Histopathologic Analysis of Canine Tumors Using Convolutional Neural Networks

Douglas Rodrigues[1], João P. Papa[1], Rebeca Scalco[3], Fabiana A. M. León[2], Alessandre Hataka[2], Clayton R. Pereira[1]
[1]*School of Sciences*, São Paulo State University (UNESP), Bauru, São Paulo, Brazil
{d.rodrigues, joao.papa, clayton.pereira}@unesp.br
[2]*Animal Pathology*, São Paulo State University (UNESP), Botucatu, São Paulo, Brazil
{a.hataka, f.leon}@unesp.br
[3]*School of Veterinary Medicine and Animal Science*, Universität Bern, Bern, Switzerland
rebeca.scalco@unibe.ch

***Abstract*—Veterinary pathology plays a fundamental role in clinical decision-making, with histopathological examination serving as the gold standard for diagnosing canine neoplasms. Among these, mast cell tumors and squamous cell carcinomas are highly prevalent, and their accurate differentiation is essential for prognosis and treatment planning. However, traditional grading workflows remain subjective, labor-intensive, and susceptible to interobserver variability. This study investigates deep learning–based approaches for automated classification of canine mast cell tumors and squamous cell carcinomas using a curated dataset of hematoxylin-and-eosin (H&E) stained images. Thirteen state-of-the-art Convolutional Neural Network architectures were systematically evaluated under two learning-rate configurations to assess the influence of network depth, connectivity patterns, and optimization hyperparameters on model performance. The results show that Xception and ResNet-152 achieved the best performance at `lr = 0.001`, whereas InceptionResNetV2 and DenseNet-121 attained the highest accuracies and perfect ROC–AUC scores at `lr = 0.0001`. These findings highlight that both architectural choice and learning-rate selection critically affect convergence stability and predictive accuracy. Explainability analyses based on Grad-CAM and feature activation visualization confirmed that the models focused on histologically meaningful regions, supporting the biological plausibility of their decision processes. Overall, this study demonstrates the potential of deep learning to enhance diagnostic consistency, objectivity, and scalability in veterinary pathology, paving the way for more reliable computational support tools in clinical workflows.**

***Index Terms*—machine learning, deep learning, veterinary pathology, mast cell tumors, squamous cell carcinoma, artificial intelligence.**

I. Introduction

Diagnostic imaging plays a crucial role in veterinary clinical evaluation and has contributed to the growing demand for specialized professionals in this field. It supports multiple disciplines, including internal medicine, surgery, neurology, oncology, and obstetrics, and rapid, accurate interpretation of imaging data is essential for timely clinical decision-making [6]. Technological advancements in recent decades have further transformed medical practice by enabling more precise diagnosis, prognosis, and treatment planning [1].

Within this context, anatomic pathology remains the gold standard for diagnosing a wide range of diseases. It relies on the microscopic examination of stained tissue sections prepared on glass slides, allowing the identification of cancerous, infectious, and autoimmune conditions [4]. Despite its central role in clinical workflows, histopathological analysis is inherently labor-intensive and subject to interobserver variability.

Among cutaneous neoplasms in dogs, mast cell tumors (MCTs) are particularly prevalent and account for approximately 7–21% of all cases. Their clinical behavior varies widely, and prognostic evaluation typically considers factors such as histological grade, clinical stage, proliferation rate, recurrence, and systemic manifestations [12]. The Patnaik grading system [14] remains the most widely adopted and classifies MCTs into three grades based on differentiation, granulation, and mitotic index [7]. However, interpretation of these criteria is often subjective and may lead to inconsistent diagnostic outcomes.

Squamous cell carcinoma (SCC) is another common and clinically significant epithelial tumor in dogs. It predominantly affects animals aged 7 to 9.8 years, with increased incidence in breeds such as Rottweilers, Giant Schnauzers, Poodles, and Dachshunds, but without sex predisposition [2]. Histologically, SCC is categorized into well and poorly differentiated forms, with less common variants including acantholytic, clear cell, spindle cell, and carcinoma arising from Bowen's disease [13]. As with MCTs, histopathological evaluation of SCC is subjective and may be influenced by the pathologist's experience.

The limitations of traditional pathology workflows, especially the subjectivity of semi-quantitative grading systems, have motivated increased interest in Artificial Intelligence (AI) and Machine Learning (ML) to improve diagnostic consistency and reproducibility [4]. Recent advances in human oncology have demonstrated that ML models can successfully predict tumor genotypes directly from Hematoxylin and Eosin (H&E) stained slides [11]. Unlike rule-based systems, ML algorithms learn directly from image data and can identify complex, high-dimensional patterns that may not be easily discernible by

São Paulo Research Foundation (FAPESP) grants 2013/07375-0, 2019/07665-4, 2023/14427-8, 2023/03726-4, 2024/02807-3, 2024/01785-6 and 2025/07171-2, and Brazilian National Council for Scientific and Technological Development (CNPq) grant 308529/2021-9

human observers [3], [5]. These models continuously refine their internal representations as new data becomes available, enhancing their predictive capability over time.

Deep learning methods, particularly Convolutional Neural Networks (CNNs), have shown exceptional performance in tasks such as tumor detection, segmentation, and subtype classification across numerous medical imaging domains [9]. By functioning as powerful feature extractors and classifiers, CNNs enable discrimination of subtle morphological patterns and support more objective, quantitative histopathological assessments [8].

This study advances veterinary pathology diagnostics by investigating the automated classification of canine MCT and SCC using a diverse set of state-of-the-art CNN architectures. The specific objectives are as follows:

- to develop and validate an automated classification system for canine MCT and SCC using deep learning approaches;
- to systematically evaluate and compare the performance of multiple CNN architectures for histopathological image analysis;
- to investigate the impact of different learning rates on model performance and convergence behavior; and
- to establish a reproducible pipeline for histopathological image processing and analysis in veterinary pathology.

The remainder of this manuscript is organized as follows. Section II presents an overview of CNNs in histopathological image analysis. Section III describes the proposed methodology, including image acquisition and preprocessing. Section IV presents the experimental setup and results. Finally, Section V discusses the findings and outlines directions for future research.

II. Related Works

Convolutional Neural Networks (CNNs) have emerged as the dominant deep learning paradigm for image-based diagnostic tasks due to their ability to learn hierarchical, increasingly abstract representations of visual patterns. This characteristic is particularly valuable in histopathology, where diagnostic interpretation depends on subtle morphological cues such as nuclear pleomorphism, tissue organization, stromal distribution, and fine-grained textural variations. By automatically extracting discriminative features from microscopic images, CNNs provide a robust alternative to traditional handcrafted descriptors and help mitigate the subjectivity and interobserver variability associated with manual evaluation.

Several families of CNN architectures have been widely adopted in medical image analysis, each incorporating distinct design principles to enhance representational capacity, gradient stability, or computational efficiency. Inception-based networks employ parallel convolutional branches with different receptive fields, enabling multi-scale feature extraction suitable for heterogeneous histological structures. Residual networks (ResNets) introduce shortcut connections that alleviate vanishing gradients and facilitate the training of substantially deeper models, improving the capture of complex morphological patterns. Densely connected networks (DenseNets) reinforce gradient flow by linking each layer to all subsequent layers, promoting feature reuse and enabling compact yet expressive architectures. Depthwise separable models such as Xception factorize convolutions into spatial and channelwise operations, reducing computational cost while preserving strong discriminative ability, an important advantage when processing large numbers of high-resolution patches.

The success of these architectures has contributed to the rapid expansion of deep learning applications in histopathology. Early studies demonstrated that CNNs surpass traditional methods in tasks including tumor subtyping, necrosis detection, mitotic figure identification, and whole-slide image classification [15]. UNet-based architectures subsequently became the standard for segmentation tasks, effectively isolating tumor regions, glands, nuclei, and other structures of interest in whole-slide images [16]. For classification, transfer learning from ImageNet has been widely adopted, with models such as VGG [17], Inception [18], and ResNet [19] demonstrating strong performance across multiple cancers and staining variations.

Applications of deep learning in veterinary histopathology, although more recent, have yielded similarly promising results. Salvi et al. [20] evaluated AlexNet, InceptionV3, and ResNet architectures for classifying canine cutaneous round cell tumors (RCTs) and grading MCTs, achieving accuracies above 91% for RCT classification and 100% for MCT grading. These findings highlight the potential of deep learning to support veterinary diagnostic workflows, particularly in environments with limited access to specialized pathology expertise.

Despite these advances, comparative analyses involving a broad set of modern CNN architectures for canine tumor classification remain scarce. Furthermore, the influence of optimization hyperparameters, particularly the learning rate, on convergence behavior and predictive performance in veterinary datasets remains poorly characterized. Motivated by these gaps, the present study systematically evaluates multiple state-of-the-art CNN families to assess their ability to distinguish between mast cell tumors and squamous cell carcinomas in canine histopathological images. The specific architectural variants and training configurations investigated in this work are detailed in Section III.

III. Methodology

This study proposes a systematic and reproducible pipeline for the automated classification of canine histopathological images using deep learning. The methodology encompasses all stages of the workflow, from tissue collection and slide preparation to model training, validation, and comparative analysis. Figure 1 summarizes the overall process.

A. Data Collection and Tissue Processing

The dataset was constructed from biopsy samples of SCC and MCTs, obtained through routine diagnostic procedures at the Veterinary Hospital of São Paulo State University. To

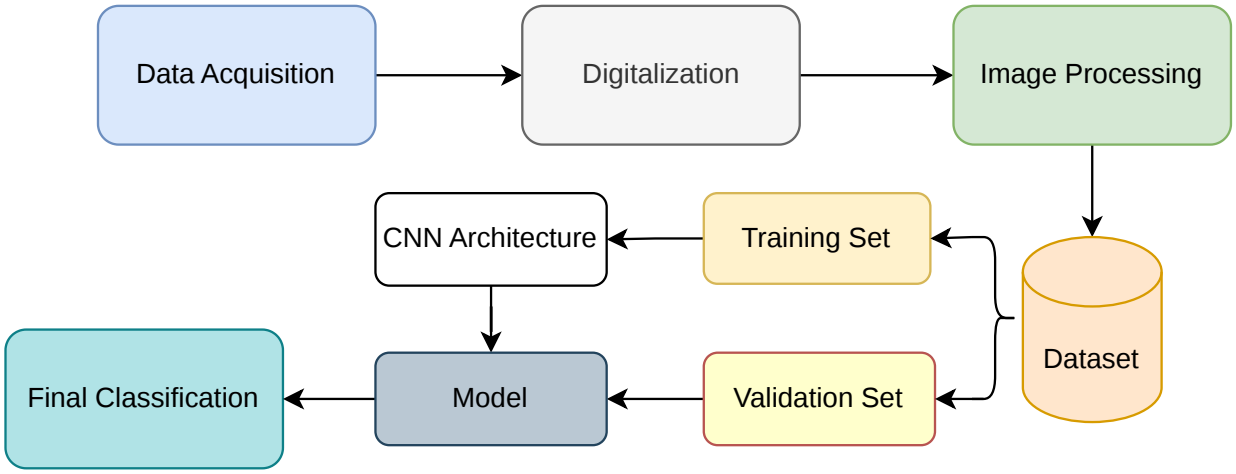


Fig. 1: Overview of the proposed pipeline for automated histopathological analysis of canine tumors using CNNs. The pipeline consists of eight stages: data acquisition, digitization, preprocessing, dataset split, architecture selection, model training, evaluation, and final classification.

maintain consistency and preserve tissue morphology, all samples underwent a standardized histological processing protocol aligned with established pathology guidelines.

The processing workflow included:

- fixation in 10% neutral buffered formalin;
- paraffin embedding following standard procedures;
- microtomy with sections cut at 4 µm thickness;
- Hematoxylin and Eosin (H&E) staining.

This protocol ensured uniform tissue preservation and staining quality, both essential for reliable microscopic interpretation and subsequent application of the deep learning model.

B. Image Acquisition and Dataset Organization

Digitization was performed using a TrueChrome 4K Pro scanner to ensure high-resolution, consistent image acquisition. Each slide was captured at 200× magnification with a native resolution of 1920×1080μm per pixel. The final dataset comprises 157 images collected throughout 2024, including 83 SCC and 74 MCT cases. Multiple optical magnifications (4×, 10×, 20×, and 40×) were recorded to preserve multiscale morphological information.

All images were manually reviewed and labeled by experienced veterinary pathologists, ensuring the reliability and diagnostic validity of the ground truth annotations. Representative examples of SCC, MCT, and normal tissue are shown in Figure 2.

C. Deep Learning Architectures

A diverse set of state-of-the-art CNN architectures was evaluated to investigate how different connectivity patterns, depth profiles, and multi-scale feature extraction strategies influence performance in veterinary histopathological classification. The following architectural families were included:

- **ResNet** (18, 34, 50, 101, 152 layers);
- **DenseNet** (121, 169, 201, 264 layers);
- **Inception** (InceptionV1, InceptionV4);
- **InceptionResNetV2**;
- **Xception**.

For each model, the final classification layer was replaced with a fully connected layer producing two outputs (SCC and MCT), ensuring compatibility with the binary classification task and reducing the risk of overfitting.

D. Training Protocol

Model training followed a standardized procedure designed to balance convergence stability and computational feasibility. All architectures were trained using the Adam optimizer under two learning rate configurations to assess their sensitivity to optimization hyperparameters:

- **lr = 0.001**: faster convergence but potentially more unstable;
- **lr = 0.0001**: slower, more conservative learning dynamics.

Training hyperparameters included:

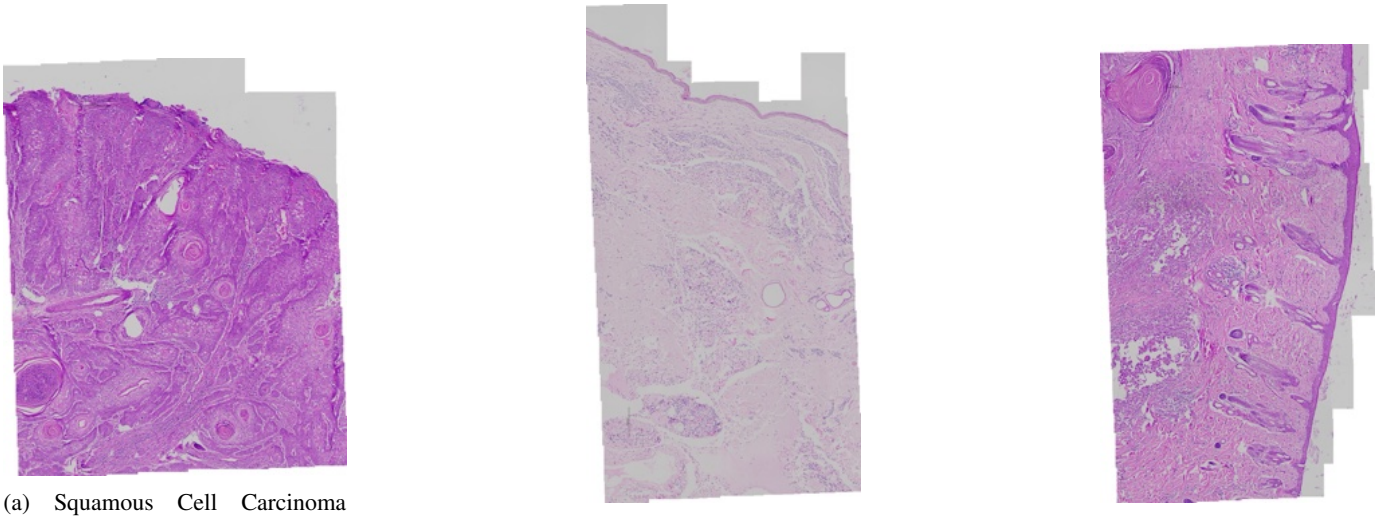
- **Loss Function:** Cross-Entropy Loss;
- **Batch Size:** 32;
- **Epochs per Fold:** 50.

A new model instance was initialized for each fold to avoid information leakage and ensure a fair comparison across architectures and learning rate settings.

E. Cross-Validation Experimental Protocol

To obtain a robust estimate of generalization performance, all architectures were evaluated using a stratified 5-fold cross-validation scheme. This procedure preserved the class distribution of SCC and MCT across all folds, reducing sampling bias and improving the stability of performance estimates.

For each architecture-learning rate configuration, models were trained on four folds, with the remaining fold used



(a) Squamous Cell Carcinoma (SCC)

(b) Mast Cell Tumor (MCT)

(c) Normal Tissue

Fig. 2: Representative histological images from the dataset.

exclusively for validation. After each fold, the following metrics were computed on the validation set:

- accuracy (ACC);
- weighted precision;
- weighted recall;
- weighted F1-score;
- ROC-AUC.

The metrics were then aggregated across folds to obtain the mean and standard deviation for each model. This cross-validation strategy provides an unbiased and stable assessment of model performance, particularly important in studies with limited datasets.

IV. Experimental Results

This section presents the results of the systematic evaluation of all CNN architectures considered in this study. We first describe the computational setup and evaluation criteria, then analyze convergence behavior across learning rates and architectures, and finally report cross-validated performance and interpretability.

A. Experimental Setup

All experiments were conducted on an NVIDIA Tesla P4 GPU with 8 GB of memory, using mixed-precision training to improve computational efficiency. The models were implemented in PyTorch 1.7.1 with custom data loaders and preprocessing routines that ensured consistent normalization and reproducible augmentation across folds.

Evaluation followed the stratified 5-fold cross-validation protocol described in Section III-E. For each fold, a fresh instance of every architecture was initialized and trained using the configurations defined in Section III-D, and validation was performed on the held-out fold. Stratification preserved the proportions of MCT and SCC samples across splits, reducing sampling bias and improving the reliability of the estimated metrics.

B. Comparative Analysis

Model performance was evaluated using five complementary metrics commonly employed in medical image classification tasks: accuracy, weighted precision, weighted recall, weighted F1-score, and ROC-AUC. Although convergence curves were omitted due to space constraints, training and validation dynamics were inspected throughout the cross-validation process to verify optimization stability under both learning rates.

Table I summarizes the cross-validated performance (mean and standard deviation over five folds) for all architectures at both learning rates. The results reveal consistent patterns regarding robustness, depth sensitivity, and architectural efficiency.

Several trends can be observed from these results:

- At `lr = 0.001`, Xception achieved the highest accuracy (0.943) and the highest ROC-AUC (0.997), followed closely by ResNet-152, which also obtained strong performance across all metrics.
- At `lr = 0.0001`, InceptionResNetV2 and DenseNet-121 stood out, with InceptionResNetV2 achieving the best overall accuracy (0.987) and both models reaching a ROC-AUC of 1.000.
- DenseNet variants exhibited remarkable robustness to learning rate changes, maintaining high accuracy and very high ROC-AUC values under both optimization regimes.
- ResNet architectures were more sensitive to learning rate variation. While ResNet-18 and ResNet-34 remained stable, deeper versions such as ResNet-50 and ResNet-101 did not consistently outperform their shallower counterparts.
- Inception-based architectures performed competitively, with InceptionV1 and InceptionV4 achieving accuracies above 0.87 and consistently high ROC-AUC scores, and InceptionResNetV2 ranking among the top models over-

TABLE I: Cross-validated performance (mean and standard deviation) of all evaluated CNN architectures for the binary classification of MCT and SCC. Results are reported for two learning rates (lr = 0.001 and lr = 0.0001) and five metrics: accuracy (Acc), weighted precision (Prec), weighted recall (Rec), weighted F1-score (F1), and area under the ROC curve (ROC-AUC). Best mean values for each metric and learning rate are highlighted in bold.

Model	lr = 0.001					lr = 0.0001				
	Acc	Prec	Rec	F1	ROC-AUC	Acc	Prec	Rec	F1	ROC-AUC
ResNet-18	0.917 (0.054)	0.932 (0.035)	0.917 (0.048)	0.916 (0.049)	0.982 (0.034)	0.911 (0.061)	0.928 (0.041)	0.911 (0.054)	0.909 (0.055)	0.981 (0.024)
ResNet-34	0.917 (0.044)	0.925 (0.034)	0.917 (0.039)	0.916 (0.040)	0.974 (0.032)	0.898 (0.086)	0.904 (0.072)	0.898 (0.077)	0.898 (0.077)	0.972 (0.022)
ResNet-50	0.872 (0.069)	0.885 (0.056)	0.872 (0.062)	0.871 (0.062)	0.942 (0.037)	0.904 (0.032)	0.908 (0.029)	0.904 (0.028)	0.904 (0.029)	0.944 (0.009)
ResNet-101	0.885 (0.059)	0.899 (0.036)	0.885 (0.053)	0.883 (0.055)	0.970 (0.016)	0.885 (0.080)	0.889 (0.069)	0.885 (0.072)	0.884 (0.072)	0.941 (0.056)
ResNet-152	0.936 (0.023)	0.941 (0.016)	0.936 (0.020)	0.936 (0.021)	0.989 (0.003)	0.936 (0.051)	0.941 (0.044)	0.936 (0.045)	0.936 (0.045)	0.978 (0.040)
DenseNet-121	0.791 (0.133)	0.861 (0.048)	0.791 (0.119)	0.769 (0.150)	0.970 (0.033)	0.975 (0.026)	0.977 (0.021)	0.975 (0.023)	0.975 (0.023)	**1.000 (0.000)**
DenseNet-169	0.917 (0.049)	0.926 (0.040)	0.917 (0.044)	0.916 (0.045)	0.979 (0.024)	0.943 (0.048)	0.947 (0.040)	0.943 (0.043)	0.943 (0.043)	0.998 (0.006)
DenseNet-201	0.853 (0.126)	0.866 (0.114)	0.853 (0.113)	0.852 (0.113)	0.918 (0.153)	0.929 (0.042)	0.940 (0.032)	0.929 (0.038)	0.929 (0.038)	0.993 (0.008)
DenseNet-264	0.892 (0.084)	0.911 (0.047)	0.892 (0.075)	0.888 (0.083)	0.962 (0.048)	0.968 (0.032)	0.970 (0.027)	0.968 (0.028)	0.968 (0.029)	0.996 (0.007)
InceptionV1	0.929 (0.074)	0.944 (0.046)	0.929 (0.066)	0.928 (0.068)	0.980 (0.019)	0.949 (0.067)	0.959 (0.045)	0.949 (0.060)	0.948 (0.061)	0.995 (0.009)
InceptionV4	0.873 (0.095)	0.890 (0.075)	0.873 (0.085)	0.871 (0.087)	0.944 (0.063)	0.898 (0.058)	0.904 (0.050)	0.898 (0.052)	0.897 (0.052)	0.945 (0.064)
InceptionResNetV2	0.929 (0.053)	0.937 (0.043)	0.929 (0.048)	0.929 (0.048)	0.978 (0.036)	**0.987 (0.017)**	**0.988 (0.015)**	**0.987 (0.016)**	**0.987 (0.016)**	**1.000 (0.000)**
Xception	**0.943 (0.060)**	**0.952 (0.040)**	**0.943 (0.054)**	**0.943 (0.055)**	**0.997 (0.008)**	0.930 (0.062)	0.938 (0.050)	0.930 (0.055)	0.929 (0.056)	0.982 (0.018)

all.

These findings highlight a trade-off between architectural depth and optimization stability. Dense connectivity (DenseNet family) and hybrid inception–residual designs (InceptionResNetV2) provided strong generalization with relatively low sensitivity to hyperparameter choices. In contrast, very deep residual networks required more careful tuning to fully exploit their representational capacity.

C. Explainability Analysis

To complement the quantitative evaluation, an explainability analysis was performed to identify the visual cues and histological structures driving the models' predictions. The main objective was to verify whether the CNNs focused on clinically meaningful regions, such as cellular morphology, keratin pearls in SCC, or the granulated cytoplasm characteristic of MCT, rather than on spurious artifacts.

The following explainability techniques were applied:

- **Grad-CAM**: generation of class-discriminative activation maps highlighting the regions that most strongly contributed to each prediction. For each architecture family (ResNet, DenseNet, Inception, Xception), Grad-CAM was computed from the last convolutional block, and the resulting heatmaps were overlaid on the original histological patches to visually assess whether the networks attended to relevant tissue structures.
- **Layer-wise activation visualization**: extraction and visualization of intermediate feature maps at different depths to provide qualitative insight into how low-, mid-, and high-level representations encode textural patterns, cellular organization and tumor–stroma interfaces.
- **Faithfulness analysis**: a quantitative measure of how well the explanations reflect the decision process of the model. Grad-CAM maps were used to mask the input images, preserving only highly activated regions, and the resulting change in predicted probability for the target class was measured. Larger reductions in confidence indicated that the highlighted regions were indeed critical for the prediction.
- **Stability analysis**: evaluation of the robustness of the explanations under small perturbations. Gaussian noise was added to the images, new Grad-CAM maps were generated and the correlation between original and perturbed maps was computed. Higher correlations indicated more stable and reliable explanations.
- **Interpretability score**: an entropy-based metric quantifying how spatially concentrated the Grad-CAM activations are. Explanations with lower entropy, that is, with activations focused on compact and anatomically coherent regions, received higher interpretability scores.

Beyond these methodological components, the explainability results were examined in detail for the four best-performing architectures: DenseNet-121 (lr = 0.0001), InceptionResNetV2 (lr = 0.0001), ResNet-152 (lr = 0.001) and Xception (lr = 0.001). Visual inspection revealed consistent attention to histologically meaningful regions across all models. DenseNet-121 produced highly localized heatmaps with concentrated activation around cohesive tumor cell groups, which is consistent with its notably higher interpretability score. InceptionResNetV2 generated sharply defined activation regions and achieved near-perfect faithfulness, indicating strong alignment between its explanations and prediction boundaries. ResNet-152 also achieved maximal faithfulness, although with slightly more diffuse activation patterns. Xception exhibited the highest stability under perturbations, suggesting that its explanations remained consistent even when the input was slightly altered.

Quantitative metrics supported these qualitative observations. Table II summarizes the faithfulness, stability, interpretability and confidence scores for the four selected architectures. Faithfulness values ranged from 0.924 to 1.000, confirming that the most activated regions were indeed critical for classification. Stability showed greater variation, with Xception demonstrating the most robust behavior and DenseNet-121 the least stable, although still within acceptable limits. Interpretability values indicated that DenseNet-121 produced the most spatially focused explanations, while the other models exhibited broader but still anatomically coherent activation

patterns. All architectures achieved extremely high confidence values, reinforcing the reliability of their outputs.

Taken together, these analyses provided converging qualitative and quantitative evidence that the best-performing models relied on histologically meaningful regions to distinguish between SCC and MCT. High faithfulness scores, stable explanations under perturbations and anatomically plausible activation patterns confirm that the CNNs captured diagnostically relevant structures rather than artifacts.

Figures 3–6 present representative explainability outputs for the four architectures, including Grad-CAM heatmaps and activation-value distributions.

V. Conclusion

This study presented a comprehensive evaluation of state-of-the-art Convolutional Neural Network architectures for the automated classification of canine mast cell tumors (MCT) and squamous cell carcinomas (SCC) using histopathological images. By systematically comparing multiple architectural families under two learning-rate configurations, we identified the models and optimization regimes that achieved the most accurate and stable performance.

The results demonstrate that architectural design plays a decisive role in classification accuracy and convergence dynamics. At `lr = 0.001`, Xception and ResNet-152 achieved the best overall performance, whereas at `lr = 0.0001`, InceptionResNetV2 and DenseNet-121 reached the highest accuracies and perfect ROC–AUC values. DenseNet variants were particularly robust across both learning rates, suggesting that dense feature reuse promotes stable generalization even under conservative optimization settings. In contrast, deeper residual architectures exhibited greater learning-rate sensitivity, emphasizing the need for careful hyperparameter tuning to ensure reliable deployment.

Beyond predictive performance, this work provides a fully reproducible pipeline encompassing standardized tissue preparation, high-resolution image digitization, dataset organization, model training, and interpretability assessment. Explainability analyses using Grad-CAM and intermediate-feature visualization confirmed that the models consistently attended to histologically meaningful structures, such as keratin pearls in SCC and the granulated cytoplasm typical of MCT. These findings reinforce the biological plausibility of the learned representations and support their clinical relevance.

The study highlights several directions for future research. Ensemble strategies that combine complementary architectural properties may further enhance robustness and predictive accuracy. Expanding the task to multi-class tumor classification or finer-grained grading systems could increase the clinical utility of automated histopathological approaches. Finally, validation on larger, multi-institutional datasets is essential to assess generalizability and to facilitate integration into real-world veterinary diagnostic workflows.

In summary, this work demonstrates the potential of deep learning to improve diagnostic efficiency and objectivity in veterinary pathology. With appropriate architectural selection, hyperparameter optimization, and interpretable model analysis, CNN-based systems can serve as valuable tools to support more consistent and reproducible histopathological evaluations.

References

[1] Acosta-Jiménez, S., González-Chávez, S. A., Camarillo-Cisneros, J., Pacheco-Tena, C. F., Ochoa-Albíztegui, R. E.: Aplicaciones de la inteligencia artificial en la medicina y la imagenología médica. Revista Anales de Radiología México **22**(2) (2023). DOI:10.24875/ARM.21000093

[2] Belluco, S., Brisebard, E., Watrelot, D., Pillet, E., Marchal, T., Ponce, F.: Digital Squamous Cell Carcinoma in Dogs. Veterinary Pathology **50**(6), 1078–1082 (2013). DOI:10.1177/0300985813490757

[3] Bi, W. L., Hosny, A., Schabath, M. B., Giger, M. L., Birkbak, N. J., Mehrtash, A., Allison, T., Arnaout, O., Abbosh, C., Dunn, I. F., Mak, R. H., Tamimi, R. M., Tempany, C. M., Swanton, C., Hoffmann, U., Schwartz, L. H., Gillies, R. J., Huang, R. Y., Aerts, H. J. W. L.: Artificial intelligence in cancer imaging: Clinical challenges and applications. CA: A Cancer Journal for Clinicians **69**(2), 127–157 (2019). DOI:10.3322/caac.21552

[4] Brodsky, V., Ullah, E., Bychkov, A., Song, A. H., Walk, E. E., Louis, P., Rasool, G., Singh, R. S., Mahmood, F., Bui, M. M., Parwani, A. V.: Generative Artificial Intelligence in Anatomic Pathology. Archives of Pathology & Laboratory Medicine (2025). DOI:10.5858/arpa.2024-0215-RA

[5] Bulusu, G., Vidyasagar, K. E. C., Mudigonda, M., Saikia, M. J.: Cancer Detection Using Artificial Intelligence: A Paradigm in Early Diagnosis. Archives of Computational Methods in Engineering (2025). DOI:10.1007/s11831-024-10209-0

[6] Burti, S., Banzato, T., Coghlan, S., Wodzinski, M., Bendazzoli, M., Zotti, A.: Artificial intelligence in veterinary diagnostic imaging: Perspectives and limitations. Research in Veterinary Science **175**, 105317 (2024). DOI:10.1016/j.rvsc.2024.105317

[7] Costa, M. C., Silva, A. L. D. A., Moreira, T. A., Gundim, L. F., Medeiros-Ronchi, A. A.: Prevalence and epidemiological and histopathological features of canine cutaneous mast cell tumours in Uberlândia, Brazil. Acta Veterinaria Brno **86**(2), 189–193 (2017). DOI:10.2754/avb201786020189

[8] Giulietti, M., Cecati, M., Sabanovic, B., Scirè, A., Cimadamore, A., Santoni, M., Montironi, R., Piva, F.: The Role of Artificial Intelligence in the Diagnosis and Prognosis of Renal Cell Tumors. Diagnostics **11**(2), 206 (2021). DOI:10.3390/diagnostics11020206

[9] Mahmood, H., Shaban, M., Rajpoot, N., Khurram, S. A.: Artificial Intelligence-based methods in head and neck cancer diagnosis: an overview. British Journal of Cancer **124**(12), 1934–1940 (2021). DOI:10.1038/s41416-021-01386-x

[10] Nemec, A., Murphy, B., Kass, P. H., Verstraete, F. J. M.: Histological Subtypes of Oral Non-tonsillar Squamous Cell Carcinoma in Dogs. Journal of Comparative Pathology **147**(2–3), 111–120 (2012). DOI:10.1016/j.jcpa.2011.11.198

[11] Puget, C., Ganz, J., Ostermaier, J., Conrad, T., Parlak, E., Bertram, C. A., Kiupel, M., Breininger, K., Aubreville, M., Klopfleisch, R.: Artificial intelligence can be trained to predict c-KIT -11 mutational status of canine mast cell tumors from hematoxylin and eosin-stained histological slides. Veterinary Pathology **62**(2), 152–160 (2025). DOI:10.1177/03009858241286806

[12] Khoo, A., Lane, A., Wyatt, K.: Intranasal mast cell tumor in the dog: A case series. Can Vet J. **58**(8), 851–854 (2017).

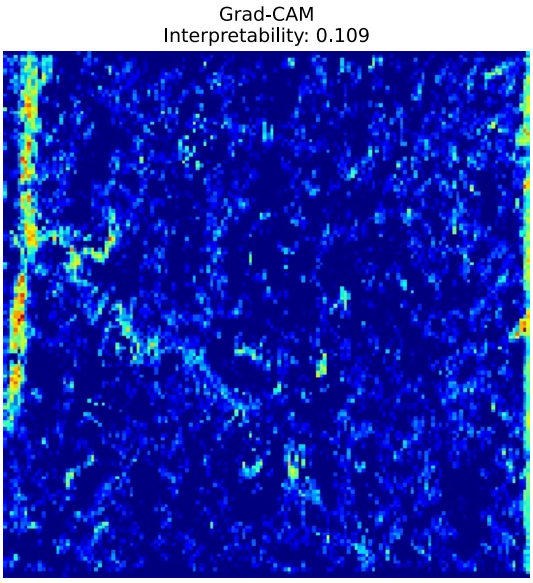
[13] Rodríguez Guisado, F., Suárez-Bonnet, A., Ramírez, G. A.: Cutaneous Spindle Cell Squamous Cell Carcinoma in Cats: Clinical, Histological, and Immunohistochemical Study. Veterinary Pathology **58**(3), 503–507 (2021). DOI:10.1177/0300985820985126

[14] Sabattini, S., Scarpa, F., Berlato, D., Bettini, G.: Histologic Grading of Canine Mast Cell Tumor. Veterinary Pathology **52**(1), 70–73 (2015). DOI:10.1177/0300985814521638

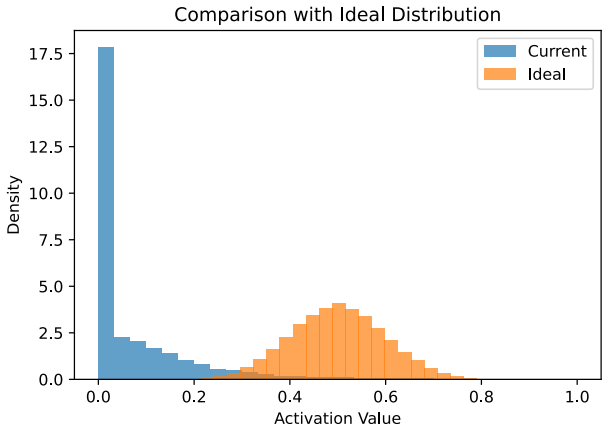
[15] Le Cun, Y., Jackel, L.D., Boser, B., Denker, J.S., Graf, H.P., Guyon, I., Henderson, D., Howard, R.E., Hubbard, W.: Handwritten digit recognition: applications of neural network chips and automatic learning. IEEE Communications Magazine **27**(11), 41–46 (1989). DOI:10.1109/35.41400

TABLE II: Explanation quality metrics for the four best-performing architectures.

Architecture	LR	Faithfulness	Stability	Interpretability	Confidence
DenseNet-121	0.0001	0.982	-0.546	0.109	1.000
InceptionResNetV2	0.0001	0.999	-0.346	0.002	1.000
ResNet-152	0.001	1.000	-0.464	0.001	1.000
Xception	0.001	0.924	-0.077	0.005	0.996

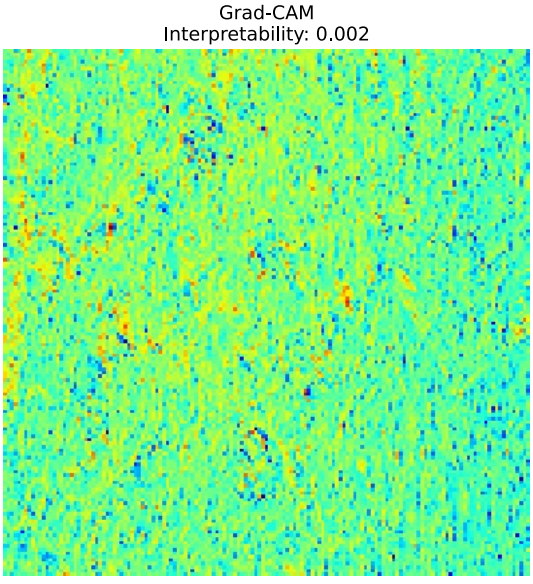


(a) Grad-CAM heatmap

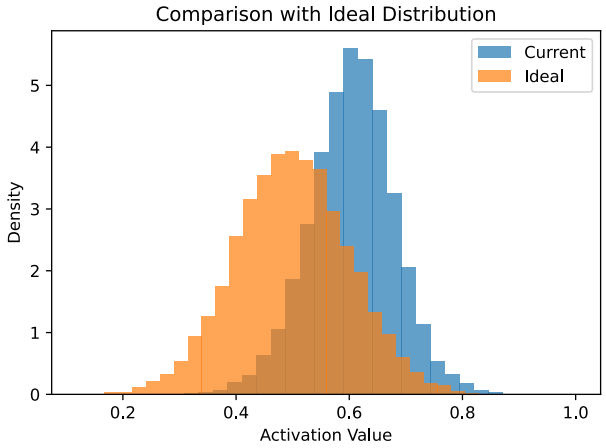


(b) Activation-value distribution

Fig. 3: Explainability results for DenseNet-121 (lr = 0.0001).



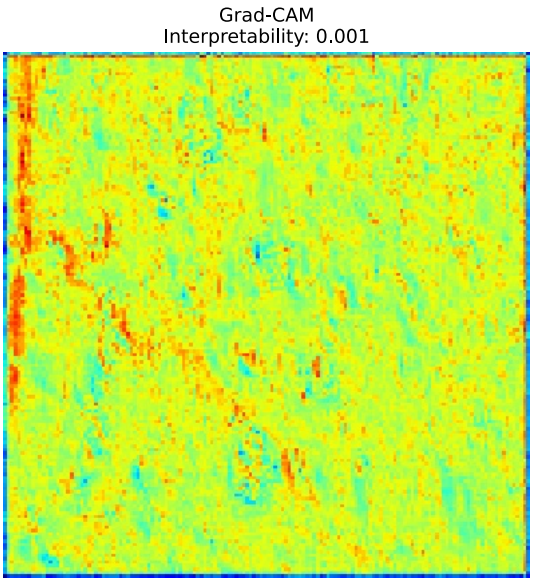
(a) Grad-CAM heatmap



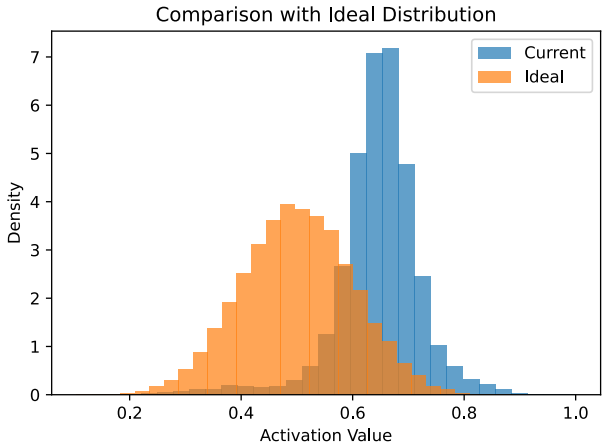
(b) Activation-value distribution

Fig. 4: Explainability results for InceptionResNetV2 (lr = 0.0001).

[16] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. DOI:10.1007/978-3-319-24574-4_28

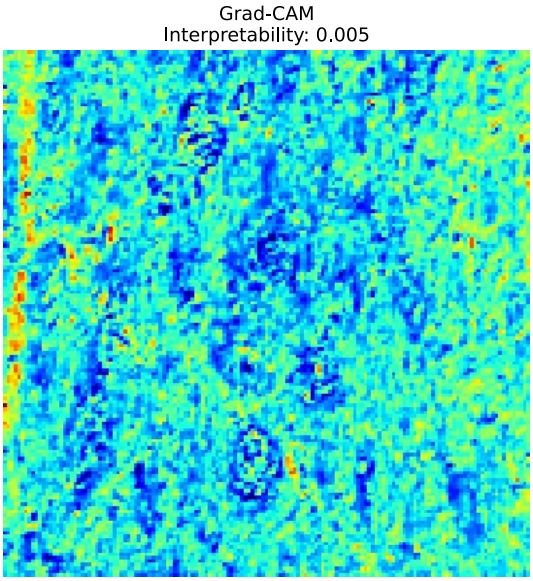


(a) Grad-CAM heatmap

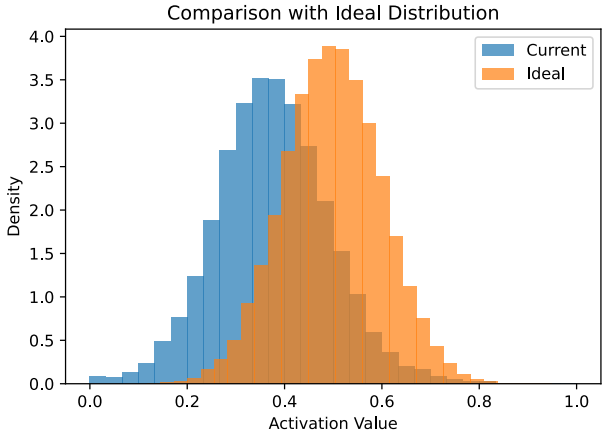


(b) Activation-value distribution

Fig. 5: Explainability results for ResNet-152 (lr = 0.001).



(a) Grad-CAM heatmap



(b) Activation-value distribution

Fig. 6: Explainability results for Xception (lr = 0.001).

[17] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014). DOI:10.48550/arXiv.1409.1556

[18] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1–9 (2015). DOI:10.1109/CVPR.2015.7298594

[19] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778 (2016). DOI:10.1109/CVPR.2016.90

[20] Salvi, M., Molinari, F., Iussich, S., et al.: Histopathological Classification of Canine Cutaneous Round Cell Tumors Using Deep Learning: A Multi-Center Study. Front Vet Sci **8**, 640944 (2021). DOI:10.3389/fvets.2021.640944

[21] Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely Connected Convolutional Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4700–4708 (2017). DOI:10.1109/CVPR.2017.243