P-71

Deep learning-based image analysis model for detecting unlearned findings in rat liver

○Taishi Shimazaki[1], Yuzo Yasui[1], Aman Shrivastava[2], Raghav Amaravadi[2], Geetank Raipuria[2], Tijo Thomas[2], Nitin Singhal[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

- We previously developed supervised models capable of accurately detecting learned abnormalities in Whole Slide Images (WSIs) of the liver, kidney, testis, epididymis, and brain in Sprague Dawley (SD) rats.
- However, the models developed based on supervised approaches were unable to identify unlearned findings with sufficient accuracy.
- We developed the models using unsupervised development approaches into the current image analysis framework for rat liver and assessed its performance in identifying various toxicity findings in the rat liver.

Discussion & Conclusion

- ✓ The model's ability to identify liver findings, including unlearned findings, was superior to that of the previous supervised learning models.
- ✓ Additionally, the model's accuracy was generally consistent with the histopathological diagnosis made by the JSTP-certified pathologists.
- ✓ Moreover, the findings overlooked by the previous supervised model could be identified with greater performance.

➢ The models' capacity to identify liver abnormalities was considered highly beneficial as a supplementary tool to histological assessment by pathologists, primarily for screening liver toxicity in non-GLP early toxicity investigations.

Materials & Methods

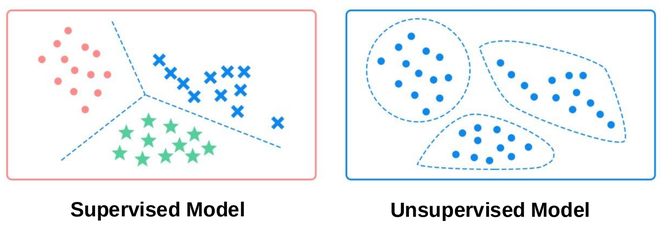

Figure 1: Supervised model utilizes labelled input and output data, whereas unsupervised model operates without such labels.

■ Tissue Preparation

WSIs of H&E-stained glass slide specimens of young SD rats (males, 6 to 8 weeks old) for developing supervised and unsupervised models and young to old SD rats (males, 8 to 104 weeks old) for validating the models were scanned at 20x.

■ Training Phase for Supervised Model with Convolutional Neural Network (CNN) (Previous model)

- The training dataset for the supervised model development consisted of 1024x1024 sized image tiles extracted at 20x magnification from 406 WSIs with annotated major histopathological findings in the liver.
- 8 findings in the liver - Necrosis, Single cell necrosis, Hypertrophy and Vacuolation of hepatocytes, Bile duct hyperplasia, Microgranuloma, Extramedullary hematopoiesis, and Mononuclear cell infiltration.

■ Training Phase for Unsupervised Model (Current model)

- 100 WSIs with untreated rat liver tissue sections for In-distribution modeling.
- Foundation model [1] trained on 100 million+ tiles obtained from different H&E-stained rat tissue images.
- In-distribution modeling to learn normal data representation
- Out-of-distribution identification on treated test data.

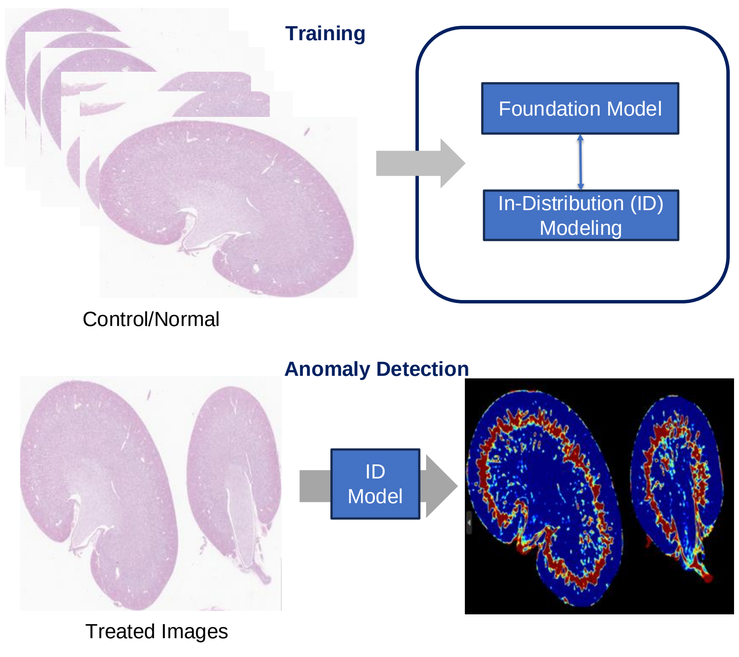

Figure 1: Design of Unsupervised Anomaly Detection Framework

■ Validation Phase for Unsupervised Model

The test dataset of group treated with vehicles (8 to 10 weeks old rats: N=30, 104 weeks old rats: N= 60) and group treated with certain compounds (8 to 10 weeks old rats: N=30) was analyzed for validation.
From the analysis of the 120 WSIs of the liver by the trained unsupervised model, 3 categories of information were gathered. The first includes heat map on the WSI and an indicator (called Anomaly score) of how many abnormal findings exist in the WSI, including findings that are not yet learned (Fig. 1A1, 1B1, 2A1, 2B1, etc.).
The second shows annotated image results with names of appropriate findings diagnosed by the models (Fig. 1A2, 1B2, 2A2, 2B2, etc.).
And the third indicates a quantification for each of the findings.
First, JSTP-certified pathologists double-checked the both anomaly and 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 models for each specimen.
The most reliable thresholds were calculated for each finding based on a receiver operating characteristic (ROC) curve. 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 (TP), false positive (FP), false negative (FN), or true negative (TN) for each finding. The statistical parameters, including the F1-score, were calculated (Table 2 and 3).
Figure 2 and 3 show the learned findings detected by the current model, and Figure 4 shows the unlearned findings that could not be detected by the previous (only supervised) models but could be detected by the current model that incorporated unsupervised learning.

Results

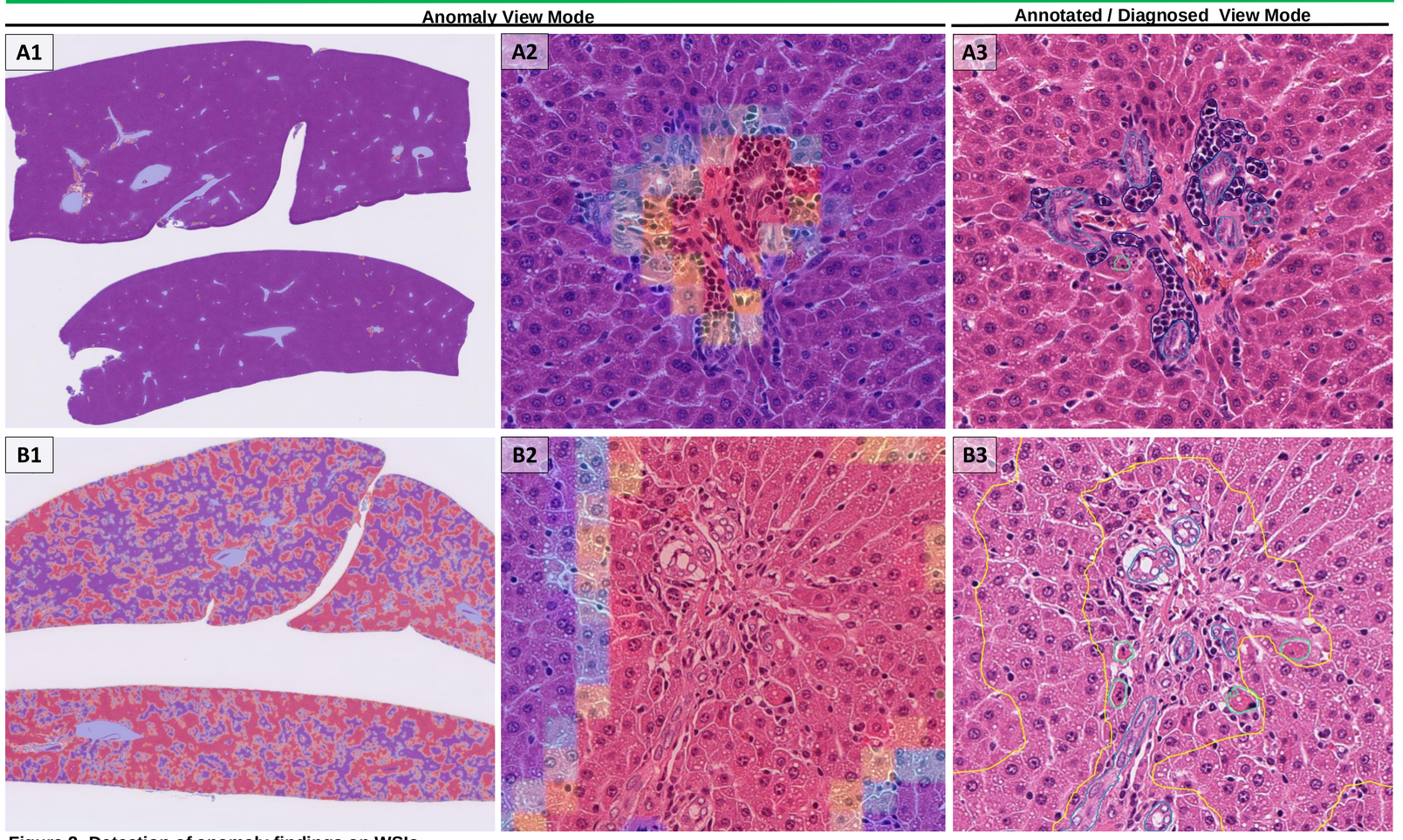

Figure 2: Detection of anomaly findings on WSIs
A: Vehicle control animal. B: Compound treated animal.
A1: In most areas, there are no anomaly areas (red). Anomaly score is 0.02.
A2: High magnification of A1. Anomaly area (spontaneous extramedullary hematopoiesis) is detected sporadically in the periportal area.
A3: Annotated image of A2. Spontaneous single cell necrosis of hepatocytes (light green), extramedullary hematopoiesis (purple), and bile ducts (light blue) are detected, classified, and diagnosed by the model and are annotated in their respective colors.
B1: Anomaly areas (yellow to red) are multifocally detected. Anomaly score is 0.58.
B2: High magnification of B1. Anomaly area (induced single cell necrosis and vacuolation of hepatocytes) is strongly detected in the periportal area.
B3: Annotated image of B2. Induced single cell necrosis (light green) and vacuolation (yellow) of hepatocytes, and bile ducts (light blue) are detected, classified, and diagnosed by the model and are annotated in their respective colors.

Table 2: Performance of lesion detection for anomaly findings

Findings	AUC-ROC	Threshold	Recall	Specificity	Precision	Balanced accuracy	F1 Score
Anomaly findings	**0.97**	**0.15**	**0.90**	**0.89**	**0.93**	**0.92**	**0.92**

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

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

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

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

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

The performance by the current (supervised with unsupervised learning) model in detecting the anomaly findings on WSIs was high and in accordance with the expert pathologist’s results.

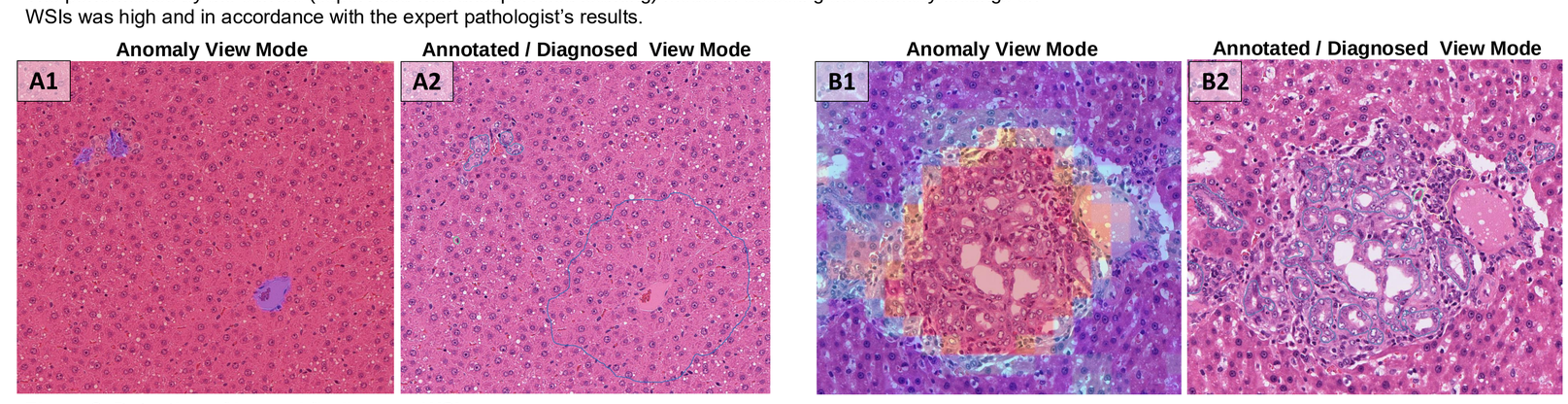

Figure 3: Detection of learned findings on WSIs
A1: Anomaly area (induced hepatocellular hypertrophy) is detected in the centrilobular to periportal area.
A2: Annotated image of A1. Induced hepatocellular hypertrophy (blue) are detected, classified, and diagnosed by the model and are annotated in blue color.
B1: Anomaly area (induced bile duct hyperplasia) is detected in the periportal area.
B2: Annotated image of B1. Induced bile duct hyperplasia (light blue) are detected, classified, and diagnosed by the model and are annotated in light blue color.

Table 3: Comparison of the detection performance of the learned findings between the previous model and the current model

	Previous model (only supervised model)							Current model (supervised with unsupervised model)						
Findings	**AUC-ROC**	**Threshold**	**Recall**	**Specificity**	**Precision**	**Balanced accuracy**	**F1 score**	**AUC-ROC**	**Threshold**	**Recall**	**Specificity**	**Precision**	**Balanced accuracy**	**F1 score**
Necrosis of hepatocytes	0.93	3**	0.84	0.93	0.88	0.86	**0.86**	0.95	0.07*	0.88	0.94	0.85	0.87	**0.87**
Single cell necrosis of hepatocytes	0.93	455**	0.84	0.93	0.80	0.82	**0.82**	0.96	0.01*	0.88	0.93	0.87	0.88	**0.88**
Vacuolation of hepatocytes	0.91	1.72*	0.84	0.86	0.81	0.82	**0.82**	0.93	1.94*	0.86	0.90	0.84	0.85	**0.85**
Hepatocellular hypertrophy	NA	NA	0.68	0.86	0.30	0.49	**0.42**	0.86	3.8*	0.78	0.81	0.80	0.79	**0.79**
Bile duct hyperplasia	0.87	0.5*	0.96	0.87	0.42	0.69	**0.59**	0.95	0.5*	0.97	0.88	0.74	0.85	**0.84**
Microgranuloma	0.84	13**	0.67	0.88	0.93	0.80	**0.78**	0.94	0.21*	0.77	0.87	0.95	0.86	**0.85**
Extramedullary hematopoiesis	0.74	1**	0.98	0.41	0.66	0.82	**0.79**	0.93	0.01*	0.98	0.64	0.77	0.87	**0.86**
Mononuclear cell infiltration	0.77	4**	0.88	0.64	0.67	0.77	**0.76**	0.89	0.01*	0.93	0.70	0.73	0.83	**0.81**
Unlearned findings	NA	NA	0.28	0.76	0.63	0.45	**0.39**	0.97	0.15*	0.90	0.89	0.93	0.92	**0.92**

*: 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 model.

In the previous (only supervised learning) model, due to the high number of false positives and false negatives, various parameters (red colored), including the F1 score, were observed to be low. However, by integrating unsupervised learning, the current model has improved detection performance for all findings, especially unlearned findings.

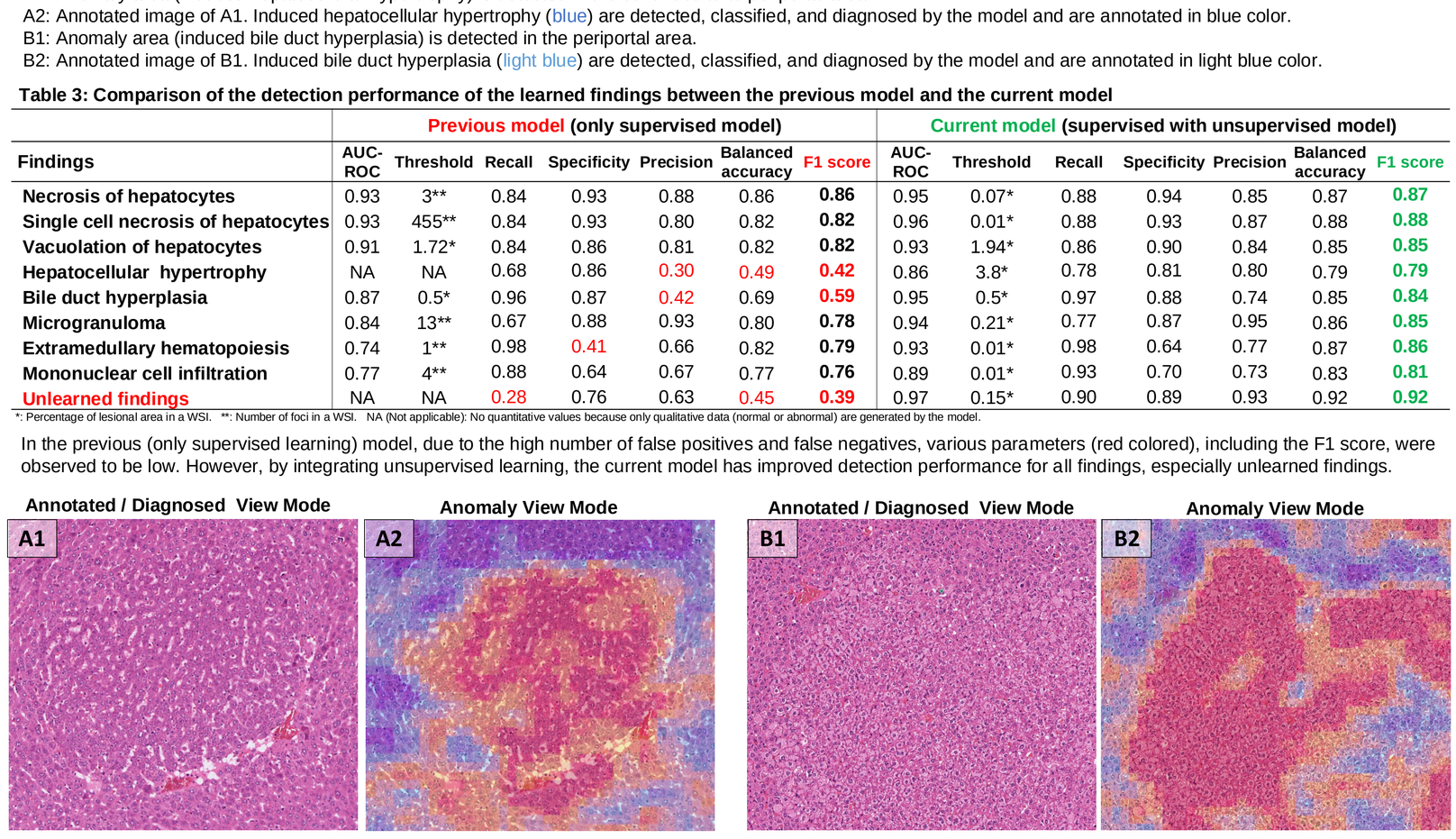

Figure 4: Detection of unlearned findings on WSIs
A: Spontaneous basophilic focus, B: Spontaneous eosinophilic focus, C: Spontaneous hepatocellular adenoma, D: Spontaneous cystic degeneration,
E: Spontaneous angiectasis, F: Spontaneous thrombosis
X1: The finding is not detected by the only supervised model.
X2: However, anomaly area (the finding) is detected by using the supervised model with unsupervised model.

References

1. Maxime Oquab et al., “DINOv2: Learning Robust Visual Features Without Supervision”, arXiv, 2304.07183, 2024

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