Article

Canine Mammary Tumor Histopathological Image Classification via Computer-Aided Pathology: An Available Dataset for Imaging Analysis

Giovanni P. Burrai [1,2], **Andrea Gabrieli** [1], **Marta Polinas** [1,*], **Claudio Murgia** [1], **Maria Paola Becchere** [3], **Pierfranco Demontis** [4] and **Elisabetta Antuofermo** [1,2]

[1] Department of Veterinary Medicine, University of Sassari, Via Vienna 2, 07100 Sassari, Italy; gburrai@uniss.it (G.P.B.); claudio97murgia@gmail.com (C.M.)

[2] Mediterranean Center for Disease Control (MCDC), University of Sassari, Via Vienna 2, 07100 Sassari, Italy

[3] Independent Researcher, 07100 Sassari, Italy

[4] Department of Chemical, Physical, Mathematical and Natural Sciences, University of Sassari, Via Vienna 2, 07100 Sassari, Italy; demontis@uniss.it

[*] Correspondence: mpolinas@uniss.it; Tel.: +39-079-229-440

Simple Summary: Digital pathology (DP) and computer-aided diagnosis (CAD) are rapidly evolving fields that have great potential for improving the accuracy and efficiency of cancer diagnosis, including that of canine mammary tumors (CMTs), the most common neoplasm in female dogs. The work presents a study on the development of CAD systems for the automated classification of CMTs utilizing convolutional neural networks (CNNs) to extract features from histopathological images of CMTs and classify them into benign or malignant tumors. The study shows that the proposed framework can accurately distinguish between benign and malignant CMTs, with testing accuracies ranging from 0.63 to 0.85. The study emphasizes how digital pathology and CAD could help veterinarians and pathologists in accurately diagnosing the tumor type, which is crucial in determining the optimal course of treatment. Overall, digital pathology and CAD are promising tools that could improve the accuracy and efficiency of cancer diagnosis, including that of canine mammary tumors.

Abstract: Histopathology, the gold-standard technique in classifying canine mammary tumors (CMTs), is a time-consuming process, affected by high inter-observer variability. Digital (DP) and Computer-aided pathology (CAD) are emergent fields that will improve overall classification accuracy. In this study, the ability of the CAD systems to distinguish benign from malignant CMTs has been explored on a dataset—namely *CMTD*—of 1056 hematoxylin and eosin JPEG images from 20 benign and 24 malignant CMTs, with three different CAD systems based on the combination of a convolutional neural network (VGG16, Inception v3, EfficientNet), which acts as a feature extractor, and a classifier (support vector machines (SVM) or stochastic gradient boosting (SGB)), placed on top of the neural net. Based on a human breast cancer dataset (i.e., BreakHis) (accuracy from 0.86 to 0.91), our models were applied to the CMT dataset, showing accuracy from 0.63 to 0.85 across all architectures. The EfficientNet framework coupled with SVM resulted in the best performances with an accuracy from 0.82 to 0.85. The encouraging results obtained by the use of DP and CAD systems in CMTs provide an interesting perspective on the integration of artificial intelligence and machine learning technologies in cancer-related research.

Keywords: breast cancer; canine mammary tumor (CMTs); CMT dataset; deep learning; machine learning; histological classification

1. Introduction

Cancer is the leading cause of death in companion animals, and canine mammary tumor (CMTs), the most common neoplasm in female dogs, represents a serious issue in worldwide veterinary practice [1–4]. Therefore, an increased number of studies in this area have been published in the last decades. As in animals, human breast cancer (HBC) is the most common malignancy among women worldwide, sharing several clinical and molecular similarities with canine lesions [5–7]. Consequentially, dogs have attracted considerable attention as potential animal models to study human cancer [8].

Detection and diagnosis of mammary tumors, alongside a clinical examination, can be accomplished via imaging procedures such as diagnostic mammograms, ultrasound, and magnetic resonance imaging [1,9], although histopathological analysis remains the gold standard for differentiating between benign and malignant neoplasms [1,3].

Histopathological analysis is a time-consuming process, requiring highly trained specialists, and could be influenced by several intrinsic and extrinsic factors, including adequate specimen fixation, laboratory handling, and the pathologists' experience [10]. In histopathology, a high percentage of cancer can be diagnosed by pathologists using hematoxylin and eosin (H&E)-stained slides. Furthermore, diagnosis based upon manual analysis of slides is prone to inter-observer variability, with approximately 75% diagnostic concordance between specialists [3,11–13]. Digital pathology (DP) is a significant modernization that changes the paradigm of microscope-based pathology, replacing the microscope with the computer screen and changing storage media from glass slides to digitalized image files [14]. Digitalized images stored in computer servers or cloud systems can be easily transmitted, thus changing the temporal and spatial domain of pathologic diagnosis [15]. Moreover, digitalized images can be further analyzed by the so-called computer-aided pathology (CAP), referred to as a computational diagnosis system or a set of methodologies that utilize computers or software to interpret pathologic images [14–17]. CAD systems using machine learning algorithms have been demonstrated to improve classification accuracy and reduce variability in interpretations, increasing the level of inter-observer agreement. Several validation studies have compared the diagnostic accuracy of DP and conventional microscopic diagnosis in the last decade [18,19]. In addition, these techniques are also useful for assisting pathologists and reducing their effort in localizing and identifying abnormalities in cancer tissue images.

In recent years, the increase in computing power due to the spread of parallel architectures based on graphical processing units (GPU) has boosted the emergence of deep learning algorithms. In particular, convolutional neural networks (CNNs) have become the elected method in the field of image analysis [20–25] and a powerful tool in the automated classification of human cancer histopathology images [26–33]. CNNs are particularly well-suited and efficient at processing data that manifest local spatial correlation with grid-like topologies [21,22]. The fundamental element is the so-called convolutional layer. In its simplest form, such a layer consists of several learnable weight matrices (kernels) of small spatial dimensions (i.e., a typical kernel for the modern architecture has a size of 3×3: 3 pixels wide and 3 pixels high). Kernels are convoluted with the input data, thus generating two-dimensional activation maps. This allows the discovery of particular aspects of the input data, and the weights are efficiently reused wherever a particular feature is located in an image. By modifying the weights, the network can learn patterns and pattern hierarchies and then learn to distinguish images.

CNNs automatically learn mid- and high-level abstractions obtained from RGB images [34], and, along with multiple-instance learning, have accomplished high performance in the binary classification of human cancers and have evolved as one method for analyzing histopathological images [35].

Despite canine mammary tumors representing a serious issue in worldwide veterinary practice, no consistent efforts have been made to automate the diagnosis of CMTs. In this study, a canine mammary tumor image database comprising images captured from 44 cases of CMTs explored with three different CNN architectures, namely VGG16, Inception v3,

and EfficientNet, associated with support vector machines (SVM) and stochastic gradient boosting (SGB) was used to investigate the ability to distinguish benign and malignant tumors on H&E-stained images, based on histopathological analysis as a gold standard. Furthermore, the models were tested on a standard human breast cancer dataset (BreakHis) and the effects of data augmentation on the performance of the proposed framework were also analyzed. Thus, a complete novel dataset, namely *CMTD*, of the most common benign and malignant mammary canine tumors was provided.

2. Materials and Methods

2.1. Canine Mammary Tumor Dataset

The canine mammary tumor image dataset (*CMTD*) is comprises 1056 H&E-stained JPEG images of a size of 1024 × 768 (width × height), acquired from 44 canine mammary tumors that were submitted to the Departments of Veterinary Medicine of the University of Sassari (UNISS). Tissue samples were fixed in 10% neutral buffered formalin, paraffin-embedded and H&E-stained for histopathological analysis. The histopathological classification of CMT tissues was performed in accordance with the recent publication of *Surgical Pathology of Tumors of Domestic Animals, Volume 2: Mammary Tumors* [36] by one board-certified and two experienced veterinary pathologists. A recorded video of a mean time of 2 min was performed for each sample. From each video, a pool of frames was programmatically chosen, and from this pool, the pathologists selected the best 24 images at a 400× magnification. Thus, 1024 × 768 high-resolution RGB images with a 24 bit color depth were captured, comprising a total of 1056 images from 20 benign and 24 malignant CMT cases (Supplementary Material: Canine Mammary Tumor Dataset—*CMTD*). Images and videos were obtained from an optical microscope (Optica c-B3) equipped with a digital camera (C-B3 optikamB3 digital camera).

Experimental permission was not required from the University's Animal Care Ethics Committee since all the samples were retrieved from the archive of the pathology laboratories and were used for diagnostic purposes.

2.2. Breast Cancer Dataset

The proposed framework was first evaluated on the standard and challenging BreakHis dataset, which is freely available (https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/ (accessed on 5 April 2021 comprising 9109 images with different magnifying factors from human breast cancer patients. The dataset included 2480 benign (fibroadenoma, tubular adenoma, and phyllodes tumor) and 5429 malignant images of breast tumors (lobular carcinoma, ductal carcinoma, papillary carcinoma, and mucinous carcinoma) [37,38].

2.3. Data Processing

To estimate the generalization error in both CMTD and BreakHis datasets, a nested 5-fold cross-validation procedure was employed [39]. The data were partitioned into five different non-overlapping training and test sets (outer-cv), so that all images belonging to the same histological slide (i.e., patient) fell in either the test or the training set but not both, to avoid information leaking [37,39,40]. For each of the aforementioned training sets, further 5-fold cross-validation (inner-cv) was performed for model tuning and selection. The tuning was performed with a random grid search. The final result was obtained by averaging the test sets of the outer loop. For details, the reader can refer to the scikit-learn documentation and the python code accompanying this work available at https://github.com/cpctools/CMTD_classification (accessed on 1 March 2023) [41].

2.4. Data Augmentation

The images were subjected to data augmentation to increase the data size and make our model robust for feature transformation. Since we used CNN as a feature extractor, the features were generated and stored before the training of the classifier. Two different

strategies of data augmentation were employed. The first one (i.e., the *base* strategy) consisted of resizing the image to 512 pixels (px), then randomly cropping it to 224 px for VGG (see below), 299 px for Inception (see below), and 380 px for EfficientNet (see below) (default crop size accepted by each model) [24], random rotation by 0, 90, 180, and 270 degrees, and random vertical reflection, similarly to what was described by Araujo et al., 2017, and Kumar et al., 2020 [30,42]. The second strategy employed (i.e., the *advanced* strategy) consisted of taking a crop of a random size between 0.08 and 1.0 of the original size and a random aspect ratio of 3/4 or 4/3 and then resizing it to 224 px for VGG, 299 px for Inception, and 380 px for EfficientNet [43]. The same random rotation and vertical reflection methods described for the first strategy were used. Finally, a random change of $\pm$ 20% of the brightness, contrast, and saturation was applied. In both cases, the images were normalized using ImageNet statistics [44]. Features extracted with the CNN were then standard-scaled (with statistics from each training set in both cross-validation loops) before the classification. The predictions were performed by using the center crop of the image (i.e., cropping the center portion of an image and using it as a new image for evaluation) or the ten crops [24], where the central crop and the four corners, as well as the horizontal flip of these five, were separately evaluated, and then the decision was made via majority voting.

2.5. Convolutional Neural Networks (CNN)

Three different CNN architectures for feature extraction, namely VGG16 [45], Inception v3 [46], and EfficientNet [47] were employed and compared to each other. The VGG16 architecture marked a fundamental step in the evolution of CNNs. By increasing the depth of the network and reducing the size of the filters, its proposers have obtained excellent results in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC, the test bench on which the new architectures proposed are compared) [48]. This architecture has simplicity among its strengths but has a high computational cost. More specifically, for VGG16 architecture features are extracted after each convolutional block just before the max-pooling layer, as reported in the Supplementary Materials (Figure S1).

Inception v3 is an architecture with great success and large performance improvement compared to the state of the art at a relatively low computational cost. It is an evolution of the so-called GoogLeNet [43]. The key to its success is the inception module which replaces a simple convolution with a composition of kernels of different sizes and pooling operations leading to a reduction in the total number of parameters and an increase in performance. For Inception v3, the extraction happens after the first two convolutional blocks and the first two inception blocks as shown in the Supplementary Materials (Figure S2).

The EfficientNet is a very recent architecture showing state-of-the-art performance. The building block of this architecture is a mobile inverted-bottleneck MBConv with added squeeze-and-excitation optimization [49]. It is built starting from a baseline network called EfficientNet-B0 (obtained using a neural architecture search [50]), and then scaled up to find an optimal combination of depth (i.e., number of layers), width (i.e., number of channels), and resolution of the input image. Features are extracted from several intermediate convolutional maps, a global average pooling (GAP) is applied along spatial dimensions, and the obtained feature vectors are concatenated [42,51,52]. This approach allows an exploitation of the hierarchical nature of CNN where each filter becomes sensitive to a different pattern via shifting the attention from global to local structures [51,53,54]. For EfficientNet-B4, the extraction occurs before every channel increase as depicted in the Supplementary Materials (Figure S3).

The extracted features are used as input for machine learning algorithms, such as support vector machines (SVMs) and stochastic gradient boosting (SGB), to perform classification.

2.6. Support Vector Machines

SVMs are a class of algorithms that were developed in the 1990s for classification tasks and have seen successful application in many areas, including breast cancer predictions [22,26,28,30,37,54–57]. The idea behind these kinds of algorithms is to find an optimal separating hyperplane between classes. In the simplest case, when classes are perfectly separable, the hyperplane is chosen to be the farthest from the observations; this leads to the maximal margin classifier, where the margin is the minimal distance from the observations to the hyperplane. The points touching the margin are known as support vectors and are the only points that affect the classifier. In the non-separable case, the choice is performed by maximizing the number of correctly assigned classes. This can be carried out with the use of a soft margin, i.e., letting some observations be misclassified. The tolerance toward violations of the margin can be tuned by using a non-negative hyperparameter, usually named C. This approach is the basis of the support vector classifier. Finally, a non-linear decision boundary is necessary, in the most complex cases, to separate classes. A general approach to cope with this situation consists of enlarging the feature space. This can be achieved using basis expansions such as polynomials, but computations can become quickly prohibitive. In SVMs, the expansion of the feature space is accomplished efficiently with kernels, which are functions that quantify the similarity between two observations. Typical kernels are polynomial or radial. In our work, we employed linear, degree three polynomial, and radial basis functions to support vector classifiers. Adding to this, and to reduce the computational burden, for polynomial and radial basis kernels (RBK), the Nystroem method [55,56], which approximates the kernel by subsampling the data, was employed.

2.7. Stochastic Gradient Boosting

The second technique employed was boosting [22,58], which is a powerful learning method based on the combination of many simple models. The basic idea is to sequentially apply a "weak" learner to modified versions of the initial data. In our work, decision trees were used as weak classifiers, but any method better than random guessing can be used. Each time a tree is built, the data were modified by applying weights to increase the influence of misclassified observations. The final classification was performed through a weighted majority vote [59,60]. This basic idea can be improved by using a stagewise gradient descent procedure [61] and by incorporating randomness by subsampling the data at each iteration. This leads to the stochastic gradient boosting methods employed here [58].

3. Results

3.1. Canine Mammary Tumors

At histopathology, a morphologically heterogeneous group of lesions were classified as follows: 20 benign tumors, including 8 benign mixed tumors, 9 complex adenomas, 3 simple adenomas, and 24 malignant neoplasms, including 11 complex and 13 simple carcinomas (4 tubular, 5 tubulopapillary, 3 solid and 1 comedocarcinoma). The representative H&E-stained images from the different CMTs showing typical benign and malignant CMTs are illustrated in Figure 1. The proposed framework was first evaluated on a standard and challenging BreakHis dataset comprising 9109 images with different magnifying factors from human breast cancer patients.
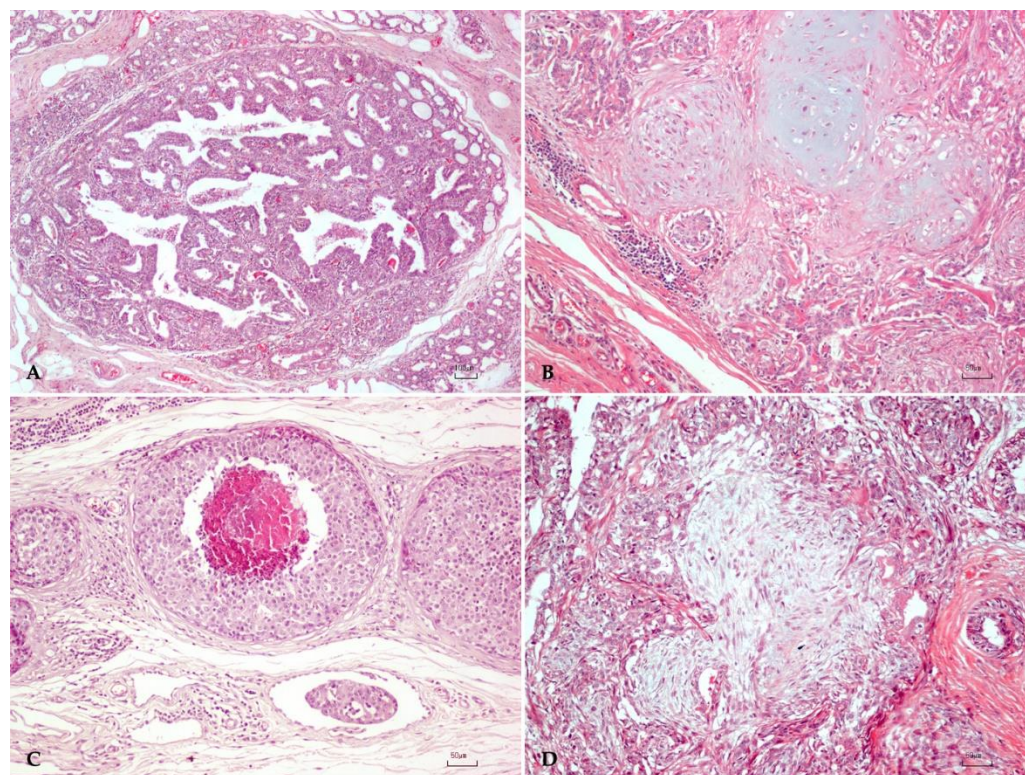
Figure 1. Canine mammary tumor. (**A**) Simple mammary adenoma characterized by a nodular, well-demarcated, focal neoplasm composed of luminal epithelial cells arranged in tubules; H&E, bar: 100 μm. (**B**) Benign mixed tumor characterized by the proliferation of luminal epithelial cells admixed with spindle myoepithelial cells and foci of cartilage; H&E, bar: 50 μm. (**C**) Simple comedocarcinoma characterized by a central necrotic area surrounded by neoplastic luminal epithelial; H&E, bar: 50 μm. (**D**) Complex carcinoma with the proliferation of malignant luminal epithelial cells and benign spindle myoepithelial cells; H&E, bar: 50 μm.

3.2. Performance of the Convolutional Neural Networks Models

The accuracy and the performances of the proposed framework were first validated on the BreakHis dataset comprising a large number of human breast tumor images [37,38]. As a result, test accuracies for distinguishing benign and malignant tumors upon H&E classification of the BreakHis dataset ranged from 0.86 to 0.91, considering all combinations of feature extractors, classifiers, and testing strategies (Table 1 and Table S1 in Supplementary Materials).

Table 1. BreakHis dataset; the mean center-crop and the ten-crop accuracies for the best combination of feature extractor, classifier and augmentation strategy.

Feature Extractor	Classifier	Augmentation	Center Crop Accuracy	Ten Crops Accuracy
VGG16	Linear SVM	Advanced 1×	0.89 ± 0.01	0.91 ± 0.01
Inception	Linear SVM	Advanced 6×	0.88 ± 0.02	0.91 ± 0.02
EfficientNet	RBF SVM	Base 6×	0.9 ± 0.02	0.91 ± 0.02

As reported in Table 1, the best performance was observed using EfficientNet as a feature extractor followed by a SVM algorithm with radial basis kernels and with the simple augmentation strategy being repeated 6 times, with a testing accuracy of 0.9 to 0.91 for the center-crop and ten-crop testing, respectively. Furthermore, VGG16 and Inception coupled with a linear SVM, both with simple and advanced repeated 6-time augmentation strategies, resulted in a comparable accuracy when ten-crop testing was used.

The application of the same strategies in the CMTD resulted in mean testing accuracies ranging from 0.63 to 0.84 for the single-crop testing and from 0.64 to 0.85 for the ten-crop testing across all architectures (Table 2 and Table S1 in Supplementary Materials).

Table 2. CMT dataset: the mean center-crop and the ten-crop accuracies for the best combination of feature extractor, classifier, and augmentation strategy.

Feature Extractor	Classifier	Augmentation	Center Crop Accuracy	Ten Crops Accuracy
VGG16	Linear SVM	Base 6×	0.78 ± 0.03	0.82 ± 0.03
Inception	Linear SVM	Advanced 6×	0.78 ± 0.03	0.81 ± 0.04
EfficientNet	Poly SVM	Base 6×	0.84 ± 0.02	0.84 ± 0.03
	RBF SVM	Base 6×	0.83 ± 0.03	0.85 ± 0.03

In particular, the framework using EfficientNet as a feature extractor coupled with a SVM and the simple augmentation strategy being used and repeated six times, resulted in the best performing one, with testing accuracies ranging from 0.83 and 0.84 for the center-crop testing and from 0.84 to 0.85 for the ten-crop testing. The other two tested architectures have similar performances to each other, with VGG16 being slightly ahead but still inferior to EfficientNet, probably as it works with a greater input size that could help to grasp more features.

The data augmentation strategies improve the CNNs' performance, especially for the more powerful classifiers, while the more complicated approach requires the generation of a higher number of images to become closer to the others. As for the testing strategy, there is a slight advantage of the ten-crop method with respect to the center-crop method.

4. Discussion

Considering the importance of a histopathological diagnosis in the management of oncologic patients, considerable efforts have been made for developing robust, precise, and automated CAD systems for humans. More specifically, CNNs are becoming the standard approach for the classification of histological images related to breast cancer [62]. In veterinary medicine, the increase in the incidence of neoplastic disease represents a relentless challenge for veterinary oncology specialists and many efforts have been made in the ongoing research to increase the earliness of diagnosis and the survival time in dogs harboring mammary tumors [1–9,63]. However, considering the high incidence of canine mammary tumors, no significant effort has been made in veterinary pathology for the development of CMT-oriented CAD systems.

In this work, three different CNN architectures (VGG16, Inception v3, and Efficient-Net), coupled with two different classifiers (support vector machines and stochastic gradient boosting), were tested and used to explore the ability to distinguish between benign and malignant canine mammary tumors on hematoxylin–eosin-stained images.

The application of the abovementioned developed architectures on the public dataset of the Breast Cancer Histopathological Database (BreakHis) comprising microscopic images of human breast tumors, with each sample labeled as either benign or malignant [8], provided a 91% classification accuracy rate. Interestingly, Li and coauthors in 2020 achieved an accuracy rate of 83%, [64], while Kumar in 2020, with a fused framework based on VGG16 CNN, used as a feature extractor for different classifiers, obtained an accuracy of 97% [42]. More recently, Liu in 2021 with concatenated VGG16, based on filtering and denoising the BreakHis images, obtained a 98.89% accuracy [65].

In the present work, the application of the same strategies to the CMT dataset resulted in mean testing accuracies ranging from 0.63 to 0.84 for single-crop testing, and from 0.64 to 0.85 for ten-crop testing across all architectures. Interestingly, Kumar and coauthors proposed a dataset of CMT histopathological images, namely CMTHis, comprising a total of 352 images from 20 benign and 24 malignant CMT cases, evaluated using a framework

based on VGGNet-16 coupled with a SVM and Random Forest, with different strategies of data augmentation, obtaining a diagnostic accuracy of 93% [42].

Furthermore, differently from our dataset, the CMTHis dataset consisted only of simple and ductal-associated, except for fibroadenoma, canine mammary tumors, while in our cases 17 benign and 11 malignant neoplasms were of the complex type, and 16 were of the simple type, reflecting the frequencies of histotypes commonly diagnosed in dogs. In particular, mixed neoplasms are the most frequent neoplasia in female dogs and are characterized by the proliferation of both luminal epithelial and interstitial myoepithelial elements admixed with foci of mesenchymal tissues such as cartilage, bone, and fat [36,66]. Thus, the different diagnostic accuracy obtained in our work could be related to the several different morphologies present in our dataset, underlining the importance of considering the complexity of histological images in veterinary medicine, in which mixed neoplasms are common.

However, our study displayed similar results to those of Kumar when analyzing images at the same magnification (i.e., 400×), ranging from 81.63 to 83.35 accuracy [42]. In machine learning, studies with inadequate samples suffer from overfitting of data, while the increase in sample size increases the accuracy of prediction. as suggested by Rajput and coauthors [67]. In our database, a great number of images were evaluated (1056 images from 44 CMT instead of the 88 CMTHis pictures), supporting the validity of our model.

In addition, the lower accuracy using the BreakHis image dataset of the evaluated framework, compared to that using the *CMTD* and CMTHis databases, could be related to the low number of canine mammary tumor images compared to breast images [42].

Several studies have examined the role of data augmentation in deep learning, as this method generates a high amount of data and the building of more generalized models. In particular, Spanhol and collaborators, using CNNs to classify images from the BreakHis database [37,38], employed an approach based on the extraction and the combination of patches of the original high-resolution images [26,37,38].

Moreover, Spanhol and collaborators made further improvements by combining different CNNs, outperforming previous machine learning approaches based on hand-crafted textural descriptors. In subsequent work, the authors employed the CNN as a feature extractor to train a classifier, showing again how classical approaches are outperformed, as are task-specific CNNs sometimes [26,28,37]. Araújo and coauthors in 2017, proposing a custom CNN to access information at different scales, obtained good results for the Bioimaging 2015 challenge dataset, both with the pure CNN and using the CNN as a feature extractor coupled with a SVM classifier [30]. Han and collaborators in 2017 developed a class structure-based deep convolutional neural network, which can be trained end-to-end and does not rely on patches [32]. The applied network dramatically improved the accuracy of previous models in both binary and multi-class tasks for the BreakHis dataset. Alom in 2019 proposed a method based on the inception recurrent residual convolutional neural network model for binary and multi-class classification tasks [33]. They explored different training and analysis strategies, obtaining good results over different benchmark datasets (e.g., accuracies larger than 97% for the BreakHis binary classification task for all magnifications).

In our study, the accuracy and the positive and negative predictive values using the EfficientNet-B4 architecture as a feature extractor had the best performances, albeit presenting only a slight difference, among the three tested options. VGG16 showed a slightly superior accuracy of performance compared to that of InceptionNetV3, but it was still inferior to that of EfficientNet, probably as the latter is more refined and works with a greater input size that could help to grasp more features.

This data are in agreement with what has been recently reported by Kallipolitis, where EfficientNets architectures, and in particular B4, have the best performances compared to InceptionNetV3 and VGG16 [68].

Data augmentation is a useful technique to enhance the performance of CNNs, especially when the available data (i.e., images) are limited or the model's performance needs

to be improved [69]. In our work, the data augmentation strategies discussed improve the CNNs' performance, especially for the more powerful classifiers. In particular, the simpler data augmentation strategy, consisting only of a random crop, rotations, and reflections, generally leads to slightly better results. The more complicated approach used in our work instead requires the generation of a higher number of images to obtain a similar performance to that of the others. An exception is the case of the linear support vector classifier, which gives good results across all options as previously described by Kumar 2020 [42]. In our study, an aggressive data augmentation strategy did not lead to substantial improvements but rather worsened the results when the number of artificially modified images was small, as in the case of VGG16, a situation that was reversed when considering the InceptionNetV3 architecture.

As for the testing strategy, there is a slight advantage of the ten-crop method compared to the center-crop method. This is tentatively explained by the fact that the ten crops strategy is probably able to acquire more information, as all regions of the image are analyzed.

Overall, our results suggest that CAD systems using deep learning techniques, specifically convolutional neural networks, have great potential for improving the accuracy of histopathological diagnoses in both human and veterinary medicine. Although the accuracy rates were not like those in human breast cancer histopathology, our results are promising, also taking in consideration the limited number of studies regarding CMT. Furthermore, we provide the larger canine mammary tumor benchmark dataset—namely *CMTD*—containing over 1000 high-resolution histological images of the most common benign and malignant tumors for developing and evaluating computer-aided diagnostic systems for testing other state-of-the-art models for histopathological image classification.

5. Conclusions

Convolutional neural networks have been used in various applications related to breast cancer, such as in medical imaging for the diagnosis and prognosis of breast cancer. In veterinary medicine, several manuscripts have described the potentiality of this technology, but no consistent efforts have been undertaken regarding canine mammary tumors [69–75]. Deep learning-based algorithms can assist the pathologist in classifying tumors on standard hematoxylin and eosin images. Therefore, publicly available datasets have become increasingly popular, as they reduce annotation costs for recurring pathological research questions and improve the comparability of computer-aided systems developed on these datasets. Overall, the results of this study demonstrate the potential of CNNs and CAD systems to aid in the diagnosis of canine mammary tumors, while the available canine mammary tumor benchmark dataset will be of great benefit to the veterinary research community. Therefore, further studies with a large number of CMT patients and histopathological images are required to prove the efficacy of the proposed framework in the binary classification of CMTs.

In conclusion, the encouraging results obtained in this study provide an interesting perspective on the integration of artificial intelligence and machine learning technologies in cancer-related research, offering a valuable starting point for further research in this area.

References

1. Sorenmo, K.U.; Worley, D.R.; Zappulli, V. Tumors of the Mammary Gland. In *Withrow & MacEwen's Small Animal Clinical Oncology*, 6th ed.; Withrow, S.J., Vail, D.V., Thamm, D.H., Liptak, J.M., Eds.; Elsevier: St. Louis, MO, USA, 2020; pp. 604–615.
2. Sleeckx, N.; de Rooster, H.; Veldhuis Kroeze, E.J.; Van Ginneken, C.; Van Brantegem, L. Canine mammary tumours, an overview. *Reprod. Domest. Anim.* **2011**, *46*, 1112–1131. [CrossRef] [PubMed]
3. Papparella, S.; Crescio, M.I.; Baldassarre, V.; Brunetti, B.; Burrai, G.P.; Cocumelli, C.; Grieco, V.; Iussich, S.; Maniscalco, L.; Mariotti, F.; et al. Reproducibility and Feasibility of Classification and National Guidelines for Histological Diagnosis of Canine Mammary Gland Tumours: A Multi-Institutional Ring Study. *Vet. Sci.* **2022**, *9*, 357. [CrossRef]
4. Burrai, G.P.; Gabrieli, A.; Moccia, V.; Zappulli, V.; Porcellato, I.; Brachelente, C.; Pirino, S.; Polinas, M.; Antuofermo, E. A Statistical Analysis of Risk Factors and Biological Behavior in Canine Mammary Tumors: A Multicenter Study. *Animals* **2020**, *10*, 1687. [CrossRef] [PubMed]
5. Antuofermo, E.; Miller, M.A.; Pirino, S.; Xie, J.; Badve, S.; Mohammed, S.I. Spontaneous Mammary Intraepithelial Lesions in Dogs—A Model of Breast Cancer. *Cancer Epidemiol. Biomark. Prev.* **2007**, *16*, 2247–2256. [CrossRef] [PubMed]
6. Mouser, P.; Miller, M.A.; Antuofermo, E.; Badve, S.S.; Mohammed, S.I. Prevalence and Classification of Spontaneous Mammary Intraepithelial Lesions in Dogs without Clinical Mammary Disease. *Vet. Pathol.* **2010**, *47*, 275–284. [CrossRef]
7. Burrai, G.P.; Tanca, A.; De Miglio, M.R.; Abbondio, M.; Pisanu, S.; Polinas, M.; Pirino, S.; Mohammed, S.I.; Uzzau, S.; Addis, M.F.; et al. Investigation of HER2 expression in canine mammary tumors by antibody-based, transcriptomic and mass spectrometry analysis: Is the dog a suitable animal model for human breast cancer? *Tumor Biol.* **2015**, *36*, 9083–9091. [CrossRef]
8. Abdelmegeed, S.M.; Mohammed, S. Canine mammary tumors as a model for human disease. *Oncol. Lett.* **2018**, *15*, 8195–8205. [CrossRef]
9. Mohammed, S.; Meloni, G.; Parpaglia, M.P.; Marras, V.; Burrai, G.; Meloni, F.; Pirino, S.; Antuofermo, E. Mammography and Ultrasound Imaging of Preinvasive and Invasive Canine Spontaneous Mammary Cancer and Their Similarities to Human Breast Cancer. *Cancer Prev. Res.* **2011**, *4*, 1790–1798. [CrossRef]
10. Veta, M.; Pluim, J.P.W.; van Diest, P.J.; Viergever, M.A. Breast Cancer Histopathology Image Analysis: A Review. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 1400–1411. [CrossRef]
11. Chu, P.Y.; Liao, A.T.; Liu, C.H. Interobserver Variation in the Morphopathological Diagnosis of Canine Mammary Gland Tumor Among Veterinary Pathologists. *Int. J. Appl. Res. Vet. Med.* **2011**, *9*, 388–391.
12. Santos, M.; Correia-Gomes, C.; Santos, A.; de Matos, A.; Dias-Pereira, P.; Lopes, C. Interobserver Reproducibility of Histological Grading of Canine Simple Mammary Carcinomas. *J. Comp. Pathol.* **2015**, *153*, 22–27. [CrossRef] [PubMed]
13. Evans, A.J.; Brown, R.W.; Bui, M.M.; Chlipala, B.E.A.; Lacchetti, M.C.; Milner, J.D.A.; Pantanowitz, L.; Parwani, A.V.; Reid, M.K.; Riben, M.W.; et al. Validating Whole Slide Imaging Systems for Diagnostic Purposes in Pathology: Guideline Update From the College of American Pathologists in Collaboration With the American Society for Clinical Pathology and the Association for Pathology Informatics. *Arch. Pathol. Lab. Med.* **2021**, *146*, 440–450. [CrossRef] [PubMed]
14. Pallua, J.; Brunner, A.; Zelger, B.; Schirmer, M.; Haybaeck, J. The future of pathology is digital. *Pathol. Res. Pract.* **2020**, *216*, 153040. [CrossRef] [PubMed]
15. Nam, S.; Chong, Y.; Jung, C.K.; Kwak, T.-Y.; Lee, J.Y.; Park, J.; Rho, M.J.; Go, H. Introduction to digital pathology and computer-aided pathology. *J. Pathol. Transl. Med.* **2020**, *54*, 125–134. [CrossRef]
16. Madabhushi, A.; Lee, G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med. Image Anal.* **2016**, *33*, 170–175. [CrossRef]

17. Amerikanos, P.; Maglogiannis, I. Image Analysis in Digital Pathology Utilizing Machine Learning and Deep Neural Networks. *J. Pers. Med.* **2022**, *12*, 1444. [CrossRef] [PubMed]

18. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef]

19. Cruz-Roa, A.; Gilmore, H.; Basavanhally, A.; Feldman, M.; Ganesan, S.; Shih, N.N.; Tomaszewski, J.; González, F.A.; Madabhushi, A. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. *Sci. Rep.* **2017**, *7*, srep46450. [CrossRef]

20. Le Cun, Y.; Jackel, L.; Boser, B.; Denker, J.; Graf, H.; Guyon, I.; Henderson, D.; Howard, R.; Hubbard, W. Handwritten digit recognition: Applications of neural network chips and automatic learning. *IEEE Commun. Mag.* **1989**, *27*, 41–46. [CrossRef]

21. Goodfellow, I.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2017; pp. 1–710.

22. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012; pp. 1–1049.

23. Rawat, W.; Wang, Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput.* **2017**, *29*, 2352–2449. [CrossRef]

24. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

25. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Correction: Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *546*, 686. [CrossRef] [PubMed]

26. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. Breast cancer histopathological image classification using Convolutional Neural Networks. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 2560–2567.

27. Yang, Y.; Guan, C. Classification of histopathological images of breast cancer using an improved convolutional neural network model. *J. X-ray Sci. Technol.* **2022**, *30*, 33–44. [CrossRef] [PubMed]

28. Spanhol, F.A.; Oliveira, L.S.; Cavalin, P.R.; Petitjean, C.; Heutte, L. Deep features for breast cancer histopathological image classification. In Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; pp. 1868–1873. [CrossRef]

29. Gupta, V.; Bhavsar, A. Sequential Modeling of Deep Features for Breast Cancer Histopathological Image Classification. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 2335–23357. [CrossRef]

30. Araújo, T.; Aresta, G.; Castro, E.M.; Rouco, J.; Aguiar, P.; Eloy, C.; Polónia, A.; Campilho, A. Classification of breast cancer histology images using Convolutional Neural Networks. *PLoS ONE* **2017**, *12*, e0177544. [CrossRef] [PubMed]

31. Nawaz, M.A.; Sewissy, A.A.; Soliman, T.H.A. Automated Classification of Breast Cancer Histology Images Using Deep Learning Based Convolutional Neural Networks. *Int. J. Comput. Sci. Netw. Secur.* **2018**, *18*, 152–160.

32. Han, Z.; Wei, B.; Zheng, Y.; Yin, Y.; Li, K.; Li, S. Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model. *Sci. Rep.* **2017**, *7*, 4172. [CrossRef]

33. Alom, Z.; Yakopcic, C.; Nasrin, M.S.; Taha, T.M.; Asari, V.K. Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network. *J. Digit. Imaging* **2019**, *32*, 605–617. [CrossRef]

34. Cruz, J.A.; Wishart, D.S. Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Inform.* **2007**, *2*, 59–77. [CrossRef]

35. Yusoff, M.; Haryanto, T.; Suhartanto, H.; Mustafa, W.A.; Zain, J.M.; Kusmardi, K. Accuracy Analysis of Deep Learning Methods in Breast Cancer Classification: A Structured Review. *Diagnostics* **2023**, *13*, 683. [CrossRef]

36. Zappulli, V.; Pena, L.; Rasotto, R.; Goldschmidt, M.H.; Gama, A.; Scruggs, J.L.; Kiupel, M. Volume 2: Mammary Tumors. In *Surgical Pathology of Tumors of Domestic Animals*; Kiupel, M., Ed.; Davis-Thompson DVM Foundation: Washington, DC, USA, 2019; pp. 1–195.

37. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. A Dataset for Breast Cancer Histopathological Image Classification. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 1455–1462. [CrossRef]

38. Breast Cancer Histopathological Database (BreakHis). Available online: https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/ (accessed on 3 May 2021).

39. Cawley, G.C.; Talbot, N.L.C. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.

40. Bussola, N.; Marcolini, A.; Maggio, V.; Jurman, G.; Furlanello, C. Not again! Data Leakage in Digital Pathology. *arXiv* **2019**, arXiv:1909.06539v2.

41. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

42. Kumar, A.; Singh, S.K.; Saxena, S.; Lakshmanan, K.; Sangaiah, A.K.; Chauhan, H.; Shrivastava, S.; Singh, R.K. Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer. *Inf. Sci.* **2020**, *508*, 405–421. [CrossRef]

43. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A.; Liu, W.; et al. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [CrossRef]

44. 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]

45. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, 7–9 May 2015; pp. 1–14.

46. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [CrossRef]

47. Tan, M.X.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; Volume 97.

48. 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. [CrossRef]

49. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

50. Tan, M.; Chen, B.; Pang, R.; Vasudevan, V.; Sandler, M.; Howard, A.; Le, Q.V. MnasNet: Platform-Aware Neural Architecture Search for Mobile. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019. [CrossRef]

51. Zheng, L.; Zhao, Y.; Wang, S.; Wang, J.; Tian, Q. Good Practice in CNN Feature Transfer. *arXiv* **2016**, arXiv:1604.00133.

52. Antropova, N.; Huynh, B.Q.; Giger, M.L. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Med. Phys.* **2017**, *44*, 5162–5171. [CrossRef] [PubMed]

53. Ng, J.Y.H.; Yang, F.; Davis, L.S. Exploiting Local Features from Deep Networks for Image Retrieval. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 7–12 June 2015.

54. Liu, Y.; Guo, Y.; Georgiou, T.; Lew, M.S. Fusion that matters: Convolutional fusion networks for visual recognition. *Multimedia Tools Appl.* **2018**, *77*, 29407–29434. [CrossRef]

55. Yang, T.; Li, Y.; Mahdavi, M.; Jin, R.; Zhou, Z. Nystroem Method vs Random Fourier Features: A Theoretical and Empirical Comparison. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012.

56. Williams, C.K.I.; Seeger, M. Using the Nystroem method to speed up kernel machines. In Proceedings of the 13th International Conference on Neural Information Processing Systems, Denver CO, USA, 1 January 2001.

57. Pêgo, A.; Aguiar, P. Bioimaging 2015. 2015. Available online: http://www.bioimaging2015.ineb.up.pt/dataset.html (accessed on 5 April 2021).

58. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [CrossRef]

59. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2009.

60. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorobot.* **2013**, *7*, 21. [CrossRef]

61. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

62. Wakili, M.A.; Shehu, H.A.; Sharif, H.; Sharif, H.U.; Umar, A.; Kusetogullari, H.; Ince, I.F.; Uyaver, S. Classification of Breast Cancer Histopathological Images Using DenseNet and Transfer Learning. *Comput. Intell. Neurosci.* **2022**, *2022*, 8904768. [CrossRef]

63. Burrai, G.P.; Baldassarre, V.; Brunetti, B.; Iussich, S.; Maniscalco, L.; Mariotti, F.; Sfacteria, A.; Cocumelli, C.; Grieco, V.; Millanta, F.; et al. Canine and feline in situ mammary carcinoma: A comparative review. *Vet. Pathol.* **2022**, *59*, 894–902. [CrossRef] [PubMed]

64. Li, X.; Shen, X.; Zhou, Y.; Wang, X.; Li, T.-Q. Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet). *PLoS ONE* **2020**, *15*, e0232127. [CrossRef] [PubMed]

65. Liu, M.; Yi, M.; Wu, M.; Wang, J.; He, Y. Breast Pathological Image Classification Based on VGG16 Feature Concatenation. *J. Shanghai Jiaotong Univ.* **2022**, *27*, 473–484. [CrossRef]

66. Dantas Cassali, G.; Cavalheiro Bertagnolli, A.; Ferreira, E.; Araujo Damasceno, K.; de Oliveira Gamba, C.; Bonolo de Campos, C. Canine Mammary Mixed Tumours: A Review. *Vet. Med. Int.* **2012**, *2012*, 274608. [CrossRef] [PubMed]

67. Rajput, D.; Wang, W.-J.; Chen, C.-C. Evaluation of a decided sample size in machine learning applications. *BMC Bioinform.* **2023**, *24*, 48. [CrossRef] [PubMed]

68. Kallipolitis, A.; Revelos, K.; Maglogiannis, I. Ensembling EfficientNets for the Classification and Interpretation of Histopathology Images. *Algorithms* **2021**, *14*, 278. [CrossRef]

69. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]

70. 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*, 16447. [CrossRef]

71. Bertram, C.A.; Klopfleisch, R. The Pathologist 2.0: An Update on Digital Pathology in Veterinary Medicine. *Vet. Pathol.* **2017**, *54*, 756–766. [CrossRef] [PubMed]

72. 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] [PubMed]

73. Salvi, M.; Molinari, F.; Iussich, S.; Muscatello, L.V.; Pazzini, L.; Benali, S.; Banco, B.; Abramo, F.; De Maria, R.; Aresu, L. Histopathological Classification of Canine Cutaneous Round Cell Tumors Using Deep Learning: A Multi-Center Study. *Front. Vet. Sci.* **2021**, *8*, 640944. [CrossRef] [PubMed]

74. La Perle, K.M.D. Machine Learning and Veterinary Pathology: Be Not Afraid! *Vet. Pathol.* **2019**, *56*, 506–507. [CrossRef]

75. Awaysheh, A.; Wilcke, J.; Elvinger, F.; Rees, L.; Fan, W.; Zimmerman, K.L. Review of Medical Decision Support and Machine-Learning Methods. *Vet. Pathol.* **2019**, *56*, 512–525. [CrossRef]