background, 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 pathologists.
- These results suggest that deep learning-based algorithms can detect, classify, and quantify multiple findings simultaneously on rat liver and kidney 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.

～Result～

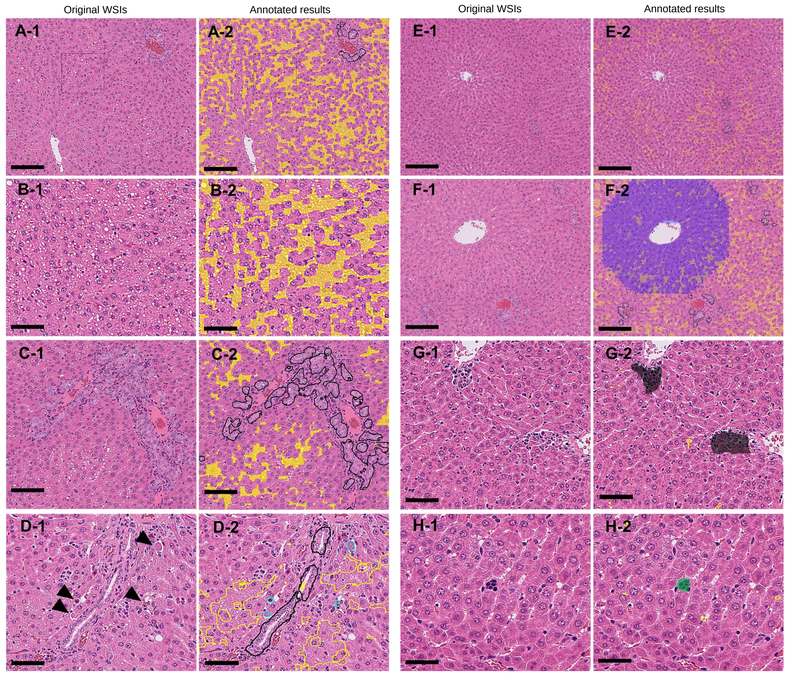

Fig. 2: Annotated results on the WSIs before/after image analysis by the algorithm of the Liver.
Original WSIs: Images before image analysis by the algorithm.
Annotated results: Images with colored annotation of findings and diagnosis, after image analysis by the algorithm.

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 = 200 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)

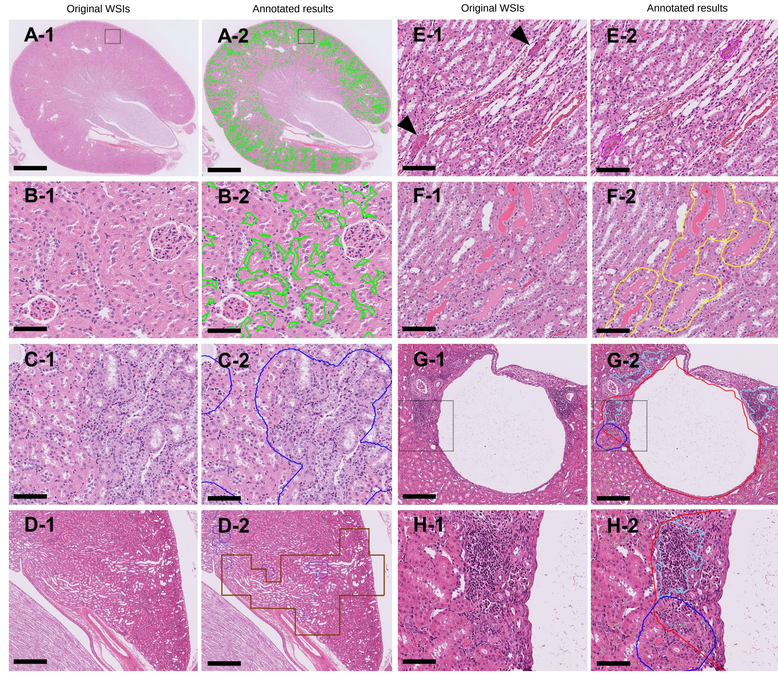

Fig. 3: Annotated results on the WSIs before/after image analysis by the algorithm of the Kidney.
Original WSIs: Images before image analysis by the algorithm.
Annotated results: Images with colored annotation of findings and diagnosis, after image analysis by the algorithm.

A-1: Vacuolation (drug-induced) at the cortex in lower magnification.
A-2: The vacuolated area was annotated with green. (Bar = 2 mm).
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 green. (Bar = 75 μm)
C-1: Degeneration/regeneration of proximal tubules (drug-induced) at the cortex.
C-2: The lesional areas (degeneration/regeneration of proximal tubules) were annotated with blue. (Bar = 150 μm)
D-1: Dilatated tubules at the cortex (drug-induced) .
D-2: The Lesional areas (dilatated tubule) were annotated with brown. (Bar = 500 μm)
E-1: Mineralization(arrowhead) of the colleting duct (spontaneous).
E-2: The areas of mineralization (spontaneous) were annotated with purple. (Bar = 150 μm)
F-1: Hyaline cast (drug-induced) at the distal tubules.
F-2: The lesional areas of hyaline cast were annotated with yellow. (Bar = 150 μm)
G-1: Cyst, mononuclear cell infiltration and basophilic tubules (spontaneous) at the cortex.
G-2: The lesional areas were annotated with red, light blue and blue, respectively. (Bar = 400 μm)
H-1: Higher magnification of the dashed area of G-1.
H-2: Higher magnification of the dashed area of G-2. Mononuclear cell infiltration and basophilic tubules were annotated with light blue and blue, respectively. (Bar = 100 μm)

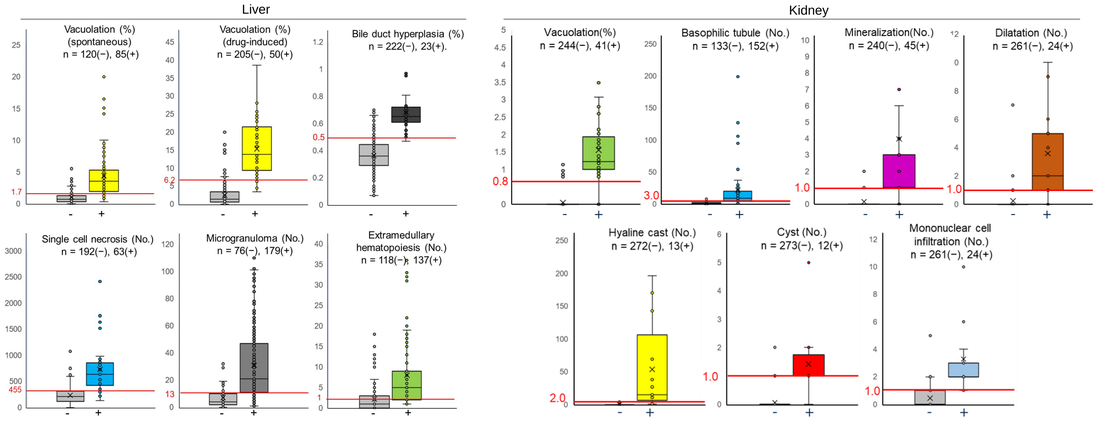

Fig. 4: Comparison of the quantitative values between binary classifications by the pathologists.
Six figures on the left represent liver findings and seven figures on the right represent kidney findings.
The horizontal axis shows binary classification judged as “no findings (−)” or “findings (+)” by 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. 4 shows that, as for the five findings other than extramedullary hematopoiesis of the liver, 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.

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
Liver							
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 hepatocytes	0.93	455**	0.84	0.93	0.80	0.89	0.82
Microgramuloma	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
Kidney							
Vacuolation	0.95	0.79*	0.93	0.93	0.70	0.93	0.80
Basophilic tubule/degeneration	0.87	3**	0.75	0.89	0.88	0.82	0.81
Dilatation	0.88	1**	0.79	0.96	0.63	0.87	0.70
Mineralization	0.84	1**	0.76	0.88	0.54	0.82	0.63
Cyst	0.94	1**	0.92	0.97	0.55	0.94	0.69
Hyaline cast	0.98	2**	0.85	0.95	0.44	0.90	0.58
Mononuclear cell infiltratino	0.89	1**	0.92	0.76	0.26	0.84	0.40

*: 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 almost all findings, except for hepatocellular hypertrophy, hyaline cast and mononuclear cell infiltration, 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.