Contents lists available at ScienceDirect

Journal of Pathology Informatics

journal homepage: www.elsevier.com/locate/jpi

Short Communication

Digital workflows for pathological assessment of rat estrous cycle stage using images of uterine horn and vaginal tissue

Shinichi Onishi[a,1], Riku Egami[b,1], Yuya Nakamura[b], Yoshinobu Nagashima[b], Kaori Nishihara[a], Saori Matsuo[a], Atsuko Murai[a], Shuji Hayashi[a], Yoshifumi Uesumi[b], Atsuhiko Kato[a], Hiroyuki Tsunoda[b], Masaki Yamazaki[c,*], Hideaki Mizuno[b,*]

[a] *Translational Research Division, Chugai Pharmaceutical Co. Ltd., 1-135 Komakado, Gotemba, Shizuoka 412-8513, Japan*
[b] *Research Division, Chugai Pharmaceutical Co. Ltd., 200 Kajiwara, Kamakura, Kanagawa 247-8530, Japan*
[c] *Translational Research Division, Chugai Pharmaceutical Co. Ltd., 200 Kajiwara, Kamakura, Kanagawa 247-8530, Japan*

ARTICLE INFO

Keywords:
Digital workflow
Estrous cycle
Image recognition
Deep learning
Pathological assessment

ABSTRACT

Assessment of the estrous cycle of mature female mammals is an important component of verifying the efficacy and safety of drug candidates. The common pathological approach of relying on expert observation has several drawbacks, including laborious work and inter-viewer variability. The recent advent of image recognition technologies using deep learning is expected to bring substantial benefits to such pathological assessments. We herein propose 2 distinct deep learning-based workflows to classify the estrous cycle stage from tissue images of the uterine horn and vagina, respectively. These constructed models were able to classify the estrous cycle stages with accuracy comparable with that of expert pathologists. Our digital workflows allow efficient pathological assessments of the estrous cycle stage in rats and are thus expected to accelerate drug research and development.

Introduction

In the process of drug development, preclinical efficacy and toxicology studies using laboratory animals such as mice, rats, dogs, and monkeys are routinely conducted. Sexually mature female animals exhibit an estrous cycle comprising the diestrous (D), proestrous (P), estrous (E), and metestrous (M) stages. This cycle entails changes in hormone levels[1] and gene expression,[2] consequently affecting responsiveness to drugs.[3] Thus, when drugs are tested *in vivo*, knowledge of the estrous cycle stage of female individuals is preferable for precise interpretation of the results. In toxicology settings, vigilance regarding drug-related risks such as disruption or prolongation of the cycle is needed from the viewpoint of reproductive safety.[2,4] Because drugs might exhibit tissue-specific effects according to their mode of action, assessment of each reproductive tissue (e.g., ovary, uterus, and vagina) is desirable. Such assessments are usually performed pathologically by experts to elucidate morphological features; however, they have the disadvantages of a laborious workload, intra- and inter-viewer variability,[5,6] subjectivity, and potential for bias.

Recent advancements in deep learning-based image recognition technologies are rapidly transforming broad pathology tasks that range from gaining new insights to assisting with routine work.[7,8] To extend the application of deep learning-based image recognition technologies to estrous cycle stage assessment, one study introduced the “Stage Estimator of estrous Cycle of RodEnt using an Image-recognition Technique” (SECREIT) framework.[9] This framework distinguishes 3 stages of the estrous cycle (D, E, and P) from rodent vaginal smear cytology images using convolutional neural network (CNN). Another study applied Faster region CNN (Faster R-CNN)[10] to whole slide images (WSIs) of hematoxylin and eosin stained ovaries to quantify 3 classes of ovarian follicles that are susceptible to the estrous cycle.[4] In this report, we present 2 alternative deep learning-based workflows for discriminating 4 stages of the estrous cycle in the uterine horns and vaginas of rats on WSIs by detecting tissue-level features and aggregating patch-level features, respectively. These approaches are robust against variations in tissue alignments on slides and enable automatic estrous cycle stage assessment with accuracy comparable with that of experienced pathologists. Together, these proposed workflows should contribute

* Corresponding authors.
E-mail addresses: onishi.shinichi58@chugai-pharm.co.jp (S. Onishi), egami.riku72@chugai-pharm.co.jp (R. Egami), nakamura.yuya84@chugai-pharm.co.jp (Y. Nakamura), nagashima.yoshinobu90@chugai-pharm.co.jp (Y. Nagashima), nishihara.kaori63@chugai-pharm.co.jp (K. Nishihara), matsuosor@chugai-pharm.co.jp (S. Matsuo), murai.atsuko79@chugai-pharm.co.jp (A. Murai), hayashisuj@chugai-pharm.co.jp (S. Hayashi), uesumi.yoshifumi80@chugai-pharm.co.jp (Y. Uesumi), katoath@chugai-pharm.co.jp (A. Kato), tsunodahry@chugai-pharm.co.jp (H. Tsunoda), yamazakimsk@chugai-pharm.co.jp (M. Yamazaki), mizunohda@chugai-pharm.co.jp (H. Mizuno).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jpi.2022.100120
Received 3 March 2022; Received in revised form 23 June 2022; Accepted 24 June 2022
Available online 29 June 2022

Table 1
Estrous cycle stages of uterine horn and vagina.

Stage	Criterion
Uterus	
D	Shortest epithelium and narrow lumen
P	Expanded lumen, round or oval stromal cells, and edematous stroma
E	Tall epithelium containing cellular debris
M	Short epithelium containing decreased apoptoses and increased mitosis
Vagina[a]	
D	Epithelium with few neutrophils and without mucus
P	Epithelium with mucus
E	Keratinized epithelium
M	Thickened epithelium with neutrophils

D, diestrous; P, proestrous; E, estrous; M, metestrous

[a] Judged in peripheral mucosa

to the efficiency of toxicological evaluation processes using laboratory animals.

Methods

Dataset

Ninety WSIs containing rat uterine horn and vagina tissues were collected from 5 independent archived toxicology studies. In those studies, all animal procedures were conducted in accordance with the Institute's Guide for the Care and Use of Laboratory Animals, and all experimental protocols were approved by the Institutional Animal Care and Use Committee.

WSIs commonly contain multiple reproductive tissues such as the ovary, uterine horn, uterine cervix, and vagina in different combinations

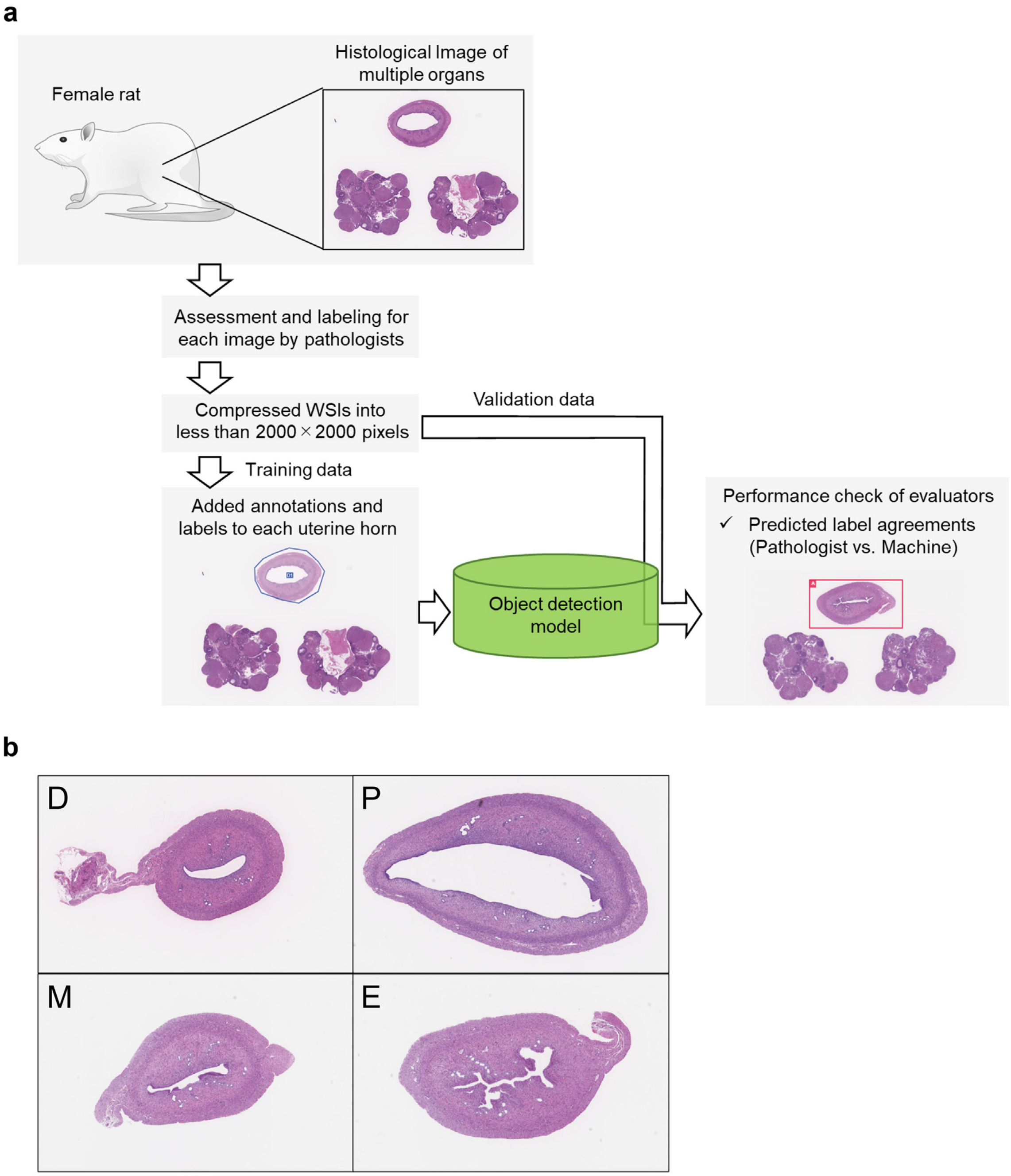

a
Female rat
Histological Image of multiple organs
Assessment and labeling for each image by pathologists
Compressed WSIs into less than 2000 × 2000 pixels
Validation data
Training data
Added annotations and labels to each uterine horn
Object detection model
Performance check of evaluators
✓ Predicted label agreements (Pathologist vs. Machine)
b
D
P
M
E

Fig. 1. (a) Overview of workflow for analyzing uterine horns in whole slide images. (b) Representative images of uterine horns in diestrous, proestrous, estrous, and metestrous. WSIs, whole slide images; D, diestrous; P, proestrous; E, estrous; M, metestrous.

depending on the purpose of the study, as shown in Fig. S1. Uterine horns and vaginas were independently examined by 5 certified pathologists using actual glass slides or WSIs and were annotated for estrous cycle stages using the criteria listed in Table 1.

Workflow for uterine horn

Model for detecting stage-specific uterine horn in WSIs

When developing a prediction model for the estrous cycle stages of reproductive tissues, WSIs containing multiple tissues should be used with caution. Tissues in the same individual normally show the same estrous cycle; thus, information from other tissues may distract models from the learning features of each tissue. In the uterine horns, the estrous cycle stage can be recognized using whole tissue at low magnification as shown in Table 1. Therefore, an object detection approach was employed. The overview is illustrated in Fig. 1a. Sixty-eight WSIs with more than 4 votes from pathologists for the same stage were randomly divided into a training dataset (48 WSIs) and a validation dataset (20 WSIs). All WSIs were compressed so that the long side was 2000 pixels. The uterine horns and their estrous cycle stage were annotated by bounding corresponding objects in images using Visual Inspection (1.3.0.1, IBM), a multi-purpose image recognition system that can be applied to tasks in bio-medical fields.[11,12] Examples are shown in Fig. 1b. Annotated images were augmented with the following parameters: color-brightness = 20, color-contrast = 20, color-

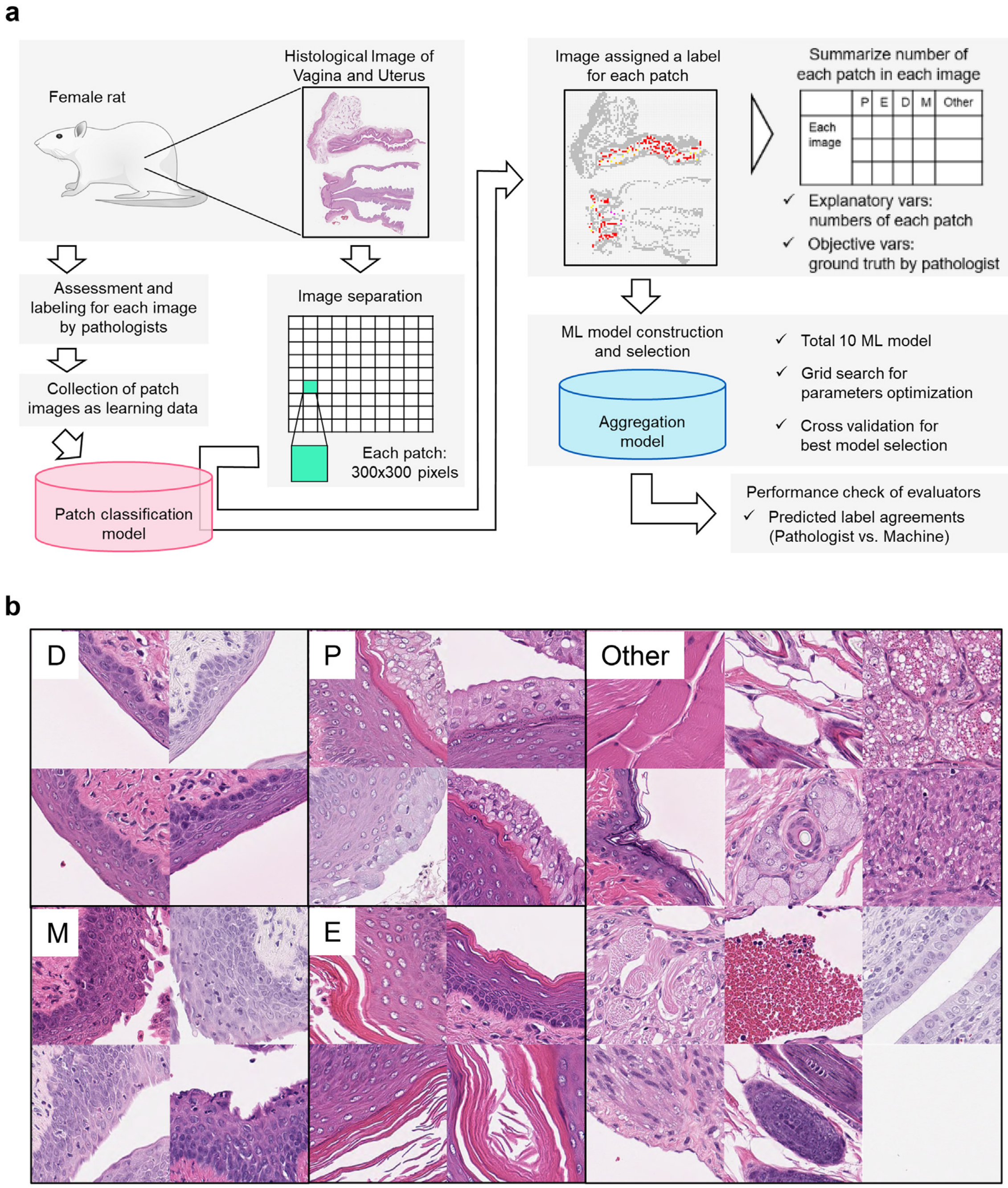

a
Female rat
Histological Image of Vagina and Uterus
Assessment and labeling for each image by pathologists
Image separation
Collection of patch images as learning data
Each patch: 300x300 pixels
Patch classification model
Image assigned a label for each patch
Summarize number of each patch in each image
P
E
D
M
Other
Each image
✓ Explanatory vars: numbers of each patch
✓ Objective vars: ground truth by pathologist
ML model construction and selection
Aggregation model
✓ Total 10 ML model
✓ Grid search for parameters optimization
✓ Cross validation for best model selection
Performance check of evaluators
✓ Predicted label agreements (Pathologist vs. Machine)
b
D
P
Other
M
E

Fig. 2. (a) Overview of workflow for analyzing vaginas in whole slide images. This workflow mainly consists of 2 processes: the former process of identifying patches that indicate the stages of the estrous cycle using an image classification model, and the latter process of assessing the final label by aggregating the number of patches classified into each stage using machine learning. The performance of the workflow for the vagina was investigated by comparing the agreement score of classification labels for the evaluation data for every pair of pathologist and machine. (b) Representative patches for diestrous, proestrous, estrous, and metestrous stages of the estrous cycle and Other. ML, machine learning; D, diestrous; P, proestrous; E, estrous; M, metestrous.

hue = 20, color-saturation = 20, blur = 20, sharpen = 20, noise = 20, vertical flip = TRUE, and horizontal flip = TRUE, yielding 1344 images. The object detection model to identify a uterine horn in the D, P, E, or M stage in a given image was then generated with the following parameters: algorithm = Faster R-CNN and iteration = 4000.

Workflow for vagina

Model for assessing vagina in WSIs

Local morphological changes in the vaginal mucosa are indicative of estrous cycle stages. A critical issue of applying image recognition technology to pathological assessment is that WSIs generally exhibit over-resolution for most available deep learning algorithms.[13,14] Image compression is often used at this process, but inevitably causes information loss, making local features of the estrous cycle stage in the vagina indistinguishable (examples are shown in Fig. S2). In the preliminary study, the object detection approach for the vagina was tested but did not give satisfactory results, unlike in the uterine horn cases (data not shown). Thus, an approach of patch-level inference followed by aggregation of patch counts was adopted. The utility of this approach in pathological assessments was described in a recent review article.[14] This workflow, as indicated in Fig. 2a, consists of: (i) distinguishing indicative patches for estrous cycle stages using the image classification model and (ii) aggregating patch counts to manifest a final label using a conventional machine learning model.

Building the patch-level classification model

Seventy-five WSIs were split into patches with 300 × 300 pixels without overlapping using image processing Python libraries (numpy v1.20.3, openslide v1.1.2, pillow v8.2.0, and PyTorch v1.8.1). Yielded patches were down-sampled and manually assessed by a pathologist with a focus on following local morphologic features of the vaginal mucosa in each stage: D, epithelium with few neutrophils and without mucus; P, epithelium with mucus; E, keratinized epithelium; and M, thickened epithelium with neutrophils (Table 1). Representative patches of the D (n = 275), P (n = 243), E (n = 266), and M (n = 179) stages of the vaginal mucosa were collected to form corresponding patch classes. To explicitly distinguish irrelevant images from estrous cycle-indicative images, 2346 patches such as blood, fatty tissue, fibrous tissue, skeletal muscle, skin, uterus, gland, and blank were also picked up to form the “Other” class. Example patches for each class are shown in Fig. 2b. The patches of D, P, E, and M were augmented using Visual Inspection with the following parameters: color-brightness = 20, color-contrast = 20, color-hue = 20, color-saturation = 20, blur = 20, sharpen = 20, noise = 20, vertical flip = TRUE, and horizontal flip = TRUE, and other patches were augmented with the following parameters: color-brightness = 20, color-contrast = 20, color-hue = 20, color-saturation = 20, vertical flip = TRUE, and horizontal flip = TRUE. These processes yielded a total of 45732 patches. The patch-level classification model for the estrous cycle stage was then generated with the following parameters: algorithm = GoogLeNet[15] and iterations = 1500.

Though the learning history from Visual Inspection suggested a sufficiently high accuracy (98%) for the resulting model (Fig. S3), the actual accuracy in practice was investigated as follows. One hundred patches with a confidence score (a number between 0 and 1 representing the confidence of prediction) >0.99 were selected from each machine-predicted D, P, E, M, and Other classes. Those 500 patches were re-labeled by one pathologist, re-filtered with range of confidence scores (0.99∼0.999). Then, the agreement and F1-score between pathologist label and machine predictions were examined (Fig. 3a, 3b, and 3c). As expected, machine predictions with higher confidence score tended to be consistent with the pathologist’s annotation. However, inevitably, the adaptation of a more stringent confidence score as a threshold resulted in decrease of the number of detected patches (Fig. 3d), making development of an aggregation model difficult. This trade-off was considered in more detail when building the aggregation model, as described in the next section.

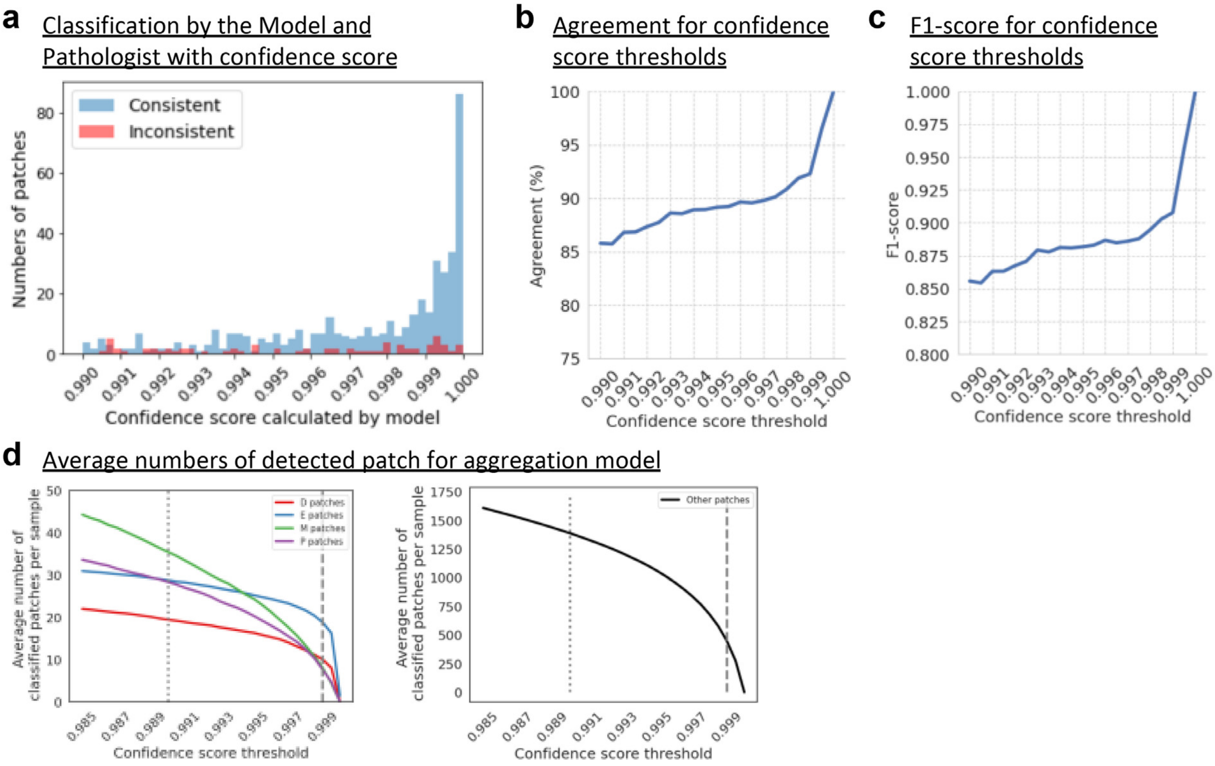
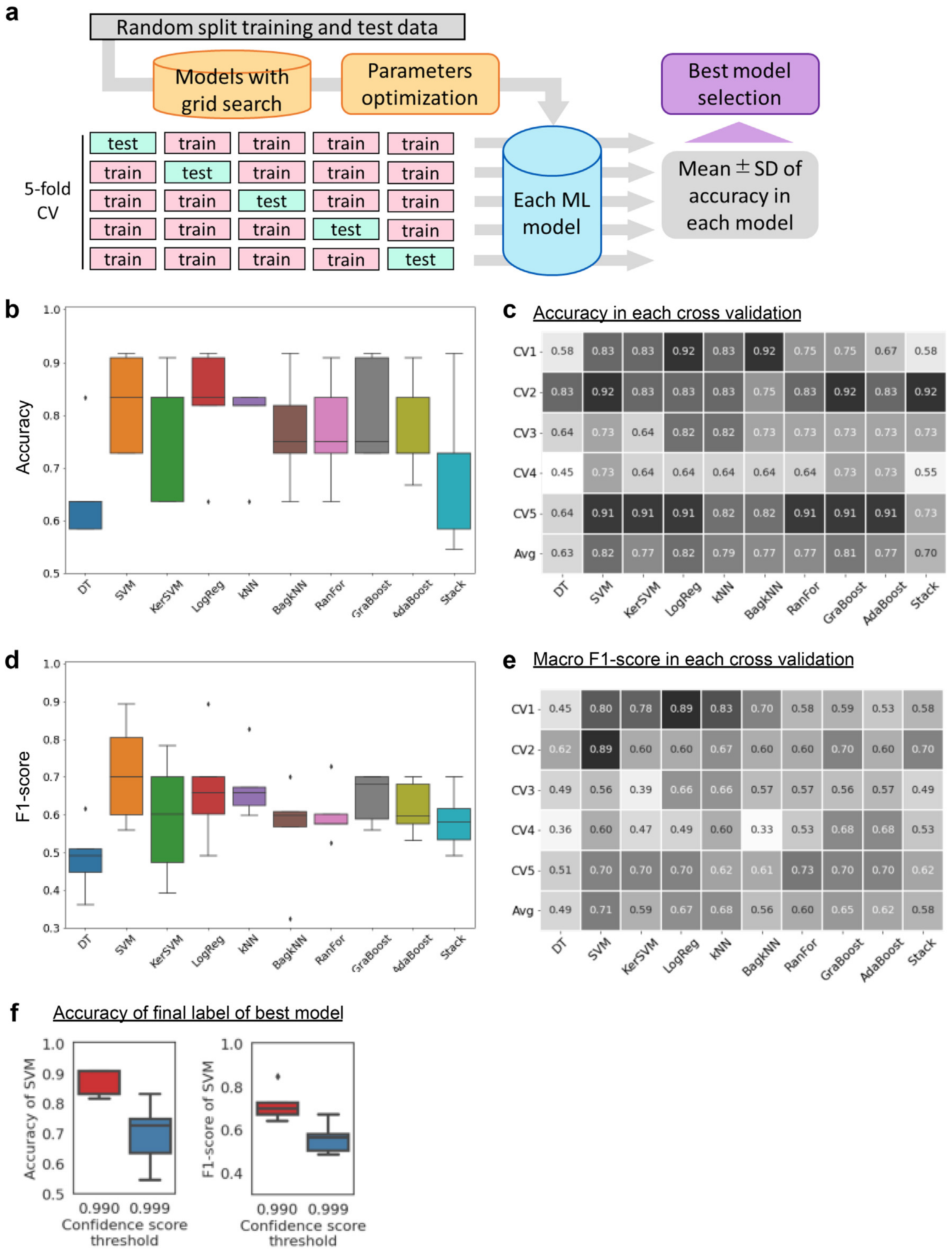


Fig. 3. Evaluation of differences between model and pathologist labels in vaginal patch classification analysis. (a) The number of consistent and inconsistent patches when the pathologist re-labeled the 500 patches determined by the classification model. x-axis: confidence score (a number between 0 and 1 representing the confidence of prediction) for each patch when Visual Inspection is applied in the classification model. y-axis: the number of patches counted. (b and c) Changes in the (b) agreement and (c) F1-score between patches classified by the model and pathologist according to the threshold of the confidence score. (d) Changes in the average number of labelled patches per sample that were detected according to the threshold of the confidence score. Dotted and dashed lines indicate a confidence score threshold of 0.990 and 0.999, respectively.

Building the aggregation model

WSIs in the training datasets were split into patches of 300 × 300 pixels with no overlap. The patch-level classification model was applied to them. In this step, patches indicative of the estrous cycle stage were detected, as shown in Fig. 2a. The number of patches for each WSI was then summarized by the predicted patch classes.



	DT	SVM	KerSVM	LogReg	kNN	BagkNN	RanFor	GraBoost	AdaBoost	Stack
CV1	0.58	0.83	0.83	0.92	0.83	0.92	0.75	0.75	0.67	0.58
CV2	0.83	0.92	0.83	0.83	0.83	0.75	0.83	0.92	0.83	0.92
CV3	0.64	0.73	0.64	0.82	0.82	0.73	0.73	0.73	0.73	0.73
CV4	0.45	0.73	0.64	0.64	0.64	0.64	0.64	0.73	0.73	0.55
CV5	0.64	0.91	0.91	0.91	0.82	0.82	0.91	0.91	0.91	0.73
Avg	0.63	0.82	0.77	0.82	0.79	0.77	0.77	0.81	0.77	0.70

	DT	SVM	KerSVM	LogReg	kNN	BagkNN	RanFor	GraBoost	AdaBoost	Stack
CV1	0.45	0.80	0.78	0.89	0.83	0.70	0.58	0.59	0.53	0.58
CV2	0.62	0.89	0.60	0.60	0.67	0.60	0.60	0.70	0.60	0.70
CV3	0.49	0.56	0.39	0.66	0.66	0.57	0.57	0.56	0.57	0.49
CV4	0.36	0.60	0.47	0.49	0.60	0.33	0.53	0.68	0.68	0.53
CV5	0.51	0.70	0.70	0.70	0.62	0.61	0.73	0.70	0.70	0.62
Avg	0.49	0.71	0.59	0.67	0.68	0.56	0.60	0.65	0.62	0.58

f
Accuracy of final label of best model
Accuracy of SVM
1.0
0.9
0.8
0.7
0.6
0.5
0.990 0.999
Confidence score threshold
F1-score of SVM
1.0
0.8
0.6
0.4
0.990 0.999
Confidence score threshold

Fig. 4. Schema for comparison of 10 conventional machine learning algorithms. (a) Overview of the optimization of each machine learning model and the selection of models to be adopted in the workflow. The hyperparameters of each model were optimized to provide the highest accuracy against the test data by grid search using randomly separated training and test data. The mean accuracy of each model to the test data by 5-fold cross validation was then compared, and the model with the highest value was selected as the algorithm to be integrated into the workflow. (b) Distribution and (c) matrix of the accuracies and their means for the test data computed by 5-fold cross validation. (d) Distribution and (e) matrix of the macro F1-scores and their means for the test data computed by 5-fold cross validation. (f) Accuracy of the prediction of the final label by the best aggregation model, SVM, when using the respective confidence score thresholds of 0.990 and 0.999. The mean accuracy of each model was then compared to the test data by 5-fold cross validation. CV, cross validation; ML, machine learning; SD, standard deviation

To build the appropriate classifier for enabling tissue-level prediction from patch counts, several parameters were considered (Fig. 4). Regarding the algorithm, 10 conventional machine learning algorithms were considered: decision tree, support vector machine (SVM), kernel SVM, linear regression, k-nearest neighbor (kNN), bagging type kNN, random forest, gradient boosting, ada boosting, and stacking model. By taking patch counts as explanatory variables and ground truth as objective variables, models were generated in a grid search manner for hyperparameters and were tested by 5-fold cross validation (Fig. 4a). Comparing their performance by accuracy and F1 score, SVM was selected as the best algorithm candidate (Fig. 4b, 4c, 4d, and 4e). Then, the influence of the selection of confidence score for patch-level classification on the accuracy of aggregation models was examined. The result indicated that an aggregation model with a confidence score >0.999 had lower accuracy than that with a confidence score >0.99 (Fig. 4f). Therefore, confidence score = 0.99 was adopted as threshold.

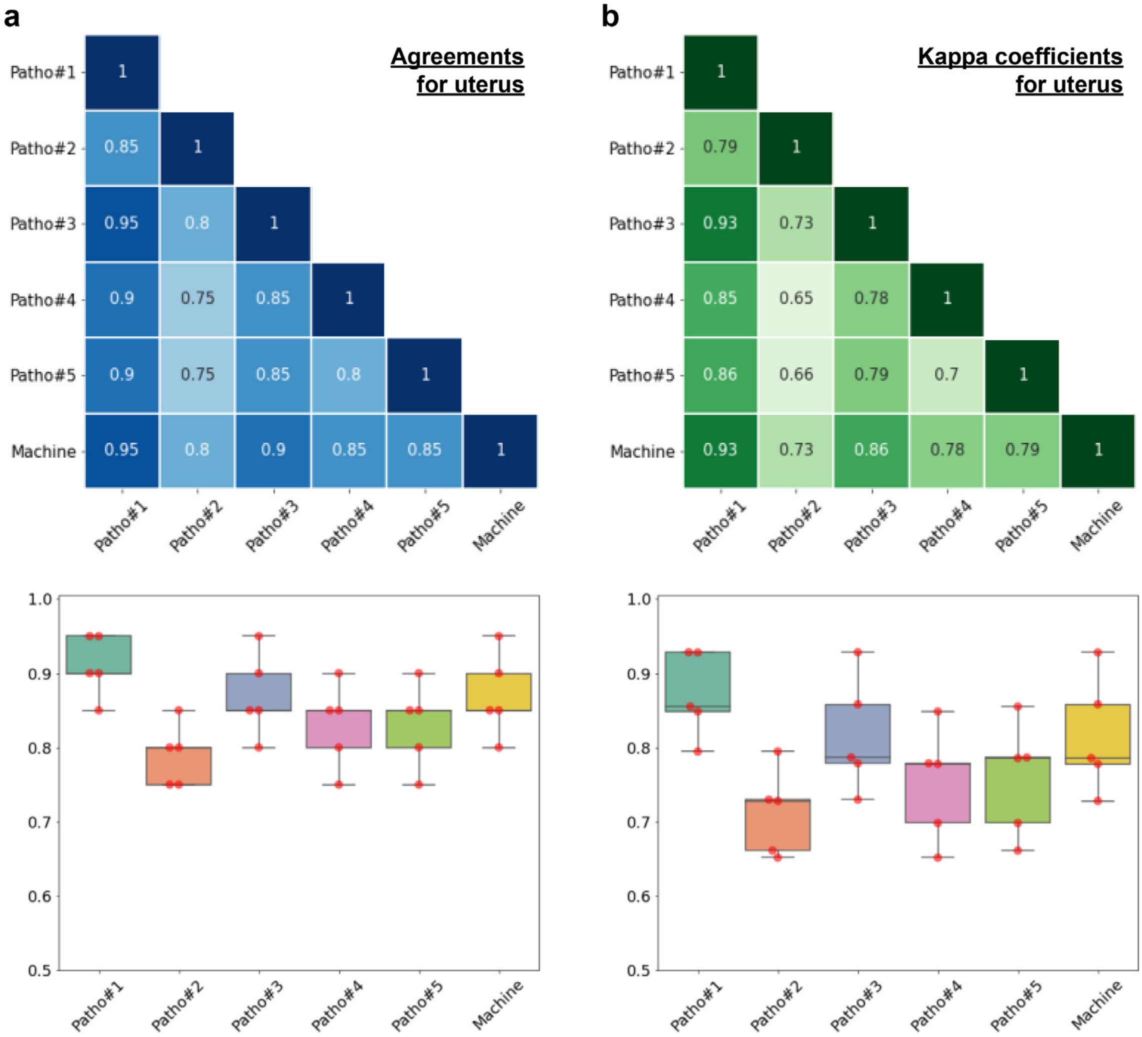
Results

Validation of model for uterine horn

The model for detecting the stage-specific uterine horn was applied to the validation dataset. When multiple objects were detected in the same WSI, the candidate with highest confidence score was adopted. The performance of the prediction was examined with the agreement coefficient and Cohen's kappa coefficient (Kcoef)[6,16] between every pair of pathologist and machine (Fig. 5a and 5b). The mean agreement of the classification for pathologist–pathologist pairs was 0.84 ± 0.06, depicting the inter-viewer variability. The mean agreement for pathologist–machine pairs was 0.87 ± 0.05. The mean Kcoef for pathologist–pathologist pairs and pathologist–machine pairs was 0.77 ± 0.08 and 0.81 ± 0.07, respectively. These matrices indicated the pathologist-level performance of our workflow for the uterine horn.

Validation of model for vagina

Using the validation dataset, the performance of the aggregation model was examined with the agreement coefficient and Kcoef for classifications between every pair of pathologist and machine (Fig. 6a and 6b). The mean agreement for pathologist–pathologist pairs was 0.77 ± 0.12. The mean agreement for pathologist–machine pairs was 0.80 ± 0.07. The mean Kcoef for pathologist–pathologist pairs and pathologist–machine pairs were 0.65 ± 0.16 and 0.69 ± 0.10, respectively. Our workflow for the vagina showed performance comparable with that of the pathologists' performance.



	Patho#1	Patho#2	Patho#3	Patho#4	Patho#5	Machine
Patho#1	1					
Patho#2	0.85	1				
Patho#3	0.95	0.8	1			
Patho#4	0.9	0.75	0.85	1		
Patho#5	0.9	0.75	0.85	0.8	1	
Machine	0.95	0.8	0.9	0.85	0.85	1

	Patho#1	Patho#2	Patho#3	Patho#4	Patho#5	Machine
Patho#1	1					
Patho#2	0.79	1				
Patho#3	0.93	0.73	1			
Patho#4	0.85	0.65	0.78	1		
Patho#5	0.86	0.66	0.79	0.7	1	
Machine	0.93	0.73	0.86	0.78	0.79	1

Fig. 5. (a) Agreement and (b) Cohen's kappa coefficient (Kcoef) for every pair of pathologist and machine for classification for the estrous cycle stage of the uterine horn. (a and b) Matrix (top) and distribution (bottom) of the (a) agreement and (b) Kcoef for every pair of pathologist and machine classification for the estrous cycle stage of the uterine horn. The Kcoef is based on the difference between the observed agreement (p_o) and the probability of chance (p_e) and is calculated as $p_o - p_e / 1 - p_e$ (see Cohen[16]). Patho, pathologist.

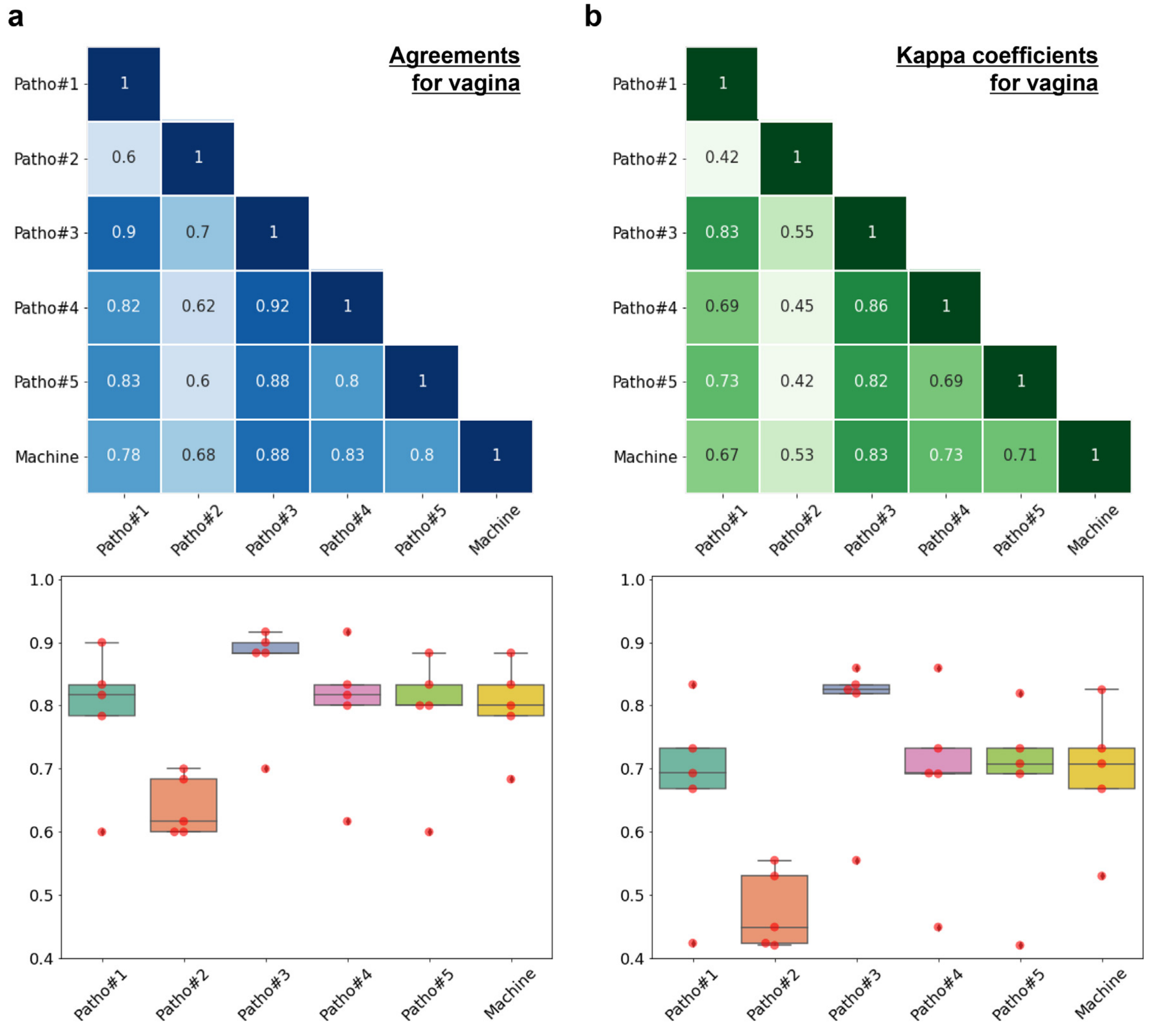


Fig. 6. (a) Agreement and (b) Cohen's kappa coefficient (Kcoef) for every pair of pathologist and machine classification for the estrous cycle stage of the vagina. (a and b) Matrix (top) and distribution (bottom) of the (a) agreement and (b) Kcoef for every pair of pathologist and machine classification for the estrous cycle stage of the vagina. The Kcoef is computed as shown in the legend of Fig. 5b. Patho, pathologist.

Discussion

We introduced digital workflows to assess the estrous cycle stage of rat uterine horns and vaginas from WSIs with attention to the following 4 considerations. First, subject WSIs are gigapixel-sized, which is overwhelming for most available image recognition algorithms. Second, subject WSIs sometimes contain other reproductive tissues in different combinations. Third, while multiple reproductive tissues from the same individual are normally governed by the same estrous cycle, each tissue must be separately assessed to detect the tissue-specific effects of drugs. Fourth, tissues sometimes gain artifacts during sample processing, exhibiting highly variable shapes on WSIs. These phenomena hamper the generation of tissue-level prediction models using entire information from WSIs. Therefore, for the uterine horns, we employed an object detection approach because the features of the estrous cycle appear at the whole-tissue level with low magnification. On the other hand, for the vagina, we adopted an approach of patch-level inference followed by aggregation of patch counts to consider the local morphologic features of the mucosa observed with high magnification in gigapixel images. These object detection and patch-level approaches were expectedly robust against variations in tissue alignment on WSIs. Although not addressed in the current study, the patch-level model would also be immune to certain artifacts such as tissue tears, shrinkage, and lack of focus, as discussed previously.[13] The mean agreement and Kcoef for pathologist–machine pairs were 0.87 and 0.81 for the uterine horn model and 0.80 and 0.69 for the vagina model, where were comparable to those for pathologist–pathologist pairs (uterine horn: 0.84 and 0.77; vagina: 0.77 and 0.65). These results may be improved by adopting alternative approaches. For example, as an aggregation method for the vagina model, thresholding and majority voting might be effective.[14] That said, the current pathologist-level results imply that our models have already nearly reached the upper limit of accuracy because pathologists' annotations showing inter-viewer variability were used as the ground truth. Therefore, our models can aid pathologists. One remaining concern is their adaptivity to unseen variations. Under controlled situations, the workflows are expected to achieve stable performance. However, their ability to deal with WSIs under other conditions (e.g., overstaining) is unknown and so would require prospective validation before implementation.

Conclusions

In this study, we have introduced digital workflows for assessing the estrous cycle stage of the uterine horns and vaginas from WSIs. Logically, these approaches can be extended to the assessment of other tissues, such as the ovary and uterine cervix, and other laboratory animals. Together, these workflows would improve efficiency in preclinical toxicology studies.

Competing interests

The authors declare that they have no competing interests.

CRediT authorship contribution statement

Shinichi Onishi: Writing – original draft. **Riku Egami:** Writing – original draft. **Yuya Nakamura:** Software. **Yoshinobu Nagashima:** Software. **Kaori Nishihara:** Methodology, Investigation. **Saori Matsuo:** Methodology, Investigation. **Atsuko Murai:** Methodology, Investigation. **Shuji Hayashi:** Methodology, Investigation. **Yoshifumi Uesumi:** Software. **Atsuhiko Kato:** Supervision. **Hiroyuki Tsunoda:** Supervision. **Masaki Yamazaki:** Writing – review & editing, Supervision. **Hideaki Mizuno:** Writing – review & editing, Supervision.

Acknowledgements

We thank Jacob Davis of Chugai Pharmaceutical Co., Ltd. for proofreading and the ASCA Corporation for their cooperation in editing this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpi.2022.100120.

References

1. Sato J, Nasu M, Tsuchitani M. Comparative histopathology of the estrous or menstrual cycle in laboratory animals. J Toxicol Pathol 2016;29(3):155–162.
2. Li S, Davis B. Evaluating rodent vaginal and uterine histology in toxicity studies. Birth Defects Res B Dev Reprod Toxicol 2007;80(3):246–252.
3. Kaur S, Benton WL, Tongkhuya SA, Lopez CMC, Uphouse L, Averitt DL. Sex differences and estrous cycle effects of peripheral serotonin-evoked rodent pain behaviors. Neuroscience 2018;384:87-100.
4. Carboni E, Marxfeld H, Tuoken H, Klukas C, Eggers T, Gröters S, et al. A workflow for the performance of the differential ovarian follicle count using deep neuronal networks. Toxicol Pathol 2021;49(4):843–850.
5. Fuchs TJ, Buhmann JM. Computational pathology: challenges and promises for tissue analysis. Comput Med Imaging Graph 2011;35(7–8):515–530.
6. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276–282.
7. Aeffner F, Adissu HA, Boyle MC, et al. Digital microscopy, image analysis, and virtual slide repository. ILAR J 2018;59(1):66–79.
8. Aeffner F, Sing T, Turner OC. Special issue on digital pathology, tissue image analysis, artificial intelligence, and machine learning: approximation of the effect of novel technologies on toxicologic pathology. Toxicol Pathol 2021;49(4):705–708.
9. Sano K, Matsuda S, Tohyama S, Komura D, Shimizu E, Sutoh C. Deep learning-based classification of the mouse estrous cycle stages. Sci Rep-Uk 2020;10(1).
10. Ren SQ, He KM, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. Adv Neur In 2015:28.
11. Bamba Y, Ogawa S, Itabashi M, et al. Object and anatomical feature recognition in surgical video images based on a convolutional neural network. Int J Comput Assist Radiol Surg 2021;16(11):2045–2054.
12. Luo SR, Kindratenko V. Hands-on with IBM visual insights. Comput Sci Eng 2020;22(5):108–112.
13. Dimitriou N, Arandjelovic O, Caie PD. Deep learning for whole slide image analysis: an overview. Front Med (Lausanne) 2019;6:264.
14. Salvi M, Acharya UR, Molinari F, Meiburger KM. The impact of pre- and post-image processing techniques on deep learning frameworks: a comprehensive review for digital pathology image analysis. Comput Biol Med 2021;128, 104129.
15. Szegedy C, Liu W, Jia YQ, et al. Going deeper with convolutions. Proc Cvpr Ieee 2015:1–9.
16. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.