MDPI

Article

Keeping Pathologists in the Loop and an Adaptive F1-Score Threshold Method for Mitosis Detection in Canine Perivascular Wall Tumours

Taranpreet Rai [1,2,*], Ambra Morisi [3], Barbara Bacci [4], Nicholas James Bacon [5], Michael J. Dark [6], Tawfik Aboellail [7], Spencer A. Thomas [8,9], Roberto M. La Ragione [3,10] and Kevin Wells [1,2]

[1] Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford GU2 7XH, UK; k.wells@surrey.ac.uk
[2] Surrey DataHub, University of Surrey, Guildford GU2 7AL, UK
[3] School of Veterinary Medicine, University of Surrey, Guildford GU2 7AL, UK; ambramorisi@hotmail.it (A.M.); r.laragione@surrey.ac.uk (R.M.L.R.)
[4] Department of Veterinary Medical Sciences, University of Bologna, 40126 Bologna, Italy; barbara.bacci@gmail.com
[5] AURA Veterinary, Guildford GU2 7AJ, UK; n.bacon@surrey.ac.uk
[6] Department of Comparative, Diagnostic and Population Medicine, College of Veterinary Medicine, University of Florida, Gainesville, FL 32611, USA; darkmich@ufl.edu
[7] Department of Diagnostic Pathology and Pathobiology, Kansas State University, Manhattan, KS 66506, USA; tawfik.aboellail@colostate.edu
[8] Department of Computer Science, University of Surrey, Guildford GU2 7XH, UK; spencer.thomas@npl.co.uk
[9] National Physical Laboratory, London TW11 0LW, UK
[10] School of Biosciences, University of Surrey, Guildford GU2 7XH, UK
* Correspondence: t.rai@surrey.ac.uk

Simple Summary: Performing a mitosis count (MC) is essential in grading canine Soft Tissue Sarcoma (cSTS) and canine Perivascular Wall Tumours (cPWTs), although it is subject to inter- and intra-observer variability. To enhance standardisation, an artificial intelligence mitosis detection approach was investigated. A two-step annotation process was utilised with a pre-trained Faster R-CNN model, refined through veterinary pathologists' reviews of false positives, and subsequently optimised using an F1-score thresholding method to maximise accuracy measures. The study achieved a best F1-score of 0.75, demonstrating competitiveness in the field of canine mitosis detection.

Abstract: Performing a mitosis count (MC) is the diagnostic task of histologically grading canine Soft Tissue Sarcoma (cSTS). However, mitosis count is subject to inter- and intra-observer variability. Deep learning models can offer a standardisation in the process of MC used to histologically grade canine Soft Tissue Sarcomas. Subsequently, the focus of this study was mitosis detection in canine Perivascular Wall Tumours (cPWTs). Generating mitosis annotations is a long and arduous process open to inter-observer variability. Therefore, by keeping pathologists in the loop, a two-step annotation process was performed where a pre-trained Faster R-CNN model was trained on initial annotations provided by veterinary pathologists. The pathologists reviewed the output false positive mitosis candidates and determined whether these were overlooked candidates, thus updating the dataset. Faster R-CNN was then trained on this updated dataset. An optimal decision threshold was applied to maximise the F1-score predetermined using the validation set and produced our best F1-score of 0.75, which is competitive with the state of the art in the canine mitosis domain.

Keywords: artificial intelligence; deep learning; canine Soft Tissue Sarcoma; canine Perivascular Wall Tumour; digital pathology; object detection; faster R-CNN; mitosis; mitosis detection; pathologists in the loop; humans in the loop

1. Introduction

Canine Soft Tissue Sarcoma (cSTS) is a heterogeneous group of mesenchymal neoplasms (tumours) that arise in connective tissue [1–6]. cSTS is more prevalent in middle-age to older and medium to large-sized breeds with the median reported age of diagnosis between 10 and 11 years old [3,7–10]. The anatomical site of cSTS can vary considerably, but it is mostly found in the cutaneous and subcutaneous tissues [9]. In human Soft Tissue Sarcoma (STS), histological grade is an important prognostic factor and one of the most validated criteria to predict outcome following surgery in canines [10–13]. General treatment consists of surgically removing these cutaneous and subcutaneous sarcomas. Nevertheless, it is the higher-grade tumours that can be problematic, as their aggressiveness can reduce treatment options and result in a poorer prognosis. The focus of this study was on one common subtype found in dogs: canine Perivascular Wall Tumours (cPWTs). Canine Perivascular Wall Tumours (cPWTs) arise from vascular mural cells and are often recognisable from their vascular growth patterns [14,15].

The scoring for cSTS grading is broken down into three major criteria: the mitotic count, differentiation and the level of necrosis [9]. Mitosis counting can be exposed to high inter-observer variability [16], depending on the expertise of the pathologist; however, the counting of mitotic figures is considered the most objective factor in comparison to tumour necrosis and cellular differentiation when grading cSTS [16]. It is routine practise to investigate mitosis using $40\times$ magnification; however, manual investigation at such high-powered fields (HPFs) is a laborious task that is prone to error, thus leading to the previously discussed inter-observer variability phenomenon.

For the purposes of this study, the focus was on creating a mitosis detection model as it is a significant criterion from the cSTS histological grading system [13] where the density of mitotic figures is also considered highly correlated with tumour proliferation [17]. Mitosis detection has been pursued in the computer vision domain since the 1980s [18]. Before 2010, relatively few studies aimed to automate mitosis detection [19–21]. However, since the MITOS 2012 challenge [22], there has been a resurgence of interest. Mitosis detection can often be considered as an object detection problem [23]. Rather than categorising entire images as in image classification tasks, object detection algorithms present object categories inside the image along with an axis-aligned bounding box, which in turn indicates the position and scale of each instance of the object category. In the case of mitosis detection, the considered objects are mitotic figures. As a result, several approaches have used object detection-related algorithms for mitosis detection. An example of an object detection algorithm is the regions-based convolutional neural network (R-CNN) [24]. At first, a selective search is performed on the input image to propose candidate regions, and then the CNN is used for feature extraction. These feature vectors are used for training in bounding box regression. There have been many developments on this type of architecture such as Fast R-CNN [25] and Faster R-CNN [26], which is the primary object detection model used in this work. One set of authors detected mitosis using a variant of the Faster R-CNN (MITOS-RCNN), achieving an F-measure score of 0.955 [27].

Several challenges have been held in order to find novel and improved approaches for mitosis detection [17,22,23,28,29]. Some of these challenges and research on mitosis detection methods have also been conducted using tissue from the canine domain [30–33].

It was made apparent by the collaborating pathologists that AI approaches for grading tasks in cSTS were desirable, and so this study aims to tackle one criterion, which is to develop methods for mitosis detection in a subtype of cSTS: cPWT. To the best of our knowledge, this is the first work in the automated detection of mitoses in cPWTs.

2. Materials and Methods

2.1. Data Description and Annotation Process

A set of canine Perivascular Wall Tumour (cPWT) slides were obtained from the Department of Microbiology, Immunology and Pathology, Colorado State University. A senior veterinary pathologist at the University of Surrey confirmed the grade of each case (patient)

and chose a representative histological slide for each patient. These histological slides were digitised using a Hamamatsu NDP Nanozoomer 2.0 HT slide scanner. A digital Whole Slide Image (WSI) was created via scanning under $40\times$ magnification (0.23 µm/pixel) with a scanning speed of approximately 150 s at $40\times$ mode (15 mm $\times$ 15 mm).

Veterinary pathologists independently annotated the WSIs for mitosis using the open-source Automated Slide Analysis Platform (ASAP) software (https://www.computationalpathologygroup.eu/software/asap/, accessed on 28 January 2024) [34]. The pathologists used different magnifications (ranging from $10\times$ to $40\times$) to analyse the mitosis before creating mitosis annotations. These annotations were centroid coordinates, which were centered on the suspecting mitotic candidate. Centroid coordinate annotations can be considered as weak annotations as they are simply coordinates placed in the centre of a mitotic figure and not fine-grained pixel-wise annotations around the mitosis. In order to categorise a mitotic figure, both pathologist annotators needed to form an agreement on the mitotic candidate. As these were centroid coordinates, an agreement was determined when two independent centroid annotations from each annotator were overlaid on one another. Any centroid annotations without agreement were dismissed from being considered as a mitotic figure. Table 1 shows the differences between the two annotators for both training and validation when counting mitotic figures in our cPWT dataset.

Table 1. The differences between the two annotators in regard to mitosis annotations for the training/validation set. The "Slide" column represents the anonymised set of slides annotated. "Anno 1" and "Anno 2" show the number of mitoses annotated per slide for each annotator. "Agreement" represents the number of agreed mitoses between each annotator. The "% agreement" for each annotator represents the percentage of the agreed mitotic count against the respective annotators mitotic count. "Avg" is the average of every WSI % agreement, which is computed for each annotator.

Slide	Anno 1	Anno 2	Agreement	% Agreement Anno 1	% Agreement Anno 2
F17-04773	31	31	23	74.19	74.19
F17-03141	69	89	55	79.71	61.80
F17-1261	45	46	41	91.11	89.13
F18-13364	695	517	444	63.88	85.88
F17-02232	331	264	218	65.86	82.58
F17-04911	49	58	37	75.51	63.79
F17-0549	157	142	112	71.34	78.87
F17-011577	27	29	23	85.19	79.31
F17-011777	449	367	290	64.59	79.02
F17-03855	97	87	70	72.16	80.46
F17-04900	91	86	75	82.42	87.21
F18-7832	496	401	346	69.76	86.28
F17-09700	202	187	139	68.81	74.33
F17-02641	59	48	43	72.88	89.58
F17-09926	77	71	62	80.52	87.32
F17-02723	49	52	40	81.63	76.92
F17-05935	55	46	44	80.00	95.65
F17-02120	58	53	43	74.14	81.13
F18-79705	132	99	87	65.91	87.88
Total:	3169	2673	2192	Avg: 74.72	Avg: 81.12

For patch extraction, downsized binary image masks (by a factor of 32) were generated, depicting tissue from the biopsy samples against background slide glass. A tissue threshold of 0.75 was applied to 512×512 patches for final patch extraction. Therefore, if a patch contained less than 75% of any tissue, it was dismissed from the dataset. This was to ensure that the patches contained relevant information for mitosis object detection.

The test set consisted of patches extracted from high-powered fields (HPFs) determined by the pathologist annotators. To replicate real-world test data, our collaborating pathologists selected 10 continuous non-overlapping HPFs from each WSI. The size of this

area was determined by loosely following the Elston and Ellis [35] criteria of an area size of 2.0 mm^2. For 20× magnification (level 1 in the WSI pyramid), the width of the 10 HPFs was 4096 pixels and the height was 2560 pixels. This produced 40 non-overlapping patches of 512 × 512 pixels, thus producing a dataset of 440 patch images from the 11 hold-out test WSIs at 20× magnification. Only patches containing mitosis were used for training and validation, whereas for testing, all extracted patches were evaluated. Details on the number of mitosis per slide in training/validation and test sets are provided in Appendix Tables A1 and A2, respectively. Details on the number of patches used for training/validation and testing for 40× magnification is provided in Appendix Table A3. Details on the number of patches used for training/validation and testing for 20× magnification is provided in Appendix Table A4.

2.2. Object Detection and Keeping the Pathologist in the Loop for Dataset Refinement

Mitosis detection is generally considered an object detection problem [23]; For this study, we used a Faster R-CNN model [26]. We initialised a Faster R-CNN model with pre-trained COCO [36] weights with the ResNet-50 head pre-trained on ImageNet. The model was fine-tuned, updating all parameters of the model using our dataset. Preliminary experiments suggested using a learning rate of 0.01 and SGD to be used as the optimiser. A batch size of 4 was also used for these experiments. Training was implemented for 30 epochs, where the the model with the lowest validation loss was saved for final evaluation. Faster R-CNN is jointly trained with four different losses; two for the RPN and two for the Fast R-CNN. These losses are RPN classification loss (for distinguishing between foreground and background), RPN regression loss (for determining differences between the regression of the foreground bounding box and ground truth bounding box), the Fast R-CNN classification loss (for object classes) and Fast R-CNN bounding box regression (used to refine the bounding box coordinates). Therefore, in our implementation of determining the lowest validation loss, at every epoch, each loss type was considered equally. We implemented 3-fold cross-validation at the patient (WSI) level to test the veracity and robustness of our approach with the training data split into three folds for training and validation. We also used an unseen hold-out test set for final evaluation and for a fair comparison of all three folds. The training, validation and hold-out test splits for each fold are depicted in Appendix Table A5.

Furthermore, as most mitotic figures from the same tissue type are generally of a similar size (dependent on the stage of mitosis, staining techniques, and slide quality), we opted to use the default anchor generator sizes provided by the PyTorch implementation of Faster R-CNN. These sizes were 32, 64, 128, 256 and 512 with aspect ratios of 0.5, 1.0 and 2.0. See Figure 1 for a depiction of the Faster R-CNN applied to the cPWT mitosis detection problem.
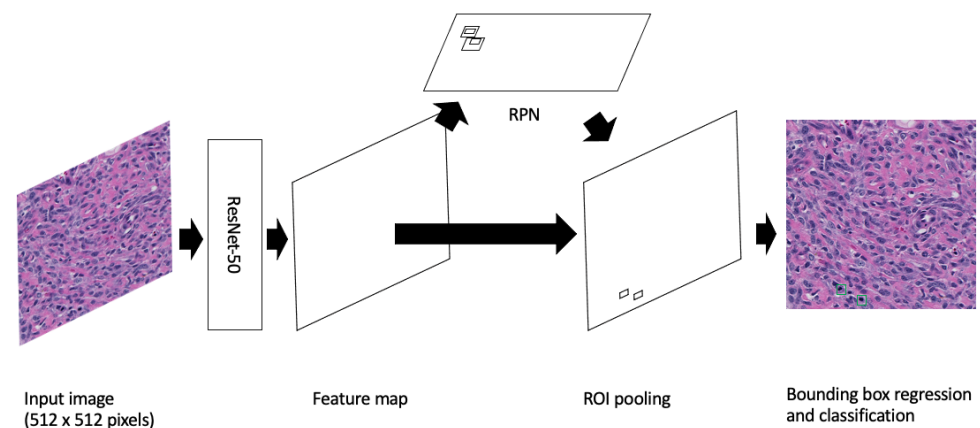
Figure 1. Image is inspired by Mahmood et al.'s depiction of Faster R-CNN [37]. A Faster R-CNN object detection model applied to the cPWT mitosis dataset. An input image of size 512×512 pixels is passed through the model where the feature map is extracted using the Resnet-50 feature-extraction network. This is then followed by generating region proposals in the Region Proposal Network (RPN) and finally mitosis detection in the classifier.

During the evaluation inference, non-maximum suppression (NMS) with an IoU value of 0.1 was applied as a post-processing step to remove low-scoring otherwise redundant overlapping bounding boxes. This post-processing method is also consistent with other mitosis detection methods in the literature [38,39].

In object detection, mean average precision (mAP) is typically used to evaluate the performance of a model depending on the task or dataset [40–43]. However, we opted to use the F1-score in order to compare our results to mitosis detection approaches in the literature. The F1-score was computed globally for each fold; thus, it was applied and determined for the entire dataset of interest. True positive (TP) detections were computed if there was an IoU of >= 0.5 between the ground truth and proposed candidate detections. Anything that did not meet the IoU threshold was considered a false positive (FP) detection. Any missed ground truth detections were considered false negatives (FNs). As a result, we could also generate the F1-score. The F1-score can be considered the harmonic mean between the precision and recall (sensitivity). Both precision (Equation (1)) and sensitivity (Equation (2)) contribute equally to the F1-score (Equation (3)):

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$F1 = 2 * \frac{Sensitivity * Precision}{Sensitivity + Precision} \tag{3}$$

where TP, FP and FN are true positives, false positives and false negatives, respectively.

The models were implemented in Python, using the PyTorch deep learning framework. The hardware and resources available for implementation used a Dell T630 system, which included 2 Intel Xeon E5 v4 series 8-Core CPUs with 3.2 GHz, 128 GB of RAM (Dell Corporation Limited, London, UK), and 4 nVidia Titan X (Pascal, Compute 6.1, Single Precision) GPUs.

The mitosis annotation process is an exhaustive and arduous process, and thus the initial annotation process may be suboptimal due to the vast area annotators needed to examine mitotic candidates. Taking inspiration from Bertram et al. [33], we used our deep

learning object detection models from these experiments to refine the dataset (see Figure 2). We hypothesised that many of the FP candidates may have been incorrectly labelled. Our collaborating pathologists reviewed all the FP candidates (irregardless of class score) from each validation fold and the hold-out test set and determined which candidates were mislabeled. As a result, we were able to formulate additional ground truth mitoses for use in the final set of experiments.
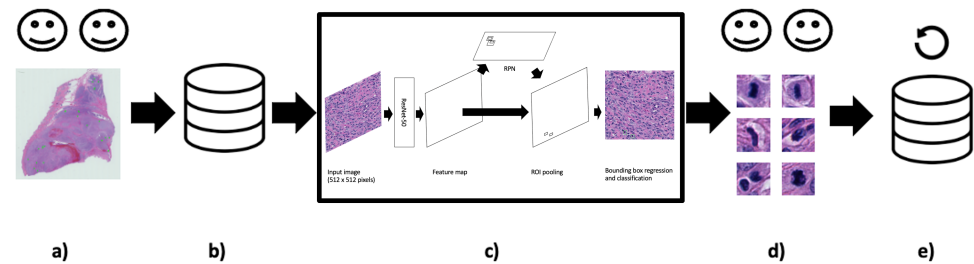
Figure 2. Keeping humans in the loop: (**a**) Two pathologist annotators independently review canine Perivascular Wall Tumour (cPWT) Whole Slide Images (WSIs) and applied centroid annotations to mitotic figures. (**b**) After initial agreement of mitoses, this formed the initial dataset of patch images with annotations. (**c**) A Faster R-CNN object detector was trained and tested on these data. (**d**) Thereafter, false positive (FP) candidates are reviewed again by the pathologist annotators where misclassified candidates are reassigned as true positives (TPs). (**e**) These new TPs are added to the updated dataset. (20× magnification images).

2.3. Adaptive F1-Score Threshold

For this method, the Faster R-CNN object detector was trained on detecting mitotic candidates using the refined (updated) dataset. The same training hyperparameters as described earlier were applied; however, we lowered the number of epochs. It was observed that the models found their optimal validation loss by epoch 7 across all three folds in the initial experiment runs. Therefore, to ensure optimality, we chose 12 epochs for training, again using the lowest validation loss as determining the "best" model. The trained Faster R-CNN model outputs potential mitosis candidates, but it also outputs probability scores relating to the strength of the object prediction. These scores ranged from 0 to 1, where 1 would highlight the model is 100% certain that the candidate is mitosis and 0.01 would describe a prediction that is very low in confidence. We optimised our models based on the F1-score [44–46]. The probability thresholds t ranged from 0.01 to 1, and so choosing the optimal threshold T for the F1-score $F1$ can be represented formally as:

$$T = \arg\max_{t} F1(t) \qquad (4)$$

We determined the optimal F1-score threshold value using the validation set and applied this threshold value to the final evaluation on the hold-out test set. Figure 3 demonstrates the entire workflow of this method from the creation of the updated mitosis dataset where the pathologists reviewed all the FP candidates all the way to the adaptive F1-score thresholds applied to the mitosis candidate predictions.
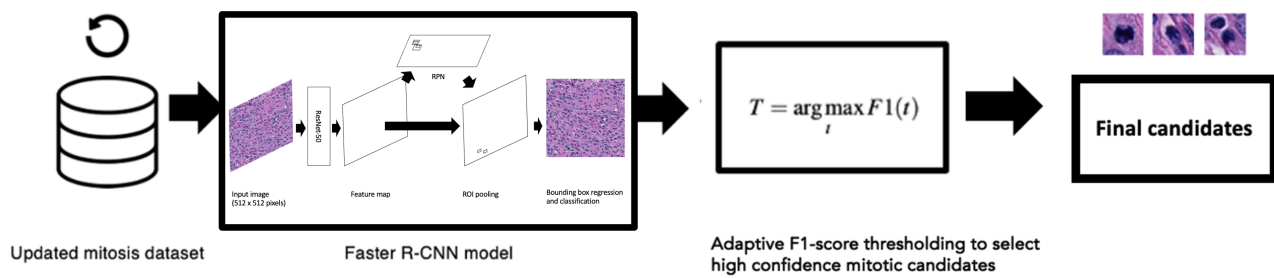
Figure 3. We used 20× magnification images and annotations from the updated mitosis dataset to train the Faster R-CNN object detection model (details from the Faster R-CNN model are also shown in Figure 1). Optimal thresholds using Equation (4) were applied on the output candidates determined from the validation set.

3. Results

The pathologists-in-the-loop approach for dataset refinement was first applied as demonstrated by Figure 2. In a preliminary investigation, two magnifications (40× and 20×) were used to determine the best resolution for our for our task (see Table 2) .

Tables A6 and A7 show the differences in mitotic candidate numbers before and after refinement (second review) for the training/validation and test sets, respectively. The first set of results from the optimised Faster R-CNN approach is depicted in Table 3. This shows a comparison of performance of the Faster R-CNN trained on the initial mitosis dataset and the updated refined mitosis dataset. It is apparent that sensitivities have improved for all folds when using the updated refined dataset; however, in some cases, such as in fold-1 validation, fold-3 validation and fold-3 test, we can see that the F1-score is lower due to a decrease in precision scores. This could be due to the updated refined dataset containing more difficult examples for the effective mitosis object detection training. The previous initial dataset may have contained more obvious mitosis examples and thus was predicting detections that closely resembled these obvious examples. Table 4 shows the Faster R-CNN results before and after F1-score thresholding was applied on the models trained using the updated mitosis dataset. The thresholds were predetermined on the validation set for each fold using Equation (4) (see Figure 4). When applying the optimal thresholds, we saw large improvements in the F1-score, which were largely due to an improvement in precision because of a reduction in FPs. This was seen on the test set with an F1-score of 0.402 to 0.750. However, this increase in precision came at the expense of some sensitivity across all three folds, where for example on the test set the mean sensitivity for all three folds reduced from 0.952 to 0.803. Nevertheless, the depreciation in sensitivity does not offset the increase in precision, where sensitivity decreased by 14.9 % and precision increased by 45.2 %. This suggests that the majority of TP detections prior to the adaptive F1-score thresholding are of a high probability confidence compared to the FP detections.

Table 2. Initial mitosis object detection results for the 40× and 20× magnification patches datasets. As the difference in performance between the two resolution datasets was of interest, we first present the initial results for 20× and 40 magnifications for validation and test sets and for all three folds. Interestingly, although the 40× magnification trained models appeared to produce better F1-scores for validation, 20× magnification models performed better across all three folds when applied to the hold-out test set. It appears that with our experimental set-up, the models trained on 20× magnification generalise across the two evaluation datasets better. As a consequence, and to also reduce computational requirements, we proceeded further with the 20× magnification extracted dataset. Results for these initial experiments also suggested that the object detector was highly sensitive for the test set (at a mean average of 0.918) but not as precise (at a mean average of 0.249 for the precision measure).

Magnification	Fold	Set	Sensitivity	Precision	F1-Score	TP	FP	FN
40×	1	Val	0.967	0.720	0.826	590	229	20
40×	1	Test	0.957	0.132	0.232	135	890	6
40×	2	Val	0.922	0.786	0.849	847	230	72
40×	2	Test	0.965	0.173	0.294	136	649	5
40×	3	Val	0.944	0.724	0.819	503	192	30
40×	3	Test	0.957	0.185	0.311	135	593	6
20×	1	Val	0.957	0.484	0.643	582	620	26
20×	1	Test	0.932	0.207	0.338	137	526	10
20×	2	Val	0.895	0.567	0.694	810	619	95
20×	2	Test	0.918	0.221	0.356	135	477	12
20×	3	Val	0.897	0.545	0.678	477	399	55
20×	3	Test	0.905	0.320	0.473	133	282	14

Table 3. A comparison of results of the models trained on the initial annotated dataset and the updated dataset. Results are shown for both the validation and test sets for folds 1, 2 and 3.

Fold	Data	Set	Sensitivity	Precision	F1-Score	TP	FP	FN
1	Initial	Val	0.957	0.484	0.643	582	620	26
1	Updated	Val	0.961	0.452	0.615	610	740	25
1	Initial	Test	0.932	0.207	0.338	137	526	10
1	Updated	Test	0.954	0.239	0.383	187	594	9
2	Initial	Val	0.895	0.567	0.694	810	619	95
2	Updated	Val	0.919	0.557	0.694	877	698	77
2	Initial	Test	0.918	0.221	0.356	135	477	12
2	Updated	Test	0.959	0.281	0.435	188	480	8
3	Initial	Val	0.897	0.545	0.678	477	399	55
3	Updated	Val	0.935	0.398	0.558	528	798	37
3	Initial	Test	0.905	0.320	0.473	133	282	14
3	Updated	Test	0.944	0.244	0.387	185	574	11
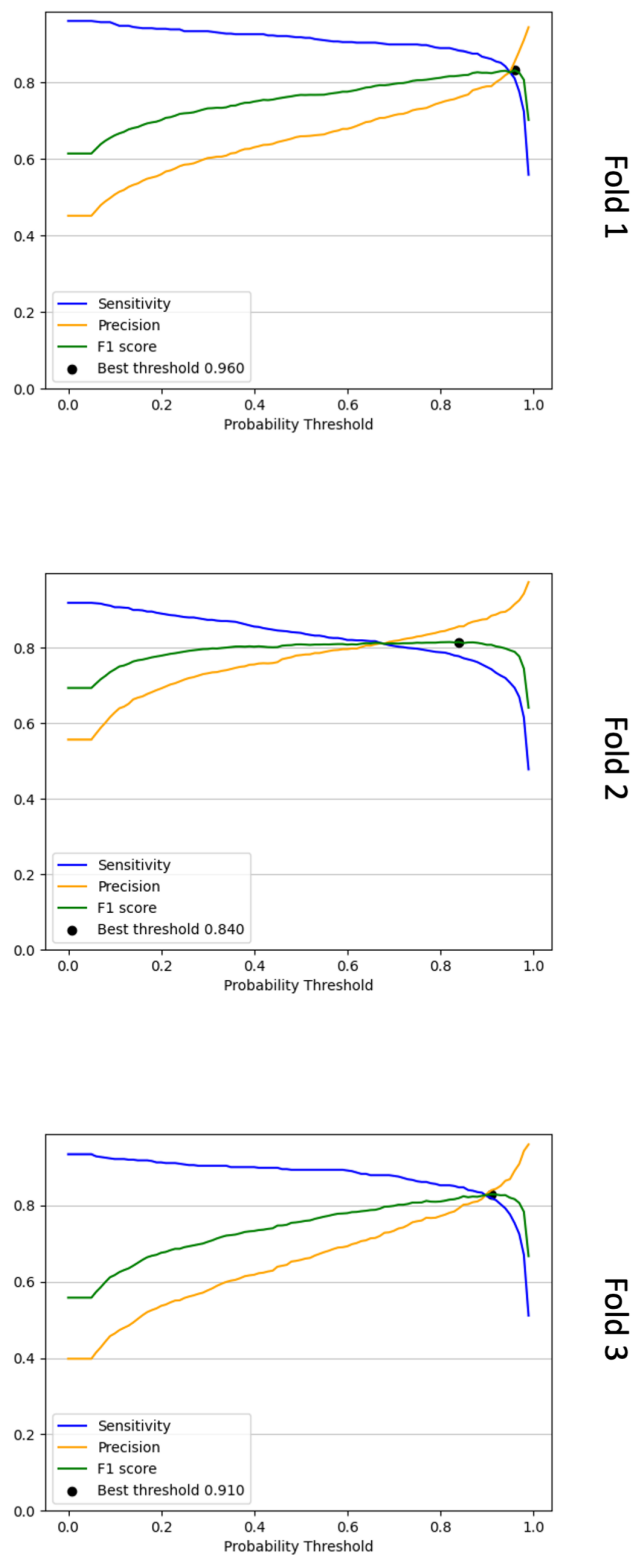
Figure 4. Line graphs that show the sensitivity, precision and F1-score calculated for each probability threshold for the three validation folds. To determine the optimal probability threshold, we choose the threshold with the highest F1-score as determined via Equation (4). In the above plots, these are denoted as "best threshold". For fold 1, this threshold was 0.96, for fold 2, it was 0.84, and for fold 3, it was 0.91.

Table 4. Results of the models trained on the updated dataset. The "Thresholds" column depict whether models were optimised using the adaptive F1-score threshold metric described in Equation (4); filled in values state the probability threshold. It is apparent that the models with optimised thresholds produced the highest F1-scores across all folds, producing a mean average F1-score of 0.750 on the test set compared to 0.402.

| Fold	Threshold	Set	Sensitivity	Precision	F1-Score
1	None	Val	0.961	0.452	0.615
1	0.96	Val	0.811	0.854	0.832
1	None	Test	0.954	0.239	0.383
1	0.96	Test	0.776	0.756	0.766
2	None	Val	0.919	0.557	0.694
2	0.84	Val	0.778	0.857	0.815
2	None	Test	0.959	0.281	0.435
2	0.84	Test	0.827	0.633	0.717
3	None	Val	0.935	0.398	0.558
3	0.91	Val	0.819	0.840	0.830
3	None	Test	0.944	0.244	0.387
3	0.91	Test	0.806	0.731	0.767
Average (mean)	None	Val	0.938	0.469	0.622
	None	Test	0.952	0.255	0.402
	Optimised	Val	0.803	0.850	0.826
	Optimised	Test	0.803	0.707	0.750

4. Discussion

This study has demonstrated a method for mitosis detection in cPWT WSIs using a Faster R-CNN object detection model, an adaptive F1-score thresholding feature on output probabilities and the refinement of a mitotic figures dataset by keeping pathologists in the loop.

Many approaches in the literature use the highest resolution images for their object detection methods (typically at 40× objective); however, we preliminarily found that 20× magnification was beneficial for our task and the dataset provided, as shown in Table 2. Nevertheless, this warrants a further investigation and additional discussions with the collaborating pathologists, who may provide reasoning as to why certain candidates were classed as mitosis at different resolutions.

Initially, solely using the outputs from a Faster R-CNN model produced promising results generating high sensitivities; however, these outputs required further post-processing to improve precision. Applying adaptive F1-score thresholds, where the optimal values were predetermined on the validation set and applied to the test set, demonstrated an effective method of reducing the number of FP predictions. This ultimately resulted in dramatically increasing the F1-score due to a stark increase in precision. However, this came at a small expense of sensitivity. Nevertheless, the rate of change of the sensitivity and the precision are not equal with the latter vastly improving. This suggests that the majority of FP detections are of lower probability confidence compared to TP detections.

Multi-stage (typically dual-stage) approaches have also become increasingly prevalent over the years where they typically take the form of selecting mitotic candidates in the first stage and then apply another classifier in the second stage [32,33,47–49]. Although not reflected in the main findings of this study, we attempted to use a second-stage classifier (Figure A1) on mitotic candidates to classify between TP and hard FPs to no avail (see results of the two-stage approach in Table A8 and its subsequent ROC curves in Figure A2). Most machine learning methods require large datasets for effective training, which in this case was not available once optimisation was applied using the adaptive F1-score threshold method. One could train models using the non-thresholded detections; however, this would result in a model that is able to distinguish between true positive mitosis and

mostly obvious FP candidates. By applying the adaptive F1-score thresholding method, we constrained the dataset and attempted to learn differences between TP and high confidence hard false positive detections, but we did not provide an adequately large dataset for training. Figure 5 depicts a 512 × 512 pixel image in the test set, highlighting FN and FP detection.
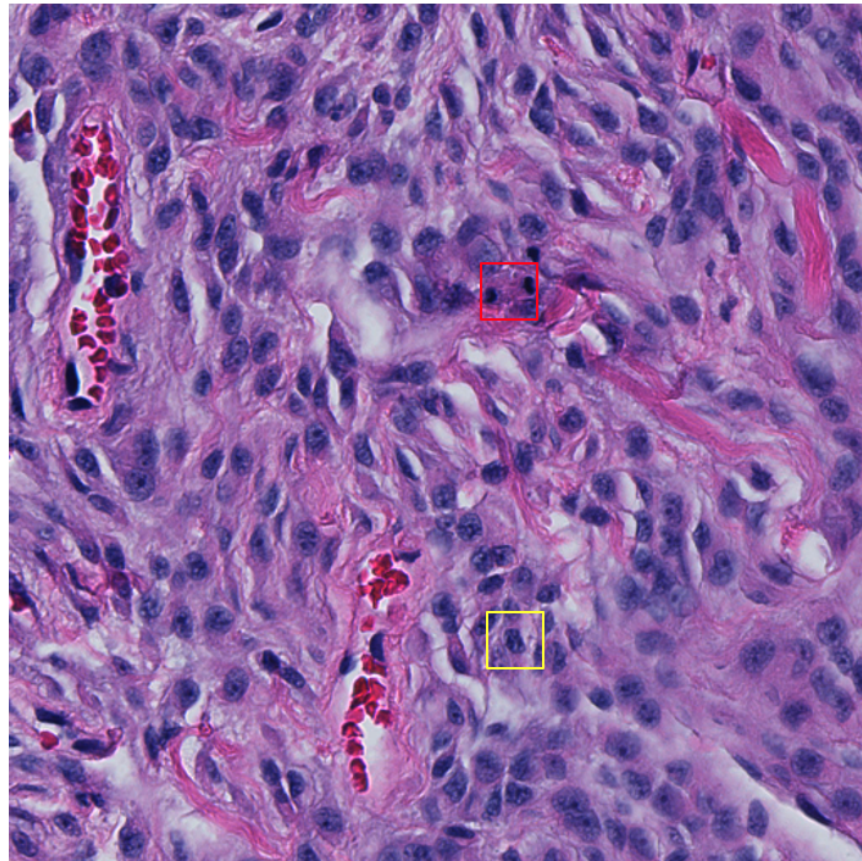
Figure 5. An example 512 × 512 pixel image from the test set with a false negative (FN) shown in the red bounding box and a false positive (FP) detection shown in the yellow bounding box (32 × 32 pixels). The FP detection provides a probability confidence score of 5.3% and so would typically be dismissed as a mitosis candidate once the adaptive F1-score threshold is applied.

Different phases and other biological phenomenon could influence the size of the mitosis region of interest. Going forward, it may also be worth labelling mitosis in regard to the phases and thus creating a multi-class problem rather than binary, as shown in this study. As a consequence, the size of the ground truth bounding boxes could also be varied depending on the target phase being classified. Nonetheless, the models were still able to predict the vast majority of mitosis in these phases.

It must be further denoted that the methodology is applied to only patches from HPFs containing mitosis that were annotated by the collaborating pathologists. Therefore, we propose expanding our dataset to include a broader range of sections, including those not initially marked by pathologists, to evaluate and enhance our model's generalisability. The data should include labels for areas containing tumour and non-tumour tissue to fully consider the overall impact of this mitosis detection method.

Our focus for this study is on cPWT; however, we could potentially adapt this method to other cSTS subtypes as well as to other tumour types. An additional study might explore the application of cPWT-trained models to different cSTS subtypes to assess if comparable outcomes are achieved. Nevertheless, given that tumour types from various domains exhibit unique challenges due to their specific histological characteristics, it may

be necessary to train or fine-tune models using tumour-specific datasets to evaluate the efficacy of this approach.

While our F1-score demonstrates competitive performance for detecting mitosis in the canine domain, the clinical relevance and applicability of this metric should be taken into account. Future work should focus on employing this method as a supportive tool, assessing its practical effectiveness and reliability in a veterinary clinical setting.

To conclude, by using our experimental set-up, the optimised Faster R-CNN model was a suitable method for determining mitosis in cPWT WSIs. To the best of our knowledge, this is the first mitosis detection model applied solely on cPWT data, and thus we consider this a baseline three-fold cross-validation mean F1-score of 0.750 for mitosis detection in cPWT.

Abbreviations

The following abbreviations are used in this manuscript:

WSI	Whole Slide Images
cPWT	Canine Perivascular Wall Tumours
cSTS	Canine Soft Tissue Sarcoma
HPF	High-Powered Fields
FP	False Positive
TP	True Positive
TN	True Negative
FN	False Negative
ROC	Receiver Operating Characteristic
CNN	Convolutional Neural Network
R-CNN	Region-Based Convolutional Neural Network
ASAP	Automated Slide Analysis Platform
NMS	Non-Maximum Suppression
IoU	Intersection over Union
mAP	Mean Average Precision
RPN	Region Proposal Network

Appendix A

Appendix A.1

Table A1. Two WSI magnification resolutions (40× and 20×) were initially investigated for determining a suitable resolution for mitosis detection using our cPWT dataset. Therefore, two separate datasets of the two resolutions were extracted. The 10 HPF size at 40× magnification (level 0 of the WSI pyramid) resulted in a width of 7680 pixels and height of 5120 pixels. In terms of physical distance, this is a width of 1.753 mm and height of 1.169 mm. When rounded to 1 decimal place, this approximately represents an aspect ratio of 3:2. When extracting 512 × 512 pixels from this area of interest, we ended up with 150 patches. This produced 150 non-overlapping patches of 512 × 512 pixels, producing a test dataset of 1650 patch images from 11 hold-out test WSIs. The details for the 20× dataset are in text. Presented below are the number of mitosis annotations per Whole Slide Image (WSI) for both 40× and 20× magnifications in the training/validation set.

Slide	Agreement after Threshold (40×)	Agreement after Threshold (20×)
F17-04773	21	18
F17-03141	35	38
F17-1261	41	41
F18-13364	437	437
F17-02232	215	216
F17-04911	37	37
F17-0549	110	106
F17-011577	23	23
F17-011777	217	212
F17-03855	69	68
F17-04900	75	75
F18-7832	335	331
F17-09700	138	134
F17-02641	40	41
F17-09926	61	62
F17-02723	38	39
F17-05935	42	41
F17-02120	43	42
F18-79705	85	84
Total:	2062	2045

Table A2. The number of mitosis annotations in 10 continuous high-powered fields (HPFs) from each Whole Slide Image (WSI) for both 40× and 20× magnifications in the hold-out test set.

Slide	Agreement after Threshold (40×)	Agreement after Threshold (20×)
F17-06348	12	13
F17-010348	1	1
F17-011490	2	2
F19-03615	54	58
F17-05256	25	35
F17-08570	1	1
F19-7408	3	3
F18-2508	10	11
F17-07510	17	17
F17-08031	5	5
F17-0260	1	1
Total:	131	147

Table A3. The number of patches per Whole Slide Image (WSI) in the train/validation and test sets for patches extracted from level 0 (40× magnification) of the WSI.

Set	Slide	No. of Patches
Train/Val	F18-7832	305
Train/Val	F17-02232	208
Train/Val	F17-011777	206
Train/Val	F17-09700	131
Train/Val	F17-0549	105
Train/Val	F18-79705	81
Train/Val	F17-03855	68
Train/Val	F17-09926	60
Train/Val	F17-1261	41
Train/Val	F17-02120	43
Train/Val	F17-03141	35
Train/Val	F17-05935	37
Train/Val	F17-02723	37
Train/Val	F17-011577	21
Train/Val	F18-13364	401
Train/Val	F17-04900	72
Train/Val	F17-02641	40
Train/Val	F17-04911	37
Train/Val	F17-04773	21
	Total	1949
Test	F17-05256	150
Test	F17-02600	150
Test	F17-07510	150
Test	F17-08031	150
Test	F17-011490	150
Test	F17-006348	150
Test	F19-03615	150
Test	F17-010348	150
Test	F18-2508	150
Test	F17-08570	150
Test	F19-7408	150
	Total	1650

Table A4. The number of patches per Whole Slide Image (WSI) in the train/validation and test sets for patches extracted from level 1 (20× magnification) of the WSI.

Set	Slide	No. of Patches
Train/Val	F18-7832	251
Train/Val	F17-02232	189
Train/Val	F17-011777	179
Train/Val	F17-09700	125
Train/Val	F17-0549	97
Train/Val	F18-79705	76
Train/Val	F17-03855	66
Train/Val	F17-09926	57
Train/Val	F17-1261	40
Train/Val	F17-02120	40
Train/Val	F17-03141	37
Train/Val	F17-05935	35
Train/Val	F17-02723	34
Train/Val	F17-011577	20

Table A4. *Cont.*

Set	Slide	No. of Patches
Train/Val	F18-13364	339
Train/Val	F17-04900	71
Train/Val	F17-02641	40
Train/Val	F17-04911	36
Train/Val	F17-04773	18
	Total	1750
Test	F17-05256	40
Test	F17-02600	40
Test	F17-07510	40
Test	F17-08031	40
Test	F17-011490	40
Test	F17-006348	40
Test	F19-03615	40
Test	F17-010348	40
Test	F18-2508	40
Test	F17-08570	40
Test	F19-7408	40
	Total	440

**Table A5.** The training, validation and hold-out test splits for each fold in the dataset.

Slide	Fold 1	Fold 2	Fold 3
F17-04773	Val	Train	Train
F17-03141	Train	Train	Val
F17-1261	Train	Val	Train
F18-13364	Val	Train	Train
F17-02232	Train	Train	Val
F17-04911	Val	Train	Train
F17-0549	Train	Train	Val
F17-011577	Train	Train	Val
F17-011777	Train	Val	Train
F17-03855	Train	Train	Val
F17-04900	Val	Train	Train
F18-7832	Train	Val	Train
F17-09700	Train	Val	Train
F17-02641	Val	Train	Train
F17-09926	Train	Val	Train
F17-02723	Train	Train	Val
F17-05935	Train	Val	Train
F17-02120	Train	Train	Val
F18-79705	Train	Val	Train
F17-06348	Test	Test	Test
F17-010348	Test	Test	Test
F17-011490	Test	Test	Test
F19-03615	Test	Test	Test
F17-05256	Test	Test	Test
F17-08570	Test	Test	Test
F19-7408	Test	Test	Test
F18-2508	Test	Test	Test
F17-07510	Test	Test	Test
F17-08031	Test	Test	Test
F17-0260	Test	Test	Test

Table A6. The updated agreed mitosis between annotator 1 and 2 for the training/validation sets. The "Agreement" column shows the number of ground truth agreed mitosis annotations for the 20× magnification dataset before refinement. "Updated Agreement" shows the number of mitosis after refinement. "Missed Mitosis" shows the difference in numbers of mitosis before and after refinement. Lastly, "% Missed Mitosis" shows the difference in percentage of mitosis before and after refinement against the updated agreed mitotic count.

Slide	Agreement	Updated Agreement	"Missed" Mitosis	% "Missed" Mitosis
F17-04773	18	18	0	0.00
F17-03141	38	39	1	2.56
F17-1261	41	41	0	0.00
F18-13364	437	460	23	5.00
F17-02232	216	236	20	8.47
F17-04911	37	39	2	5.13
F17-0549	106	112	6	5.36
F17-011577	23	24	1	4.17
F17-011777	212	227	15	6.61
F17-03855	68	71	3	4.23
F17-04900	75	76	1	1.32
F18-7832	331	350	19	5.43
F17-09700	134	138	4	2.90
F17-02641	41	42	1	2.38
F17-09926	62	62	0	0.00
F17-02723	39	39	0	0.00
F17-05935	41	45	4	8.89
F17-02120	42	44	2	4.55
F18-79705	84	91	7	7.69
Total:	2045	2154	109	Avg: 3.93

Table A7. The updated agreed mitosis between annotator 1 and 2 for the hold-out test set. The "Agreement" column shows the number of ground truth agreed mitosis annotations for the 20× magnification dataset before refinement. "Updated Agreement" shows the number of mitosis after refinement. "Missed Mitosis" shows the difference in numbers of mitosis before and after refinement. Lastly, "% Missed Mitosis" shows the difference in percentage of mitosis before and after refinement against the updated agreed mitotic count.

Slide	Agreement	Updated Agreement	"Missed" Mitosis	% "Missed" Mitosis
F17-06348	13	16	3	18.75
F17-010348	1	5	4	80.00
F17-011490	2	3	1	33.33
F19-03615	58	81	23	28.40
F17-05256	35	39	4	10.26
F17-08570	1	1	0	0.00
F19-7408	3	3	0	0.00
F18-2508	11	16	5	31.25
F17-07510	17	26	9	34.62
F17-08031	5	5	0	0.00
F17-0260	1	1	0	0.00
Total:	147	196	49	Avg: 21.51
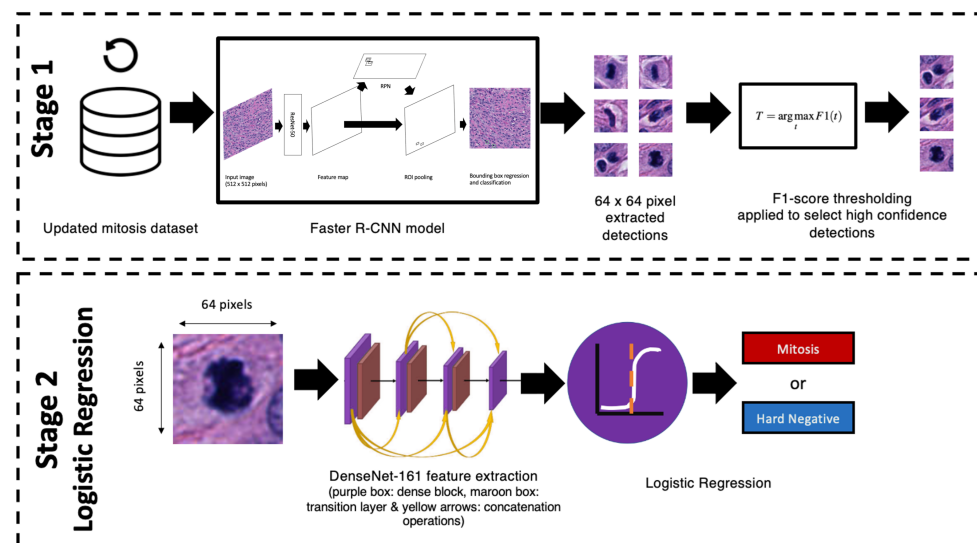
Figure A1. A depiction of the two-stage mitosis detection approach. On the top, in stage 1, 20×
magnification images and annotations from the updated refined mitoses dataset are used to train
a Faster R-CNN model (the model is also presented in Figure 1). Optimal probability thresholds
are applied on the output candidates, which are determined from the validation set (based on
Equation (4)). These selected candidates are then extracted (size 64 × 64 pixels) at 40× magnification
from the original Whole Slide Images (WSIs) and passed into the second stage. On the bottom
shows stage 2 where the extracted patches are fed into a DenseNet-161 ImageNet pre-trained feature
extractor, where the outputs are fed into a logistic regression classifier to determine whether the
candidates are mitosis or difficult false positives.

Table A8. Results from the stage 2 logistic regression model. Across all fold datasets, the sensitivity
has dramatically decreased, and it is offset with a large increase in precision when compared to the
results in Table 4. The mean average F1-scores for the validation and test sets are 0.654 and 0.611,
respectively.

Fold	Set	Sensitivity	Precision	F1-Score	TP	FP	FN
1	Val	0.561	0.906	0.693	356	37	279
	Test	0.526	0.844	0.648	103	19	93
2	Val	0.487	0.906	0.634	465	48	489
	Test	0.592	0.773	0.671	116	34	80
3	Val	0.487	0.920	0.637	275	24	290
	Test	0.367	0.857	0.514	72	12	124
Avg. (mean)	Val	0.512	0.911	0.654			
	Test	0.495	0.825	0.611			

ROC curve

Fold 1

ROC curve

Fold 2

ROC curve

Fold 3
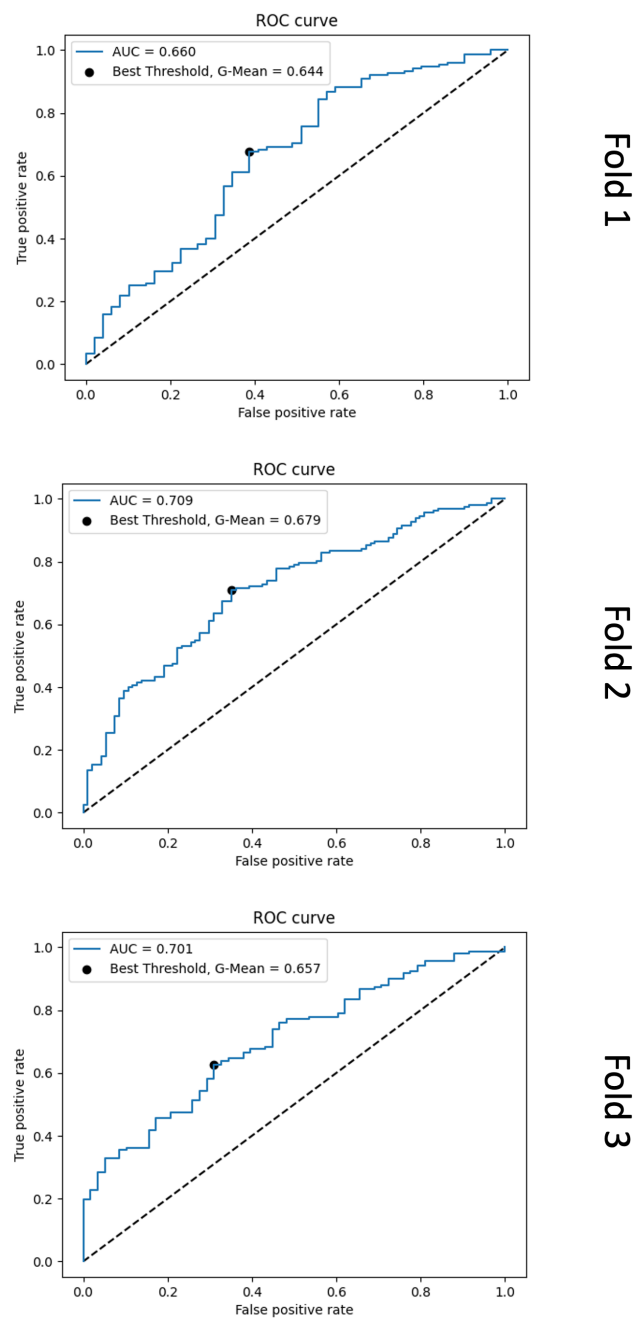
Figure A2. Receiver operating characteristic (ROC) curve plots from the second-stage logistic regression model results for each cross-validation fold. For each fold, it is evident that the models do not effectively learn the differences between true positive (TP) and false positive (FP) mitosis detections.

References

1. Bostock, D.; Dye, M. Prognosis after surgical excision of canine fibrous connective tissue sarcomas. *Vet. Pathol.* **1980**, *17*, 581–588. [CrossRef] [PubMed]
2. Dernell, W.S.; Withrow, S.J.; Kuntz, C.A.; Powers, B.E. Principles of treatment for soft tissue sarcoma. *Clin. Tech. Small Anim. Pract.* **1998**, *13*, 59–64. [CrossRef] [PubMed]
3. Ehrhart, N. Soft-tissue sarcomas in dogs: A review. *J. Am. Anim. Hosp. Assoc.* **2005**, *41*, 241–246. [CrossRef] [PubMed]
4. Mayer, M.N.; LaRue, S.M. Soft tissue sarcomas in dogs. *Can. Vet. J.* **2005**, *46*, 1048. [PubMed]
5. Cavalcanti, E.B.; Gorza, L.L.; de Sena, B.V.; Sossai, B.G.; Junior, M.C.; Flecher, M.C.; Marcolongo-Pereira, C.; dos Santos Horta, R. Correlation of Clinical, Histopathological and Histomorphometric Features of Canine Soft Tissue Sarcomas. *Braz. J. Vet. Pathol.* **2021**, *14*, 151–158. [CrossRef]

6. Torrigiani, F.; Pierini, A.; Lowe, R.; Simčič, P.; Lubas, G. Soft tissue sarcoma in dogs: A treatment review and a novel approach using electrochemotherapy in a case series. *Vet. Comp. Oncol.* **2019**, *17*, 234–241. [CrossRef] [PubMed]

7. Stefanello, D.; Avallone, G.; Ferrari, R.; Roccabianca, P.; Boracchi, P. Canine cutaneous perivascular wall tumors at first presentation: clinical behavior and prognostic factors in 55 cases. *J. Vet. Intern. Med.* **2011**, *25*, 1398–1405. [CrossRef]

8. Chase, D.; Bray, J.; Ide, A.; Polton, G. Outcome following removal of canine spindle cell tumours in first opinion practice: 104 cases. *J. Small Anim. Pract.* **2009**, *50*, 568–574. [CrossRef]

9. Dennis, M.; McSporran, K.; Bacon, N.; Schulman, F.; Foster, R.; Powers, B. Prognostic factors for cutaneous and subcutaneous soft tissue sarcomas in dogs. *Vet. Pathol.* **2011**, *48*, 73–84. [CrossRef]

10. Bray, J.P.; Polton, G.A.; McSporran, K.D.; Bridges, J.; Whitbread, T.M. Canine soft tissue sarcoma managed in first opinion practice: Outcome in 350 cases. *Vet. Surg.* **2014**, *43*, 774–782. [CrossRef]

11. Kuntz, C.; Dernell, W.; Powers, B.; Devitt, C.; Straw, R.; Withrow, S. Prognostic factors for surgical treatment of soft-tissue sarcomas in dogs: 75 cases (1986–1996). *J. Am. Vet. Med. Assoc.* **1997**, *211*, 1147–1151. [CrossRef]

12. McSporran, K. Histologic grade predicts recurrence for marginally excised canine subcutaneous soft tissue sarcomas. *Vet. Pathol.* **2009**, *46*, 928–933. [CrossRef] [PubMed]

13. Avallone, G.; Rasotto, R.; Chambers, J.K.; Miller, A.D.; Behling-Kelly, E.; Monti, P.; Berlato, D.; Valenti, P.; Roccabianca, P. Review of histological grading systems in veterinary medicine. *Vet. Pathol.* **2021**, *58*, 809–828. [CrossRef]

14. Avallone, G.; Helmbold, P.; Caniatti, M.; Stefanello, D.; Nayak, R.; Roccabianca, P. The spectrum of canine cutaneous perivascular wall tumors: Morphologic, phenotypic and clinical characterization. *Vet. Pathol.* **2007**, *44*, 607–620. [CrossRef]

15. Loures, F.; Conceição, L.; Lauffer-Amorim, R.; Nóbrega, J.; Costa, E.; Torres, R.; Clemente, J.; Vilória, M.; Silva, J. Histopathology and immunohistochemistry of peripheral neural sheath tumor and perivascular wall tumor in dog. *Arq. Bras. Med. Vet. Zootec.* **2019**, *71*, 1100–1106. [CrossRef]

16. Mathew, T.; Kini, J.R.; Rajan, J. Computational methods for automated mitosis detection in histopathology images: A review. *Biocybern. Biomed. Eng.* **2021**, *41*, 64–82. [CrossRef]

17. Aubreville, M.; Stathonikos, N.; Bertram, C.A.; Klopfleisch, R.; Ter Hoeve, N.; Ciompi, F.; Wilm, F.; Marzahl, C.; Donovan, T.A.; Maier, A.; et al. Mitosis domain generalization in histopathology images—The MIDOG challenge. *Med. Image Anal.* **2023**, *84*, 102699. [CrossRef] [PubMed]

18. Kaman, E.; Smeulders, A.; Verbeek, P.; Young, I.; Baak, J. Image processing for mitoses in sections of breast cancer: A feasibility study. *Cytom. J. Int. Soc. Anal. Cytol.* **1984**, *5*, 244–249. [CrossRef] [PubMed]

19. Gallardo, G.M.; Yang, F.; Ianzini, F.; Mackey, M.; Sonka, M. Mitotic cell recognition with hidden Markov models. In *Proceedings of the Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display*; SPIE: Bellingham, WA, USA, 2004; Volume 5367, pp. 661–668.

20. Tao, C.Y.; Hoyt, J.; Feng, Y. A support vector machine classifier for recognizing mitotic subphases using high-content screening data. *SLAS Discov.* **2007**, *12*, 490–496. [CrossRef]

21. Liu, A.; Li, K.; Kanade, T. Mitosis sequence detection using hidden conditional random fields. In Proceedings of the 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Rotterdam, The Netherlands, 14–17 April 2010; pp. 580–583.

22. Roux, L.; Racoceanu, D.; Loménie, N.; Kulikova, M.; Irshad, H.; Klossa, J.; Capron, F.; Genestie, C.; Le Naour, G.; Gurcan, M.N. Mitosis detection in breast cancer histological images An ICPR 2012 contest. *J. Pathol. Inform.* **2013**, *4*.

23. Aubreville, M.; Bertram, C.; Veta, M.; Klopfleisch, R.; Stathonikos, N.; Breininger, K.; ter Hoeve, N.; Ciompi, F.; Maier, A. Quantifying the Scanner-Induced Domain Gap in Mitosis Detection. *arXiv* **2021**, arXiv:2103.16515.

24. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 142–158. [CrossRef]

25. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

26. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv* **2015**, arXiv:1506.01497.

27. Rao, S. Mitos-rcnn: A novel approach to mitotic figure detection in breast cancer histopathology images using region based convolutional neural networks. *arXiv* **2018**, arXiv:1807.01788.

28. Veta, M.; Van Diest, P.J.; Willems, S.M.; Wang, H.; Madabhushi, A.; Cruz-Roa, A.; Gonzalez, F.; Larsen, A.B.; Vestergaard, J.S.; Dahl, A.B.; et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med. Image Anal.* **2015**, *20*, 237–248. [CrossRef]

29. Roux, L.; Racoceanu, D.; Capron, F.; Calvo, J.; Attieh, E.; Le Naour, G.; Gloaguen, A. Mitos & atypia. Detection of Mitosis and Evaluation of Nuclear Atypia Score in Breast Cancer Histological Images. 2014, Volume 1, pp. 1–8. Available online: http://ludo17.free.fr/mitos_atypia_2014/icpr2014_MitosAtypia_DataDescription.pdf (accessed on 28 January 2024).

30. Aubreville, M. MItosis DOmain Generalization Challenge 2022 (MICCAI MIDOG 2022), Training Data Set (PNG version) (1.0) [Data Set]. Zenodo. Available online: https://zenodo.org/records/6547151 (accessed on 28 January 2024).

31. Aubreville, M.; Wilm, F.; Stathonikos, N.; Breininger, K.; Donovan, T.A.; Jabari, S.; Veta, M.; Ganz, J.; Ammeling, J.; van Diest, P.J.; et al. A comprehensive multi-domain dataset for mitotic figure detection. *Sci. Data* **2023**, *10*, 484. [CrossRef]

32. Aubreville, M.; Bertram, C.A.; Marzahl, C.; Gurtner, C.; Dettwiler, M.; Schmidt, A.; Bartenschlager, F.; Merz, S.; Fragoso, M.; Kershaw, O.; et al. Deep learning algorithms out-perform veterinary pathologists in detecting the mitotically most active tumor region. *Sci. Rep.* **2020**, *10*, 1–11.

33. Bertram, C.A.; Aubreville, M.; Marzahl, C.; Maier, A.; Klopfleisch, R. A large-scale dataset for mitotic figure assessment on whole slide images of canine cutaneous mast cell tumor. *Sci. Data* **2019**, *6*, 1–9. [CrossRef] [PubMed]

34. Litjens, G. Automated Slide Analysis Platform (ASAP). 2017. Available online: https://www.computationalpathologygroup.eu/software/asap/ (accessed on 28 January 2024).

35. Elston, C.W.; Ellis, I.O. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology* **1991**, *19*, 403–410. [CrossRef]

36. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.

37. Mahmood, T.; Arsalan, M.; Owais, M.; Lee, M.B.; Park, K.R. Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs. *J. Clin. Med.* **2020**, *9*, 749. [CrossRef] [PubMed]

38. Halmes, M.; Heuberger, H.; Berlemont, S. Deep Learning-based mitosis detection in breast cancer histologic samples. *arXiv* **2021**, arXiv:2109.00816.

39. Zhou, Y.; Mao, H.; Yi, Z. Cell mitosis detection using deep neural networks. *Knowledge-Based Systems* **2017**, *137*, 19–28. [CrossRef]

40. Henderson, P.; Ferrari, V. End-to-end training of object class detectors for mean average precision. In Proceedings of the Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016; pp. 198–213.

41. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]

42. Everingham, M.; Eslami, S.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes challenge: A retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [CrossRef]

43. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]

44. Rai, T.; Morisi, A.; Bacci, B.; Bacon, N.J.; Dark, M.J.; Aboellail, T.; Thomas, S.A.; Bober, M.; La Ragione, R.; Wells, K. Deep learning for necrosis detection using canine perivascular wall tumour whole slide images. *Sci. Rep.* **2022**, *12*, 10634. [CrossRef]

45. Morisi, A.; Rai, T.; Bacon, N.J.; Thomas, S.A.; Bober, M.; Wells, K.; Dark, M.J.; Aboellail, T.; Bacci, B.; La Ragione, R.M. Detection of Necrosis in Digitised Whole-Slide Images for Better Grading of Canine Soft-Tissue Sarcomas Using Machine-Learning. *Vet. Sci.* **2023**, *10*, 45. [CrossRef] [PubMed]

46. Rai, T.; Papanikolaou, I.; Dave, N.; Morisi, A.; Bacci, B.; Thomas, S.; La Ragione, R.; Wells, K. Investigating the potential of untrained convolutional layers and pruning in computational pathology. In Proceedings of the Medical Imaging 2023: Digital and Computational Pathology, San Diego, CA, USA, 19–23 February 2023; Volume 12471, pp. 323–332.

47. Aubreville, M.; Bertram, C.A.; Donovan, T.A.; Marzahl, C.; Maier, A.; Klopfleisch, R. A completely annotated whole slide image dataset of canine breast cancer to aid human breast cancer research. *Sci. Data* **2020**, *7*, 1–10. [CrossRef] [PubMed]

48. Piansaddhayanon, C.; Santisukwongchote, S.; Shuangshoti, S.; Tao, Q.; Sriswasdi, S.; Chuangsuwanich, E. ReCasNet: Improving consistency within the two-stage mitosis detection framework. *arXiv* **2022**, arXiv:2202.13912.

49. Çayır, S.; Solmaz, G.; Kusetogullari, H.; Tokat, F.; Bozaba, E.; Karakaya, S.; Iheme, L.O.; Tekin, E.; Özsoy, G.; Ayaltı, S.; et al. MITNET: A novel dataset and a two-stage deep learning approach for mitosis recognition in whole slide images of breast cancer tissue. *Neural Comput. Appl.* **2022**, *34*, 17837–17851. [CrossRef]