

Article

Detection of Necrosis in Digitised Whole-Slide Images for Better Grading of Canine Soft-Tissue Sarcomas Using Machine-Learning

Ambra Morisi ^{1,*}, Taran Rai ², Nicholas J. Bacon ^{1,3}, Spencer A. Thomas ^{2,4} , Mirosław Bober ², Kevin Wells ², Michael J. Dark ⁵ , Tawfik Aboellail ⁶ , Barbara Bacci ⁷  and Roberto M. La Ragione ^{1,8}

- ¹ Department of Comparative Biomedical Sciences, School of Veterinary Medicine, University of Surrey, Guildford GU2 7AL, UK
² Centre for Vision Speech and Signal Processing, University of Surrey, Guildford GU2 7XH, UK
³ AURA Veterinary, Guildford GU2 7AJ, UK
⁴ National Physical Laboratory, Teddington TW11 0LW, UK
⁵ Department of Comparative, Diagnostic, and Population Medicine, College of Veterinary Medicine, University of Florida, Gainesville, FL 32610, USA
⁶ Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, CO 80523, USA
⁷ Department of Veterinary Medical Sciences, University of Bologna, 40126 Bologna, Italy
⁸ Department of Microbial Sciences, School of Biosciences, University of Surrey, Guildford GU2 7XH, UK
* Correspondence: ambramorisi@hotmail.it

Simple Summary: Canine soft-tissue sarcomas are a group of tumours that arise from the skin and subcutaneous connective tissue. The most common method used to predict the behaviour of these tumours is grading. The grading system used for soft-tissue sarcomas is derived from a combined score calculated by evaluating the mitotic count, percentage of tumour necrosis and degree of cellular differentiation. However, these parameters are highly subjective and a high inter-observer variability has been reported in grading these tumours, which can result in complications regarding treatment plans. Manual identification of areas of necrosis is a time-consuming task that is prone to observer error. Artificial-intelligence algorithms and, in particular, machine learning, can help improve grading by automatically detecting regions of necrosis. The aim of this study was to differentiate image regions in order to automatically identify tumour necrosis in digitised canine soft-tissue sarcoma slides. This method showed an accuracy of 92.7% which represents the number of correctly classified data instances over the total number of data instances. Therefore, the proposed method is a promising tool to minimise human error in the evaluation of necrosis in soft-tissue sarcomas, and hence increase the efficiency and accuracy of histopathological grading of canine soft-tissue sarcomas.

Abstract: The definitive diagnosis of canine soft-tissue sarcomas (STSs) is based on histological assessment of formalin-fixed tissues. Assessment of parameters, such as degree of differentiation, necrosis score and mitotic score, give rise to a final tumour grade, which is important in determining prognosis and subsequent treatment modalities. However, grading discrepancies are reported to occur in human and canine STSs, which can result in complications regarding treatment plans. The introduction of digital pathology has the potential to help improve STS grading via automated determination of the presence and extent of necrosis. The detected necrotic regions can be factored in the grading scheme or excluded before analysing the remaining tissue. Here we describe a method to detect tumour necrosis in histopathological whole-slide images (WSIs) of STSs using machine learning. Annotated areas of necrosis were extracted from WSIs and the patches containing necrotic tissue fed into a pre-trained DenseNet161 convolutional neural network (CNN) for training, testing and validation. The proposed CNN architecture reported favourable results, with an overall validation accuracy of 92.7% for necrosis detection which represents the number of correctly classified data instances over the total number of data instances. The proposed method, when vigorously validated represents a promising tool to assist pathologists in evaluating necrosis in canine STS tumours, by increasing efficiency, accuracy and reducing inter-rater variation.



Citation: 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. <https://doi.org/10.3390/vetsci10010045>

Received: 10 November 2022

Revised: 9 December 2022

Accepted: 21 December 2022

Published: 8 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: canine; digital pathology; grading; machine learning; necrosis detection; soft-tissue sarcoma

1. Introduction

Soft-tissue sarcomas (STSs) are tumours derived from mesenchymal tissues [1–4]. In dogs, they develop most frequently in the subcutis where they represent between 9 and 15% of all cutaneous or subcutaneous tumours [3,5,6]. Histological assessment of canine STSs is traditionally performed by microscopic analysis of tissue sections on glass slides. The most-used histological parameter to prognosticate canine STSs and predict their outcome following surgery is the tumour grade, derived from a combined score calculated based on cellular differentiation, mitotic index and percentage of tumour necrosis [7–10]. The application of these histologic criteria allows individual STSs to be categorised into three distinct grades (I, low grade; II, intermediate grade or III, high grade) [5,9–12]. Tumour necrosis is a common feature of solid tumours caused by ischaemic injury, owing to rapid rates of tumour growth [13]. However, manual identification and calculation of necrotic regions by visual inspection can be a time-consuming and error-prone task for large whole-slide images [14] and can lead to inter-observer variability [15]. Recent technological advances, on the other hand, would allow histological tumour slides to be converted into digital image datasets for automated analysis. Utilising machine learning to interrogate patterns in these digital histological images may address some of the limitations of manual grading [16]. Machine-learning algorithms have been evaluated with success in the field of human oncology and histopathology for glioma, renal clear cell carcinoma, breast cancer, gastric carcinoma, prostate cancer, and non-small-cell lung cancer [16–25]. Machine-learning methods have been pivotal to investigating the degree of necrosis in pathology images. Sharma et al. [14] used machine learning for necrosis detection in gastric carcinomas with the best average cross-validation rate (which estimates how accurately a predictive model will perform in practice) of 85.3%. In a second study published by the same authors [14], the proposed CNN architecture for necrosis detection in human gastric carcinoma had the best overall rate of 81.4%. In a more recent study performed by Arunachalam et al. [26] on human osteosarcomas, the accuracy in detecting areas of necrosis using their proposed model was 92.7%.

In this study, we attempt to differentiate image regions in order to identify tumour necrosis in the haematoxylin and eosin (H&E)-stained WSIs of canine STSs (cSTSs). There were several motivations for this study. We previously published [27] the first report on the use of deep learning to detect cSTSs in haematoxylin and eosin (H&E)-stained whole slides. However, the study reported here builds on the initial study and focuses on grading. Necrosis is a specific determinant in assigning a histological grade to cSTSs, and so it is important to recognise and quantify necrosis in STS sections. More generally, the presence of necrosis is considered a characteristic of malignancy and subjectively influences the pathologist's judgement on tumour behaviour. Furthermore, there are no specific histological stains for necrosis which makes automatic detection of necrosis a highly desirable objective. Automatic necrosis detection could decrease viewing times for pathologists, reduce inter-observer variabilities and hence increase the accuracy of diagnosis and prognosis. Lastly, these methods could also be applied to other histopathology datasets. In this study we applied a pre-trained DenseNet161 CNN model to automatically detect necrosis in cSTSs from WSIs.

2. Materials and Methods

2.1. Dataset and Slide Annotation

A total of 90 WSIs of canine perivascular-wall tumours (cPWTs), a subtype of cSTSs, were collected from the Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins. The digitised slides were reviewed, and the diagnoses and grades were confirmed by two board-certified veterinary pathologists.

Of the 90 cPWTs, necrosis, which was characterised by loss of cellular detail and presence of eosinophilic amorphous material, was identified in 21 cases. The 21 WSIs containing necrosis (one WSI for each case) were manually annotated by a board-certified veterinary pathologist. The same slides were annotated separately by a veterinary surgeon after training in the annotation of necrosis in WSIs. The pathologist had regular follow-up discussions with the veterinary surgeon regarding the annotation procedure and annotation rules. The annotations were made using the open source Automated Slide Analysis Platform (ASAP) software by delineating the contour of necrotic areas using mouse clicks. To avoid potential human bias and increase accuracy, only areas of consensus where the veterinary pathologist and the veterinary surgeon agreed were included in the analysis [16].

2.2. Pre-Processing

Patches or “tiles” of 256×256 pixel size, were extracted from a large WSI and fed into a pre-trained convolutional neural network. Every patch containing at least 30% necrosis derived from the expert mark-ups was extracted to create the necrosis class. As a proportion of the WSIs did not contain necrosis, and to avoid class imbalance between necrosis versus non-necrosis patches, a subset of all the non-necrosis patches was extracted. It should be noted that these non-necrosis patches were chosen at random. An example of the patch extraction of necrosis is given in Figure 1.

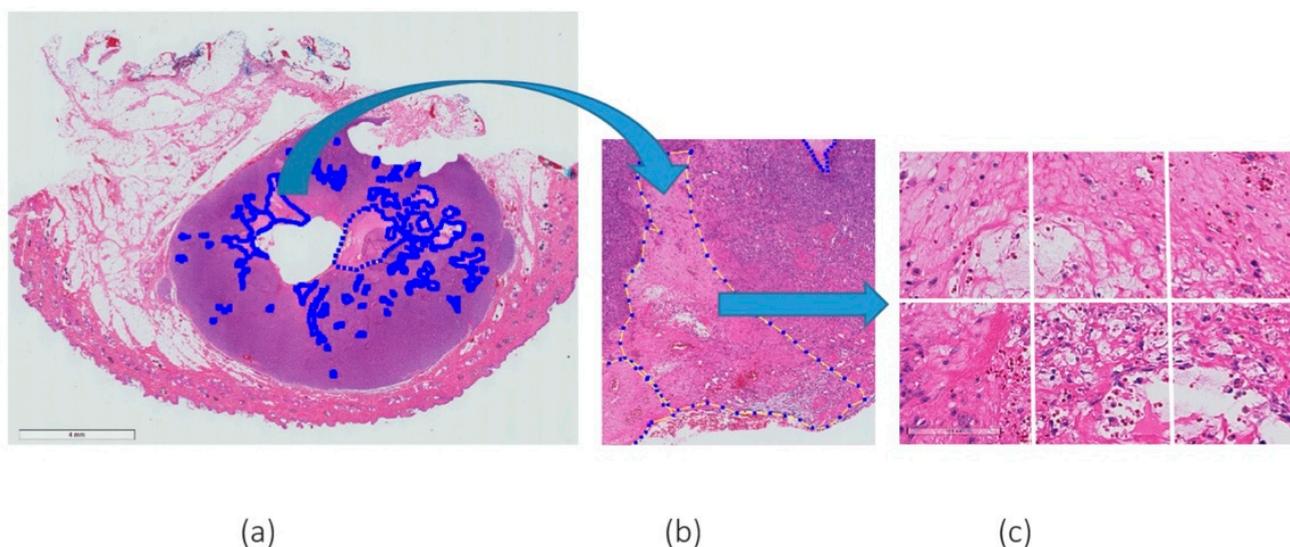


Figure 1. Example of H&E WSI with (a) a few annotations of necrosis marked by expert pathologists, (b) a magnified (magnification $5\times$) region of agreement of both pathologists and (c) an example of patches extracted from the annotated region (magnification $20\times$).

Patch-based approaches are required as whole-slide images are too large to pass through CNN architectures.

The aim of this study was to classify entire patch images; therefore, as patches may contain a mix of necrotic and non-necrotic tissue, a threshold needed to be considered for determining patches containing necrosis (positive patches) and patches that did not contain necrosis (negative patches). Histopathological images with multiple levels of magnification can depict various types of information. For this study, $20\times$ objective magnification was deemed appropriate following a discussion with the pathologists. The higher magnification allows clear identification of necrosis in a WSI image while a lower magnification resulted in loss of the spatial information. However, higher magnifications include sub-cellular regions that are not relevant to the task and may negatively affect the segmentation process.

For the $20\times$ magnification, the threshold applied for negative patch was 0.75 (75% of the patch must not contain necrosis for it to be labelled as non-necrosis). This means that at least 25% of the patch must contain necrosis for it to be labelled as necrosis (positive patch).

The total percentage of necrosis present in a slide was calculated from the intersection of the ground truth labelling from human experts. The four grade 3 cSTS slides had 10.92%, 3.05%, 10.06% and 10.59% of necrosis. The grade 2 cSTS slides contained 0.74%, 1.24%, 0.00% and 0.07% of necrosis. None of the grade 1 cSTS slides contained necrosis. These data were used for training, validating and testing the algorithm.

A pre-trained DenseNet161 CNN model was implemented in Python using the PyTorch library and experiments were performed on Dell T630 system, including two Intel Xeon E5 v4 series 8-Core CPUs, four Nvidia Titan X GPUs and 128GB of RAM.

2.3. DenseNet161

Several pre-trained networks were investigated [28,29] and DenseNet161 CNN was deemed appropriate. According to the results reported in a study conducted by Talo [30], DenseNet161 can be used for fast and accurate classification of histopathology images to assist pathologists in their daily clinical tasks. The DenseNet161 model is one of the DenseNet group of models designed to perform image classification [31]. In our deep learning set-up, there were two components: a feature extractor model that was pretrained on the ImageNet dataset [32] and a dense classification layer model. The weights in the convolutional layers of the feature extractor were frozen and the dense classification layer was amended for a binary classification task. Features were extracted using the feature extractor and passed into the dense classification layer model. Training was implemented using the Adam optimiser [33], a learning rate of 0.0001, and a batch size of 32. Test-time normalization was also implemented.

2.4. Training, Validation and Testing

The dataset was divided into several subsets as follows: a *training dataset* which is the set of data that are used to train and make the model learn the features/patterns in the data; a *validation dataset* which is a set of data, separate from the training set, that is used to validate the model performance during training; and a *test dataset* which is a separate set of data used to test the model *after* completion of the training. Although the validation and test datasets are similar, test datasets are “unseen” whereas validation is used as an informative dataset during training.

To ensure statistical robustness, 3-fold cross validation was implemented. Three experiments named fold 1, fold 2 and fold 3 were run. The three experiments, fold 1, fold 2 and fold 3, contained 3824, 3754 and 3990 non-overlapping patches of necrosis, respectively, for training, and 1960, 2030 and 1794 patches of necrosis, respectively, for validation (Table 1).

Table 1. The data in the table relate to information for the training and validation datasets for each experiment (folds 1, 2 and 3). Each table (folds 1, 2 and 3) contains information about WSI number, tumour grade, number of patches extracted from necrosis areas (positive patches) and normal tissue (negative patches) for each WSI. Each table also indicates the total number of positive and negative patches used for training and validation.

Slide Code	Grade	Fold 1	No. of Necrosis Patches 20×	No. of Negative Patches 20×
#1	1	Validation	0	2856
#2	1	Validation	0	3379
#3	2	Validation	0	2606
#4	2	Validation	314	1417
#5	3	Validation	86	2542
#6	3	Validation	1560	2324
Total			1960	15,124

Table 1. Cont.

Slide Code	Grade	Fold 1	No. of Necrosis Patches 20×	No. of Negative Patches 20×
#7	1	Training	0	800
#8	1	Training	0	800
#9	1	Training	0	800
#10	2	Training	41	800
#11	2	Training	0	800
#12	2	Training	0	800
#13	2	Training	0	800
#14	2	Training	1696	800
#15	2	Training	0	800
#16	3	Training	57	800
#17	3	Training	934	800
#18	3	Training	210	800
#19	3	Training	742	800
#20	3	Training	144	800
Total			3824	11,200
Slide Code	Grade	Fold 2	No. of Necrosis Patches 20×	No. of Negative Patches 20×
#9	1	Validation	0	2132
#8	1	Validation	0	2280
#11	2	Validation	0	4944
#19	3	Validation	742	2367
#17	3	Validation	934	1332
#20	3	Validation	144	1727
#18	3	Validation	210	2280
Total			2030	17,062
Slide Code	Grade	Fold 3	No. of Necrosis Patches 20×	No. of Negative Patches 20×
#1	1	Training	0	800
#2	1	Training	0	800
#7	1	Training	0	800
#4	2	Training	314	800
#10	2	Training	41	800
#4	2	Training	0	800
#12	2	Training	0	800
#13	2	Training	0	800
#14	2	Training	1696	800
#15	2	Training	0	800
#16	3	Training	57	800
#6	3	Training	1560	800
#5	3	Training	86	800
Total			3754	10,400
Slide Code	Grade	Fold 3	No. of Necrosis Patches 20×	No. of Negative Patches 20×
#7	1	Validation	0	2259
#12	2	Validation	0	3562
#13	2	Validation	0	2551
#14	2	Validation	1696	2119
#15	2	Validation	0	2936
#10	2	Validation	41	2983
#16	3	Validation	57	2379
Total			1794	18,789

Table 1. *Cont.*

Slide Code	Grade	Fold 3	No. of Necrosis Patches 20×	No. of Negative Patches 20×
#1	1	Training	0	800
#9	1	Training	0	800
#2	1	Training	0	800
#8	1	Training	0	800
#4	2	Training	314	800
#3	2	Training	0	800
#20	3	Training	144	800
#5	3	Training	86	800
#18	3	Training	210	800
#19	3	Training	742	800
#17	3	Training	934	800
#11	3	Training	0	800
#6	2	Training	1560	800
Total			3990	10,400

For training, 100 epochs were used. The best model selected was the one with the lowest validation loss.

After training and validation, the algorithm was tested on a test dataset (Table 2).

Table 2. Test dataset for folds 1, 2 and 3. The data represented in the table contain information about slide code, tumour grade, number of patches containing necrosis (positive), number of positive and negative patches (total) and % of necrosis present in each WSI.

Slide Code	Grade	Positive	Total	Necrosis %
#21	1	0	4371	0.00
#22	1	0	1611	0.00
#23	1	0	2798	0.00
#24	1	0	3040	0.00
#25	2	14	1883	0.74
#26	2	0	3368	0.00
#27	2	20	1618	1.24
#28	2	2	2714	0.07
#29	3	302	3003	10.06
#30	3	138	4528	3.05
#31	3	369	3378	10.92
#32	3	306	2890	10.59

The test dataset contained 35,202 patches extracted from 12 WSIs (four slides for each tumour grade). The patches that contained necrosis were identified as “positive” and the sum of the patches extracted from the slide containing necrosis and normal tissue were identified as “total”. The total percentage of necrosis present in a slide was then calculated. A probability map for each slide was also generated for a WSI for visualisation purposes.

3. Results

The results from the study are presented below in Figures 2–4 and Tables 3–5. After training the algorithm, the estimated model’s accuracy was computed using 3-fold cross-validation of the dataset. Training accuracy ranged from 89.2% to 98.0%. The training loss ranged from 0.058 to 0.295 (Figure 2). The proposed CNN architecture reported favourable results, with an overall validation accuracy of 92.7% for necrosis detection which represents the number of correctly classified data instances over the total number of data instances (Table 3).

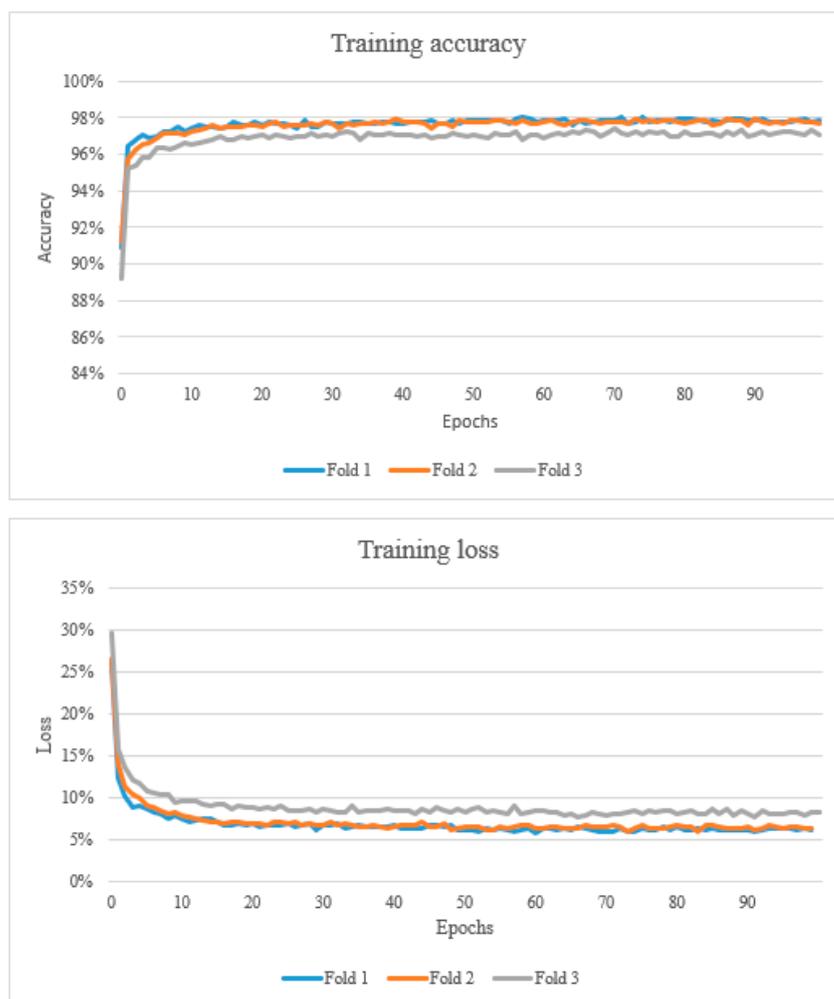


Figure 2. The top graph shows the training accuracy (learning curve). A learning curve is a plot that shows epochs on the x-axis and learning or improvement on the y-axis. The bottom graph shows the training loss which represents the summation of errors in the model.

Overall, the validation accuracy ranged from 88.6% to 94.7%. Validation average class-wise accuracy ranged from 89.9% to 94.4%. The validation loss ranged from 0.143 to 0.269 (Figure 3).

After training and validating the algorithm, we computed *accuracy*, *sensitivity*, *precision* and *F1 score* (Table 3) using the number of true positives (TPs), true negatives (TNs), false positives (FPs) and false negatives (FNs).

Accuracy represents the number of correctly classified data instances over the total number of data instances that was calculated as follows:

$$Accuracy : \frac{TP + TN}{TP + TN + FP + FN}$$

The results ranged from 89.1% to 92.7%.

Recall, also known as *sensitivity* or *true positive rate*, is defined as follows:

$$Recall : \frac{TP}{TP + FN}$$

Recall ranged from 93.4% to 94.6%.

Precision (positive predictive value) in classifying the data instances is defined as follows:

$$Precision : \frac{TP}{TP + FP}$$

Precision ranged from 22.4% to 30.0%.

F1-score is a metric, which considers both *precision* and *recall* and is defined as follows:

$$F1\text{-score} : 2 * \frac{Precision \times Recall}{Precision + Recall}$$

F1-score ranged from 36.2% to 45.4%.

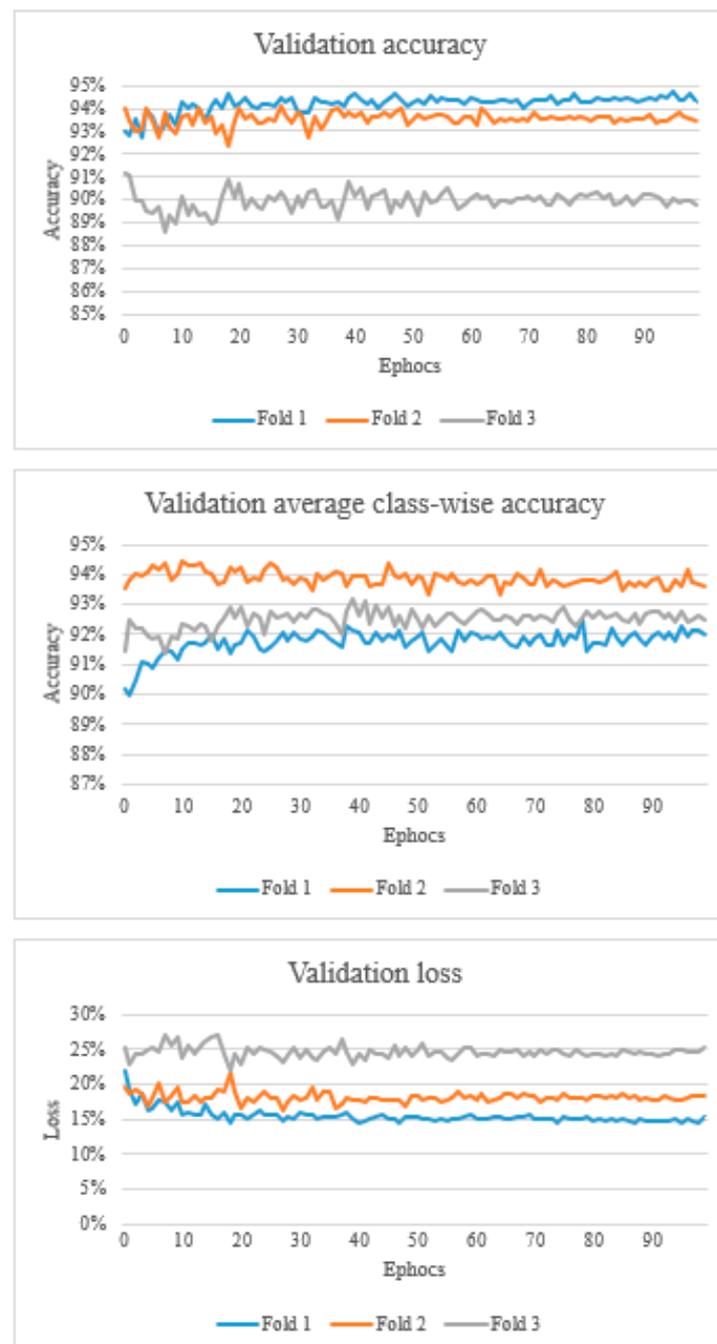


Figure 3. The top graph shows the validation accuracy, the middle graph shows the validation average class-wise accuracy, and the bottom graph shows the validation loss.

Table 3. Accuracy, sensitivity precision and F1 score for validation and test for folds 1, 2 and 3.

Slide Code	Sensitivity/Recall (%)	Precision (%)	Accuracy (%)	F1-Score (%)
Fold1_validation	88.5	70.05	94.4	78.5
Fold1_test	93.4	30.0	92.7	45.4
Fold2_validation	94.3	63.2	93.6	75.7
Fold2_test	94.0	25.4	90.8	39.9
Fold3_validation	95.9	46.8	90.1	62.9
Fold3_test	94.6	22.4	89.1	36.2

Table 4. Predicted positive labels and % of necrosis that the algorithm detected in each slide after training.

Fold 1			
Slide	Predicted Positive	Total	Necrosis %
#21	165	4371	3.77
#22	343	1883	18.22
#23	147	3040	4.84
#24	44	3368	1.31
#25	22	2798	0.79
#26	45	1618	2.78
#27	56	2714	2.06
#28	56	1611	3.48
#29	535	3003	17.82
#30	905	4528	19.99
#31	873	3378	25.84
#32	395	2890	13.67
Fold 2			
Slide	Predicted Positive	Total	Necrosis %
#21	192	4371	4.39
#22	343	1883	18.22
#23	270	3040	8.88
#24	61	3368	1.81
#25	31	2798	1.11
#26	53	1618	3.28
#27	55	2714	2.03
#28	76	1611	4.72
#29	537	3003	17.88
#30	1153	4528	25.46
#31	997	3378	29.51
#32	498	2890	17.23
Fold 3			
Slide	Predicted Positive	Total	Necrosis %
#21	388	4371	8.88
#22	412	1883	21.88
#23	474	3040	15.59
#24	70	3368	2.08
#25	30	2798	1.07
#26	46	1618	2.84
#27	74	2714	2.73
#28	76	1611	4.72
#29	621	3003	20.68
#30	1130	4528	24.96
#31	1172	3378	34.70
#32	368	2890	12.73

Table 5. Predicted % of necrosis that the algorithm detected in each slide after training and the % of necrosis annotated by the pathologists.

Slide	Fold 1	Fold 2	Fold 3	Pathologists' Annotations
#21	3.77	4.39	8.88	0.00
#22	18.22	18.22	21.88	0.00
#23	4.84	8.88	15.59	0.00
#24	1.31	1.81	2.08	0.00
#25	0.79	1.11	1.07	0.74
#26	2.78	3.28	2.84	0.00
#27	2.06	2.03	2.73	1.24
#28	3.48	4.72	4.72	0.07
#29	17.82	17.88	20.68	10.06
#30	19.99	25.46	24.96	3.05
#31	25.84	29.51	34.70	10.92
#32	13.67	17.23	12.73	10.59

The data in Table 4 represent the test data for the three experiments (folds 1, 2 and 3) with the number of predicted positive patches (patches that contain necrosis), the total number of patches and the percentage of necrosis in each slide. The data in Table 5 illustrate the predicted % of necrosis that the algorithm detected in each slide after training and the % of necrosis annotated by the pathologists.

A tumour necrosis-prediction map for each slide was also generated for a WSI for visualisation purposes (Figure 4). The prediction for true positive, false positive, true negative and false negative are expressed in red, orange, clear and green, respectively.

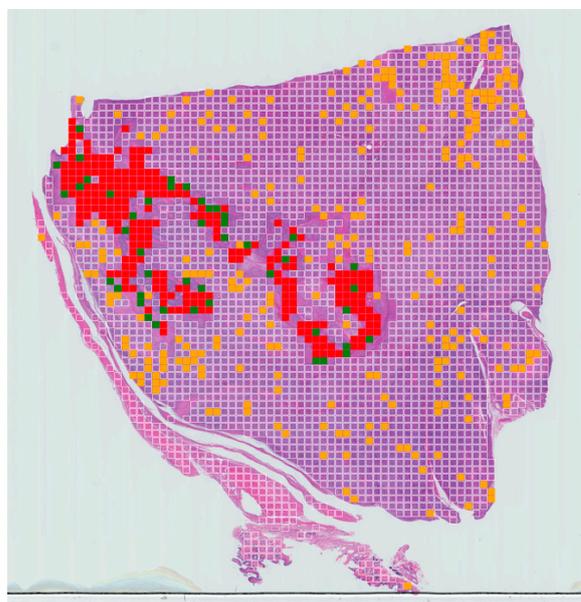


Figure 4. Example of slide-level confusion map showing areas of necrosis within a WSI of canine soft-tissue sarcoma. True positives (TPs) are displayed in red, false negatives (FNs) in green, false positives (FPs) in yellow and true negatives (TNs) in clear. These maps can be used to calculate the percentage of tumour necrosis in a patient and visualise the extent of the tumour necrosis over the whole-slide image.

4. Discussion

In this study, we present a detailed report describing the automatic assessment of WSIs for the detection and quantification of necrosis in cSTSs, providing further insight and analysis from our baseline approach as previously published [27]. The experiments pre-

sented in this study confirmed that DenseNet161 is able to recognise areas of necrosis with high accuracy (92.7%). Recent studies have shown that CNN can be substantially deeper, more accurate, and more efficient to train than other models [31]. In the current study we used a dense convolutional network (DenseNet). This CNN has several advantages including reducing the number of parameters that need to be learnt in the training phase. At the 0th and close to 0th epoch, the training and validation accuracy were close to 90%. From the experimental datasets it was observed that all three folds produced remarkably consistent results (fold 1 accuracy 92.7%, fold 2 90.8% and fold 3 89.1%). However, the F1 score and precision had lower results (fold 1 F1 score 78.5%, fold 2 75.7% and fold 3 62.9%; fold 1 precision 70.5%, fold 2 63.2% and fold 3 46.8%). This could be due to the fact that this method may fail to generalise to smaller datasets.

On further analysis of the results for necrosis detection, it appears that DenseNet161 had high overall accuracy rates (92.7%) in detecting areas of necrosis in canine STSs from a WSI. This model currently outperforms several methods described in the literature [14,16,26], as we detail below.

In the first study conducted by Sharma et al. [14] the machine-learning model used for necrosis detection in gastric carcinomas had the best average cross-validation rate of 85.3%. In a second study by the same authors in 2017, the proposed CNN architecture for necrosis detection in human gastric carcinoma had the best overall rate of 81.4%. In a more recent study performed by Arunachalam et al. on human osteosarcoma, the accuracy in detecting areas of necrosis using their proposed model was 92.7%, which is identical to the results presented here. Hence, our proposed model DenseNet161 achieved a comparatively favourable performance outcome. These results are unlikely to be influenced by tumour type as the histopathological appearance of coagulative necrosis is the same, however further validation studies should be performed to confirm this. It was also interesting to note that half the number of WSIs were required to train the deep convolutional models compared to similar studies on detection of viable and necrotic tumours in human osteosarcoma [26]. This could be due to the fact that our model is pre-trained and therefore requires less data compared to a non-pre-trained model to achieve similar results.

Furthermore, from the data presented in the learning curves, it can be observed that they are generally constant with a decreasing training error and increasing validation accuracy. It can be also noted that the validation loss is nearly constant due to characteristics of validation data, but training loss, which represents the summation of errors in the model, decreases and validation accuracy increases to become constant. This is a desired characteristic in training as it indicates the model is not overfitting the training data.

As a final validation step, tumour necrosis-prediction maps were generated to display the necrotic regions. These maps can be used to calculate the percentage of tumour necrosis in each patient and visualise the extent of the tumour necrosis over the whole-slide image.

The number of false positive and false negatives are controlled by the threshold, and they can be adjusted as required to suit a specific problem. In our case, we wanted to minimise the number of false negatives. From the data presented here, it appears there is a very low rate of false negatives meaning that very few areas of necrosis are likely to be missed, but that there is a tendency to overestimate areas of detected necrosis (false positives). This can be seen in our results where the algorithm detected more areas of necrosis compared to the ground truth labelling from human experts. However, this overall approach can provide a rapid first pass through a given WSI, which, whilst not intending to replace the skilled expert pathologist, can provide a rapid early indication of areas requiring expert attention.

Analysis of the datasets revealed several limitations to the study as follows: firstly, like all other deep-learning applications published to date in human medicine [14,16,26], our method also requires training with large-scale datasets containing thousands of images. This issue could be addressed using data-augmentation strategies and/or by increasing the number of cases from different sources. Secondly, the ground-truth data were generated by expert pathologists annotating WSIs, which demands a significant time investment from

specialists, in common with the vast majority of other supervised deep-learning approaches in clinical sciences. Thirdly, training the proposed model requires approximately four to five days and the availability of significant GPU computer resource. Both issues could be rectified with a large increase in resources, both human and computational, with the additional required financial investment. Lastly, tissue slides can vary in appearance due to the biopsy technique, slide preparation, staining, processing and scanning techniques used in different pathology laboratories [34]. However, our dataset only used slides collected from one institution in order to reduce these variations as much as possible. Future studies should focus on implementing methods for automatic colour and intensity normalisation of WSIs.

There are currently multiple aspects of necrosis assessment that need to be defined. First of all, the current methods for determining percent of tumour necrosis in cSTSs have been poorly defined, meaning that it is difficult for others to replicate, leading to intra- and inter-observer variability [35]. In veterinary medicine, it has not yet been demonstrated whether the percentage of tumour necrosis should be determined grossly (which would have to be confirmed microscopically), histologically or both [35]. The current methods for calculating the percent of tumour necrosis are not well defined [34]. In addition, the number of sections of cSTSs that should be obtained in order to get an accurate representation of necrosis is, as yet, unknown [35]. Roccabianca et al. [7] suggested using one tissue block for each 2 cm diameter. The introduction of standardised trimming would aid in reducing the variability. However, this would require the generation of a large number of sections in order to improve accuracy, and thus would significantly add to the workload of the pathologist. The use of machine-learning algorithms in assessing the percentage of tumour necrosis would allow for a larger number of sections to be assessed with minimal input from the pathologist. This approach would allow a larger number of slides to be evaluated and therefore a more representative sample to be checked without impacting on the time taken by the pathologist to report each case. This approach would also help eliminate the current scenario where technicians may avoid creating sections of tumours in areas where they appear necrotic, haemorrhagic or oedematous when trimming samples [35]. Future studies should investigate these areas by including outcome data.

5. Conclusions

In conclusion, DenseNet161 was able to recognise areas of necrosis in cSTSs with high accuracy. This suggests that such an approach could improve the performance of pathologists by offering high sensitivity and reducing inter-observer variability. These results demonstrate that AI can potentially be used as an effective diagnostic support tool to grade cSTSs with more accuracy. Future studies could investigate the accuracy of the prediction maps and correlate the results with patient outcome to better define tumour grade. In addition, researchers could investigate optimising thresholds to improve the sensitivity, specificity, precision and F1-score.

Author Contributions: Conceptualisation, A.M., R.M.L.R., K.W. and T.R.; methodology, A.M. and T.R.; software, T.R.; validation, A.M. and T.R.; formal analysis, A.M. and T.R.; investigation, A.M.; resources, M.J.D., N.J.B. and T.A.; data curation, A.M.; writing—original draft preparation, A.M.; writing—review and editing, T.R., N.J.B., B.B., K.W., M.B., S.A.T., M.J.D., T.A. and R.M.L.R.; visualisation, R.M.L.R.; supervision, R.M.L.R.; project administration, R.M.L.R. and K.W.; All authors have read and agreed to the published version of the manuscript.

Funding: The study was funded by the Doctoral College, University of Surrey (UK) (UoS - DTC PhD studentship awarded to Ambra Morisi), National Physical Laboratory (UK) and Zoetis through the vHive initiative.

Institutional Review Board Statement: For this study a Non-Animals Scientific Procedures Act 1986 (NASPA) form was submitted, and successfully approved by the University of Surrey. The approval number is NERA-1819-045.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets and algorithm code generated for this study are available on request to the corresponding author.

Acknowledgments: This research was supported by the Doctoral College, University of Surrey (UK), National Physical Laboratory (UK) and Zoetis through the vHive initiative.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bostock, D.E.; Dye, M.T. Prognosis after surgical excision of canine fibrous connective tissue sarcoma. *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.* **1988**, *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](#)]
4. Mayer, M.N.; Larue, S.M. Soft tissue sarcomas in dogs. *Can. Vet. J.* **2005**, *46*, 040–1052.
5. Dennis, M.M.; McSporran, K.D.; Bacon, N.J.; Schulman, F.Y.; Foster, R.A.; Powers, B.E. Prognostic factors for cutaneous and subcutaneous soft tissue sarcomas in dogs. *Vet. Pathol.* **2011**, *48*, 73–84. [[CrossRef](#)]
6. Liptak, J.M.; Forrest, L.J. Soft tissue sarcomas. In *Withrow and McEwen's Small Animal Clinical Oncology*, 5th ed.; Withrow, S.J., Vail, D.M., Page, R.L., Eds.; Elsevier: St. Louis, MO, USA, 2013.
7. Roccabianca, P.; Schulman, Y.; Avallone, G.; Foster, R.; Scruggs, J.; Dittmer, K.; Kiupel, M. *Surgical Pathology of Tumors of Domestic Animals: Volume 3: Tumors of Soft Tissue*; Kiupel, M., Ed.; Davis Thompson Foundation: Gurnee, Illinois, 2020.
8. Bray, J.P. Soft tissue sarcoma in the dog—Part 1: A current review. *J. Small Anim. Pract.* **2016**, *57*, 510–519. [[CrossRef](#)] [[PubMed](#)]
9. Coindre, J.M. Grading of soft tissue sarcomas: Review and update. *Arch. Pathol. Lab. Med.* **2006**, *130*, 1448–1453. [[CrossRef](#)]
10. Yap, F.W.; Rasotto, R.; Priestnall, S.L.; Parsons, K.J.; Stewart, J. Intra- and inter-observer agreement in histological assessment of canine soft tissue sarcoma. *Vet. Comp. Oncol.* **2017**, *15*, 1553–1557. [[CrossRef](#)]
11. Kuntz, C.A.; Dernell, W.S.; Powers, B.E.; Devitt, C.; Straw, R.C.; Withrow, S.J. Prognostic factors for surgical treatment of soft-tissue sarcomas in dogs: 75 cases (1986–1996). *J. Am. Vet. Med. Assoc.* **1997**, *211*, 1147–1151.
12. McSporran, K.D. Histologic grade predicts recurrence for marginally excised canine subcutaneous soft tissue sarcomas. *Vet. Pathol.* **2009**, *46*, 928–933. [[CrossRef](#)]
13. Swinson, D.E.; Jones, J.L.; Richardson, D.; Cox, G.; Edwards, J.G.; O'Byrne, K.J. Tumour necrosis is an independent prognostic marker in non-small cell lung cancer: Correlation with biological variables. *Lung Cancer* **2002**, *37*, 235–240. [[CrossRef](#)] [[PubMed](#)]
14. Sharma, H.; Zerbe, N.; Klempert, I.; Lohmann, S.; Lindequist, B.; Hellwich, O.; Hufnagl, P. Appearance-based necrosis detection using textural features and SVM with discriminative thresholding in histopathological whole slide images. In Proceedings of the IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE), Belgrade, Serbia, 2–4 November 2015; pp. 1–6.
15. Kang, J.W.; Shin, S.H.; Choi, J.H.; Moon, K.C.; Koh, J.S.; Jung, C.; Park, Y.K.; Lee, K.B.; Chung, Y.G. Inter- and intra-observer reliability in histologic evaluation of necrosis rate induced by neo-adjuvant chemotherapy for osteosarcoma. *Int. J. Clin. Exp. Pathol.* **2017**, *10*, 359–367.
16. Sharma, H.; Zerbe, N.; Klempert, I.; Hellwich, O.; Hufnagl, P. Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. *Comput. Med. Imaging Graph.* **2017**, *61*, 2–13. [[CrossRef](#)]
17. Fuchs, T.J.; Wild, P.J.; Moch, H.; Buhmann, J.M. Computational pathology analysis of tissue microarrays predict survival of renal clear cell carcinoma patients. In *Medical Image Computing and Computer-Assisted Intervention, Proceedings of the 11th International Conference, New York, NY, USA, 6–10 September 2008*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–8.
18. Yu, K.-H.; Zhang, C.; Berry, G.J.; Altman, R.B.; Ré, C.; Rubin, D.L.; Snyder, M. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* **2016**, *7*, 12474. [[CrossRef](#)]
19. Ertosun, M.G.; Rubin, D.L. Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks. In *AMIA Annual Symposium Proceedings*; American Medical Informatics Association: Bethesda, MD, USA, 2015; Volume 2015, p. 1899.
20. Petushi, S.; Garcia, F.U.; Haber, M.M.; Katsinis, C.; Tozeren, A. Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. *BMC Med. Imaging* **2006**, *6*, 14. [[CrossRef](#)] [[PubMed](#)]
21. Luo, X.; Zang, X.; Yang, L.; Huang, J.; Liang, F.; Rodriguez-Canales, J.; Wistuba, I.I.; Gazdar, A.; Xie, Y.; Xiao, G. Comprehensive computational pathological image analysis predicts lung cancer prognosis. *J. Thorac. Oncol.* **2017**, *12*, 501–509. [[CrossRef](#)] [[PubMed](#)]
22. Cruz-Roa, A.; Basavanthally, A.; González, F.; Gilmore, H.; Feldman, M.; Ganesan, S.; Shih, N.; Tomaszewski, J.; Madabhushi, A. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In *Medical Imaging 2014: Digital Pathology, Proceedings of the SPIE Medical Imaging, San Diego, CA, USA, 15–20 February 2014*; SPIE: Bellingham, WA, USA, 2014; p. 904103.
23. Litjens, G.; Sánchez, C.I.; Timofeeva, N.; Hermsen, M.; Nagtegaal, I.; Kovacs, I.; Hulsbergen-van de Kaa, C.; Bult, P.; Van Ginneken, B.; Van Der Laak, J. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci. Rep.* **2016**, *6*, 26286. [[CrossRef](#)]

24. Cireşan, D.C.; Giusti, A.; Gambardella, L.M.; Schmidhuber, J. Mitosis detection in breast cancer histology images with deep neural networks. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2013, Proceedings of the 16th International Conference, Nagoya, Japan, 22–26 September 2013*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 411–418.
25. Hou, L.; Samaras, D.; Kurc, T.M.; Gao, Y.; Davis, J.E.; Saltz, J.H. Patch-based Convolutional Neural Network for Whole Slide Tissue Image Classification. *arXiv* **2016**, arXiv:1504.07947.
26. Arunachalam, H.B.; Mishra, R.; Daescu, O.; Cederberg, K.; Rakheja, D.; Sengupta, A.; Leonard, D.; Hallac, R.; Leavey, P. Viable and necrotic tumor assessment from whole slide images of osteosarcoma using machine-learning and deep-learning models. *PLoS ONE* **2019**, *14*, e0210706. [[CrossRef](#)]
27. 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](#)]
28. Rai, T.; Morisi, A.; Bacci, B.; Bacon, N.J.; Thomas, S.; La Ragione, R.; Bober, M.; Wells, K. Can imagenet feature maps be applied to small histopathological datasets for the classification of breast cancer metastatic tissue in whole slide images? *Proc. SPIE* **2019**, *10956*, 109560V.
29. Rai, T.; Morisi, A.; Bacci, B.; Bacon, N.J.; Thomas, S.; La Ragione, R.; Bober, M.; Wells, K. An investigation of aggregated transfer learning for classification in digital pathology. *Proc. SPIE* **2019**, *10956*, 109560U.
30. Talo, M. Automated Classification of Histopathology Images Using Transfer Learning. *Artif. Intell. Med.* **2019**, *101*, 101743. [[CrossRef](#)]
31. Huang, G.; Liu, Z.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
32. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Li, F.-F. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
33. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015.
34. 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](#)] [[PubMed](#)]
35. Meuten, D.J.; Moore, F.M.; Donovan, T.A.; Bertram, C.A.; Klopffleisch, R.; Foster, R.A.; Smedley, R.C.; Dark, M.J.; Milovancev, M.; Stromberg, P.; et al. International Guidelines for Veterinary Tumor Pathology: A Call to Action. *Vet. Pathol.* **2021**, *58*, 766–794. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.