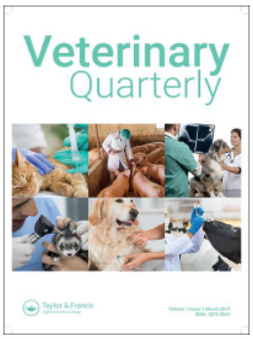

Veterinary Quarterly

ISSN: 0165-2176 (Print) 1875-5941 (Online) Journal homepage: www.tandfonline.com/journals/tveq20

Automated segmentation of canine pulmonary masses in CT imaging using AI

Artur Jurgas, Silvia Burti, Marek Wodziński, Caterina Puccinelli, Giunio Bruto Cherubini, Simonetta Citi, Giulia Poloni, Nicolò Mastromattei, Margherita Bendazzoli, Diane Wilson, Alessandro Zotti & Tommaso Banzato

To cite this article: Artur Jurgas, Silvia Burti, Marek Wodziński, Caterina Puccinelli, Giunio Bruto Cherubini, Simonetta Citi, Giulia Poloni, Nicolò Mastromattei, Margherita Bendazzoli, Diane Wilson, Alessandro Zotti & Tommaso Banzato (2025) Automated segmentation of canine pulmonary masses in CT imaging using AI, Veterinary Quarterly, 45:1, 2573449, DOI: 10.1080/01652176.2025.2573449

To link to this article: https://doi.org/10.1080/01652176.2025.2573449

VETERINARY QUARTERLY
2025, VOL. 45, NO. 1, 1–10
https://doi.org/10.1080/01652176.2025.2573449

Taylor & Francis
Taylor & Francis Group

Automated segmentation of canine pulmonary masses in CT imaging using AI

Artur Jurgas[a], Silvia Burti[b], Marek Wodziński[a,c], Caterina Puccinelli[d], Giunio Bruto Cherubini[d], Simonetta Citi[d], Giulia Poloni[b], Nicolò Mastromattei[b], Margherita Bendazzoli[b], Diane Wilson[e], Alessandro Zotti[b] and Tommaso Banzato[b]

[a]Department of Measurement and Electronics, AGH University of Krakow, Krakow, PL, Poland; [b]Department of Animal Medicine, Production and Health, University of Padua, Legnaro, Padua, Italy; [c]Information Systems Institute, University of Applied Sciences - Western Switzerland (HES-SO Valais), Sierre, Switzerland; [d]Department of Veterinary Sciences, University of Pisa, San Piero a Grado, Pisa, Italy; [e]Antech Imaging Services, Fountain Valley, California, US

ABSTRACT

Primary pulmonary lung cancer is rare in dogs, and clinicians increasingly rely on advanced imaging for diagnosis and treatment planning. However, manual lesion segmentation can be time-consuming and subject to operator variability. This retrospective study compiled a multicenter dataset of canine CT scans containing at least one pulmonary mass measuring more than 2 cm. Data were collected from two university veterinary hospitals and a teleradiology service, encompassing varying acquisition protocols and scanner types. Lesions were manually segmented to create ground truth masks, and an AI model was trained and evaluated using the nnUNet v2 framework with a 5-fold cross-validation approach. Performance on a separate test set of 30 scans was quantified using the Dice Similarity Coefficient (DSC) and Average Symmetric Surface Distance (ASSD). The databse was made of 217 cases. The training/validation set comprised 187 cases. The model's segmentation accuracy was tested on 30 cases. The trained model had a high segmentation accuracy on the test set, with a mean DSC of 0.91 and an ASSD of 1.88 mm. The model had high performance on homogeneous, well-defined masses, whereas the presence of intralesional mineralisation or pleural effusion had a negative impact on the model's performance.

ARTICLE HISTORY
Received 2 May 2025
Accepted 29 September 2025

KEYWORDS
Computed tomography; U-Net; Dice score; lung mass; segmentations

1. Introduction

Primary pulmonary lung cancer is a rare condition in dogs, accounting for approximately 1% of all canine cancers (Marcinowska et al. 2024), and it is often detected incidentally on survey radiographs performed for unrelated reasons (Vail et al. 2020). Although benign lung tumors, such as adenomas, papillomas, haemangioma, and granulomas, have been reported in dogs, they are exceedingly rare (Marcinowska et al. 2024). Complete disease staging is typically performed after the initial diagnosis using computed tomography (Vail et al. 2020). Treatment options include surgery, chemotherapy, and radiation therapy, or a combination of those, depending on the location and stage of the disease (Vail et al. 2020). The response to treatment, especially when chemotherapy or radiotherapy are chosen, is evaluated following the Response Evaluation Criteria in Solid Tumors criteria (Nguyen et al. 2015).

AI-driven algorithms for the automatic analysis of diagnostic images are becoming increasingly popular in veterinary medicine (Burti et al. 2024). Presently, various subscription-based online platforms utilize AI models for the automatic analysis of radiographs in dogs and cats (Müller et al. 2022). Research and market applications of AI tools predominantly concentrate on plain radiographs (Banzato et al. 2017; Banzato et al., 2018; Adrien-Maxence et al. 2022; Boissady et al. 2020). Lesion detection, particularly in

CONTACT Tommaso Banzato tommaso.banzato@unipd.it; tommaso.banzato@unipd.it Department of Animal Medicine, Production and Health, University of Padua, Viale dell'Università 16, 35020 Legnaro, Padua, Italy.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/01652176.2025.2573449.

thoracic radiographs (Adrien-Maxence et al. 2022; Burti et al. 2024; Burti et al. 2020; Boissady et al. 2020; Valente et al. 2023), and orthopedics (McEvoy et al. 2021; Begum Ergun and Guney 2021), represent the most extensively studied applications. In terms of lesion classification, AI models have been created to assess the quality of thoracic radiographic images (Banzato et al. 2023; Tahghighi et al. 2024), predict the histologic association and grade with meningiomas based on MRI images (Banzato et al. 2017; Banzato et al., 2018), and classify lesions based on MRI (Banzato et al., 2018). Semantic AI algorithms have been designed for the automatic segmentation of lesions, including retropharyngeal lymph nodes (Schmid et al. 2022), head and neck tumors (Park et al. 2021), and normal kidneys (Ji et al. 2022), and hepatic masses.

In both human and veterinary medicine, lesion segmentation (LS) is crucial, yet labor-intensive, as it significantly influences cancer clinical practices. The planning of treatment extensively relies on the dimensions and positioning of lesions (Park et al. 2021). Furthermore, it is crucial for assessing treatment response in longitudinal imaging analyses. Traditionally, LS is manually conducted by skilled professionals using specialized tools. Although manual segmentation generally achieves high accuracy, it is time-consuming and varies with the operator. As a result, AI-based segmentation methods are gaining preference over manual LS (Ansari et al. 2022).

This research utilized a dataset comprising canine CT scans, each featuring one or more lung masses, to train and evaluate an AI-based algorithm designed to automate the segmentation of lung tumors in canine CT scans. The aim was to develop a model that could identify and segment lung tumors with nearly expert-level accuracy on CT scans originating from a variety of medical institutions. During this phase of the study, our focus was restricted to larger lesions, specifically those with a primary diameter exceeding 2 cm.

2. Material and methods

This retrospective study included CT scans performed between June 2019 and July 2024. Ethical approval was waived due to the retrospective nature of the study, and informed consent for personal data processing was obtained from the owners. All experiments were conducted in accordance with relevant guidelines and regulations.

2.1. Datasets

The primary objective of this data collection was to build a comprehensive dataset of CT studies featuring lung masses in dogs. This dataset would be used to train, test, and validate an automatic segmentation algorithm. To enhance the algorithm's generalizability and minimize selection bias, we included data from the widest possible range of CT scanners.

To assemble the dataset, we retrospectively searched the databases of the Veterinary Hospital of the University of Padua, the Veterinary Hospital of the University of Pisa, and Antech Imaging Services for CT studies performed in dogs with lung masses or nodules between January 2019 and December 2023.

Keywords like 'Lung Mass,' 'Pulmonary Mass,' 'Lung Nodule,' and 'Primary Lung Tumor' were employed to retrieve pertinent studies. Due to the absence of a universally recognized differentiation between nodules and masses, both terms were included in the search strategy. The broad nature of these search terms resulted in many irrelevant cases in the initial findings, which were then manually filtered using CT reports to pinpoint pertinent studies. The selected cases were stored in Digital Imaging and Communications in Medicine (DICOM) format in a dedicated folder, maintaining complete anonymization of the CT images and associated reports. Only post-contrast thoracic CT scans were utilized for lesion segmentation. As of now, a universally accepted size threshold to differentiate nodules from masses does not exist. In this research, we have chosen to classify all nodular lesions with the largest axial size exceeding 2 cm as masses. The study included only cases with a maximum of 2 lesions.

2.2. Ground truth determination

Creating effective segmentation masks, which are binary volumes outlining lesions within a dataset, is vital for training segmentation algorithms. We used the open-source platform 3D Slicer to produce

these masks. Initially, we imported and visually assessed the chosen volumes in 3D Slicer to confirm good image quality and lack of notable artifacts. Once imported in Slicer all DICOM metadata is lost and, thereby, the studies were completely anonymized. Volumes were excluded whenever breathing artifacts were detected or if the edges of the mass were not well-defined (as in cases where extensive lung consolidation enveloped the mass). The primary challenge lies in the inability to discern the tumour from the lung consolidation in the initial stage. Consequently, the creation of high-quality annotations becomes impossible, rendering the accurate evaluation of the algorithm's performance unfeasible. The contouring commenced with manually tracing lesions in a subset of images. Automatic tools were then employed to fill in gaps, but given the complex shapes of lesions, these tools frequently needed manual adjustment to achieve precise contours. This refinement is crucial as lesions appear over multiple CT slices. Manually delineating large lesions crossing many slices would be extremely time-intensive. Ultimately, both the original volume and the generated segmentation mask were saved separately.

2.2.1. Model architecture

The nnUNet v2 framework (Isensee et al. 2021) was employed for medical image segmentation due to its robust self-configuring capabilities and adaptability to diverse datasets. This platform automatically tailors its U-Net-based (Ronneberger et al. 2015) architecture—including preprocessing, network topology, and training parameters—based on the characteristics of the input data, streamlining the segmentation pipeline without requiring manual intervention.

In this study we employed the Residual Encoder U-Net in the large configuration (nnU-Net ResEnc L) as implemented in the nnU-Net v2 framework (Isensee et al. 2021). This architecture follows the standard encoder–decoder design but replaces the plain convolutional blocks in the encoder with residual blocks, which stabilize gradient flow and facilitate the training of deeper networks. The 'L' preset refers to the large-scale configuration optimized for modern GPUs with approximately 22 GB VRAM usage and a typical training runtime of 35 GPU hours. Compared to the smaller 'M' preset, ResEnc L uses larger patch sizes and higher feature map dimensionality, thereby improving segmentation accuracy on complex datasets at the cost of increased memory and computation.

We specifically selected the 3D full-resolution (fullres) model variant. The 3D fullres model processes volumetric image patches at the target resolution, preserving detailed spatial information across all three dimensions. This is advantageous for our dataset, where accurate lesion characterization depends on fine structural continuity and subtle intensity variations that could be lost under downsampling (3D low-resolution variant) or 2D slice-wise analysis (2D variant). In practice, the fullres model offered the most favorable trade-off between segmentation accuracy and computational feasibility in our experiments, outperforming lighter 2D and low-resolution cascaded models without exceeding the available GPU memory. We provide the initial testing of different configurations as supplementary material in Tables 1 and 2.

2.2.2. Training configuration

Cases were resampled to a uniform spacing of 1.25 mm x 0.58 mm x 0.58 mm before training. The raw image intensities underwent a normalization process where lesion intensity values were recorded. The mean (−267), standard deviation (313), along with the 0.5 and 99.5 percentiles (−900, 325) were calculated. Subsequently, intensities were clipped to these percentiles, the mean was subtracted, and the result was divided by the standard deviation.

Throughout training, the network processed randomly chosen 3D patches sized 128×224×224 from the resampled volumes. This approach optimizes anatomical context capture within GPU memory limitations and acts as data augmentation by presenting the network with diverse input scan subregions. To mitigate class imbalance 66.7% of samples are picked from random spots within a training case, and 33.3% of patches are ensured to include the foreground class.

For optimization, nnU-Net employed the batch Dice loss in combination with cross-entropy. With batch Dice enabled, the Dice score is computed as if all samples in a mini-batch constitute a single volume, thereby stabilizing the gradient signal in cases where individual patches contain few or no lesion

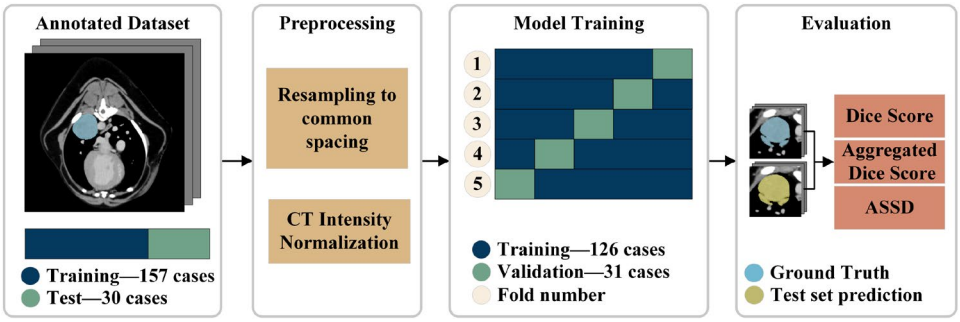

Figure 1. The model training and evaluation involved first dividing the data into separate training and test sets. Within the training phase, the data was divided into five distinct folds. The model underwent training five times, each time with a unique subset, while performance in each instance was determined using a distinct validation set. The final predictions were achieved by averaging the outputs from all folds applied to the test images.

voxels. This is particularly relevant in highly imbalanced segmentation tasks, where treating each patch independently could otherwise lead to unstable updates and slower convergence.

A 5-fold cross-validation approach was utilized. Each fold used 80% of the data for training and 20% for validation. A distinct set of 30 cases was set aside for testing the model's generalization. Refer to Figure 1 for the training and evaluation sequence. Training parameters, including learning rates and batch sizes, were automatically calibrated according to data characteristics. Training employed 12 CPU cores with 15 Gb of memory each and a GPU powered by an Nvidia A100-SXM4-40GB using CUDA version 12.1. The software used included Python 3.10, PyTorch 2.5, nnUNetv2 2.5.1, SimpleITK 2.4.0 under Ubuntu 22.04.5 LTS.

2.2.3. Evaluation metrics

Evaluating automatic segmentation methods uses specific metrics to ensure accuracy. The pixel-wise comparison between expert-provided ground truth and algorithm-predicted segmentation employs three key metrics: Dice Similarity Coefficient (DSC), Aggregated Dice Score (DSCAgg), and Average Symmetric Surface Distance (ASSD). The DSC measures the similarity between the expert's and the algorithm's segmentations. DSCAgg considers an aggregated sum of True Positives, False Negatives, and False Positives, providing a more balanced score by reducing the influence of small lesions. Consequently, it serves as a single consolidated metric for the entire test dataset and is not presented on Figures 3 and 4. ASSD evaluates the average distance between the surfaces of two shapes, offering a balanced gauge of their overall similarity.

3. Results

3.1. Dataset

In the first phase, 403 CT studies featuring nodular (nodules/masses) lung lesions were retrieved. Of these 403 cases only 217 were used for training and testing the algorithm. 186 examinations were discarded due to: 1- movement artifacts, 2- presence of diffused consolidation hindering clear visualization of the borders of the mass. The average lesion dimensions in the final dataset were 179.28 × 46.08 × 20.63 mm. Of these 217 cases, 187 were included in the training/validation dataset and 30 were used as a hold-out test set.

The dataset included a total of 217 animals (mean age: 10.92, standard deviation: 2.68 years). There were 117 males and 100 females, including both intact and neutered individuals. The canine population encompassed a wide range of breeds, including: Alaskan Husky, American Cocker Spaniel, Australian Kelpie, Australian Shepherd, Basset Hound, Beagle, Bernese Mountain Dog, Bichon Frise, Border Collie, Boston Terrier, Boxer, Brittany, Bull Terrier, Cairn Terrier, Chihuahua, Chow Chow, Cocker Spaniel, Collie Mixed, Corgi Mixed, Dachshund, Doberman Pinscher, English Bulldog, English Setter, Flat-Coated Retriever, Fox Terrier, French Bulldog, German Shepherd, Giant Schnauzer, Golden Retriever, Goldendoodle, Great

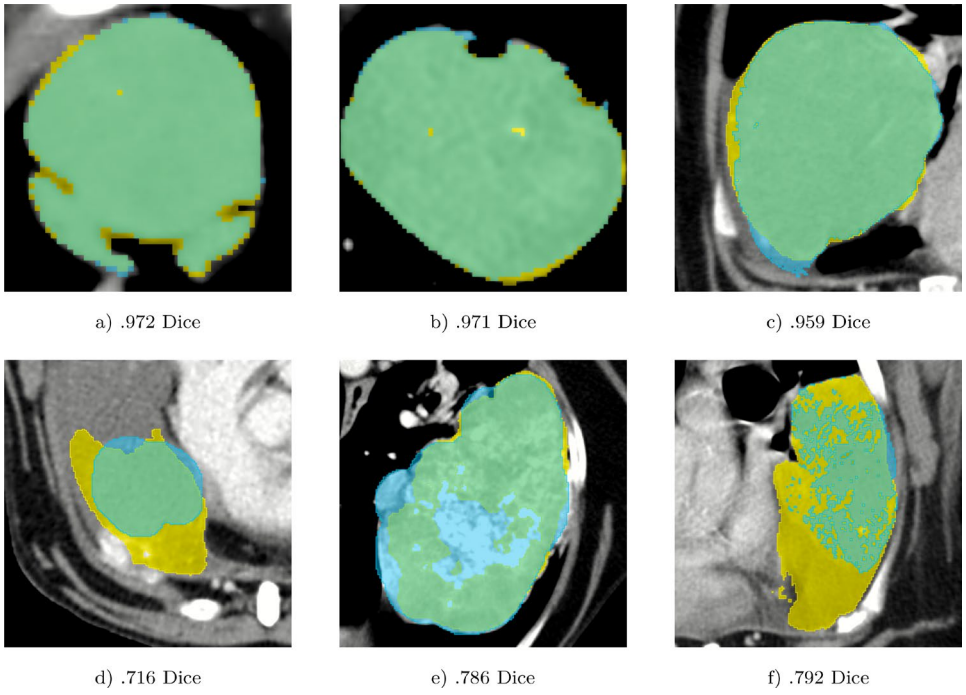

Figure 2. Representative cases highlighting the variability of the automated segmentation (yellow) relative to the ground-truth lesion contours (blue). Regions of overlap appear in green. The dice similarity coefficient (DSC) for each case is indicated below the corresponding image. Examples are ordered from the highest DSC (a) to the lowest (f). All CT scans are shown in a soft-tissue window (window level = 40 HU, window width = 350 HU).

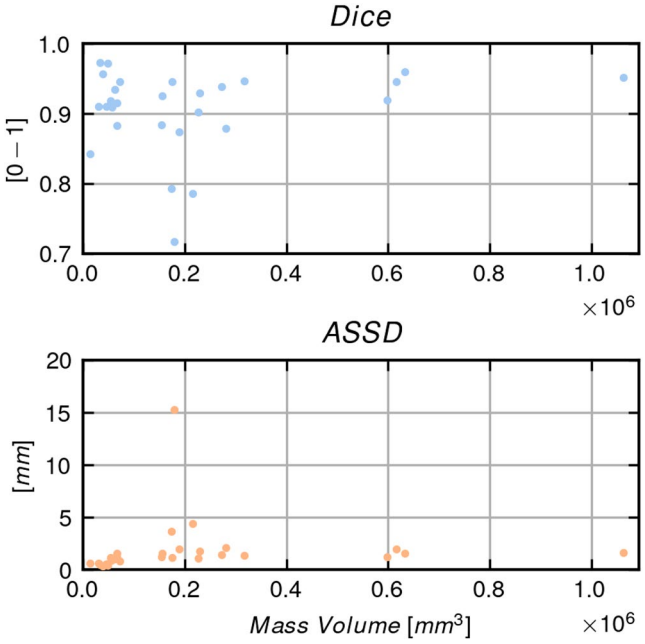

Figure 3. Relationship between lesion size and segmentation accuracy. Across more than three orders of magnitude in mass volume, dice scores remain uniformly high (≥ 0.85 for nearly all cases), indicating that volumetric overlap is essentially volume-independent. Likewise, ASSD stays below 3 mm for most lesions, with only a few larger-volume outliers showing appreciable surface error. Overall, neither metric displays a systematic degradation with increasing volume, suggesting that the algorithm maintains robust accuracy from very small to very large masses.

Dane Mixed, Great Pyrenees, Havanese, Jack Russell Terrier, King Charles Spaniel, Kuvasz, Labradoodle, Labrador Retriever, Maltese, Mastiff, Miniature Dachshund, Miniature Pinscher, Miniature Poodle, Miniature Schnauzer, Mixed, Norfolk Terrier, Papillon, Pit Bull Terrier, Pomeranian, Portuguese Water Dog, Pudelpointer,

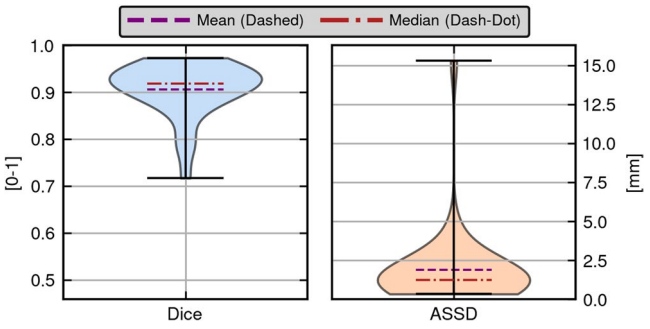

Figure 4. The smoothed violins visualize the full distribution, black bars mark the extrema (excluding any outliers beyond them), the **purple dashed** lines denote the mean, and the **red dash-dot** lines indicate the median. High dice values and low ASSD values correspond to better agreement with the ground-truth; the data show a tight cluster around dice > 0.9 and ASSD < 3 mm, with only a few isolated outliers widening the tails.

Pug, Rat Terrier, Rhodesian Ridgeback, Rottweiler, Saluki, Schipperke, Schnoodle, Scottish Terrier, Shepherd Mixed, Shih Tzu, Siberian Husky, Smooth Fox Terrier, Staffordshire Bull Terrier, Terrier Mixed, Utonagan, Weimaraner, Welsh Corgi, West Highland White Terrier, Whippet, and Yorkshire Terrier. The breed was not specified in 10 cases.

In this research, the CT studies were acquired in over 60 different institutions and were obtained from diverse CT scanners, resulting in significant variability in the acquisition parameters, particularly concerning slice thickness and radiological settings. The CTs were procured using the following CT scanners(GE MEDICAL SYSTEMS Brivo CT385 Series, GE MEDICAL SYSTEMS LightSpeed Plus, GE MEDICAL SYSTEMS LightSpeed Pro 32, GE MEDICAL SYSTEMS LightSpeed Ultra, GE MEDICAL SYSTEMS LightSpeed VCT, GE MEDICAL SYSTEMS LightSpeed16, GE MEDICAL SYSTEMS Revolution ACT, International Medical Solutions, Neurologica CereTom, SIEMENS Emotion 16, SIEMENS Spirit, TOSHIBA Aquilion, TOSHIBA Aquilion Lightning). The slice thickness ranged from 1 to 3 mm, KV ranged from 100 to 150, mAs ranged from 100 to 140.

Manual segmentation took between 10 and 40 min per case, though precise times were not consistently recorded. Segmentation time fluctuated based on the size and characteristics of the lesions. Radiologists in this study reported that contouring masses with regular shapes and smooth margins were completed faster than those that were irregular. Typically, the pronounced difference in attenuation between the mass and surrounding lung tissue allows for an easy separation of the mass from the lungs. However, when masses exhibit heterogeneous attenuation, the selection tools based on HU thresholds often fall short, necessitating manual adjustments for complete lesion segmentation.

3.2. Model performance

The model was tested on the test set of 30 CT scans containing lung masses. All the cases in the test set came from Antech Imaging Services' database and were acquired in 21 different institutions with 6 different scanners (GE MEDICAL SYSTEMS LightSpeed Pro 32, GE MEDICAL SYSTEMS LightSpeed VCT, GE MEDICAL SYSTEMS LightSpeed16, SIEMENS Emotion 16, TOSHIBA Aquilion, TOSHIBA Aquilion Lightning). The model showed high precision in the segmentation of lung masses in the test set with a **DSC of 0.91** (range), **DSCAgg of 0.92**, and an **ASSD of 1.88 mm** (range). A more detailed analysis of the results is reported in Figures 2–4. The training time of the model took 42h.

Lung consolidation was identified in 24 instances and pleural effusion in 15 instances within the training/validation dataset, whereas the test dataset revealed lung consolidation in 6 instances and pleural effusion in 3. Although assessing the impact of these lesions on the accuracy of the developed algorithm would be highly complex due to the small size of the test set, examining the results individually could provide valuable insights. Masses characterized by homogeneous parenchyma, a regular shape, and a distinct boundary with the lungs produced the most detailed results. Notably, the algorithm consistently avoided mistaking lung consolidation for a lung mass. In two instances where the mass was accompanied by pleural effusion, the Dice score was lower because some of the pleural effusion was incorrectly

identified as a mass. Lastly, the algorithm produced poor segmentations when dealing with masses containing significant intralesional mineralization, as these mineralization were not recognized as part of the mass.

4. Discussion

This is the first paper describing the development of an AI-based algorithm for the automatic segmentation of pulmonary masses in dogs, and, therefore, a side-by-side comparison with different approaches on the same topic is not possible. On the other hand, many papers describing different approaches to this problem have been published in human medicine (Hiraman et al. 2024). These studies primarily rely on the LIDC-IDRI (https://www.cancerimagingarchive.net/collection/lidc-idri/) and the Data Science Bowl 2017 (DSB-17) databases (Hiraman et al. 2024). These databases are significantly larger than the one utilized in this study, with the LIDC-IDRI encompassing 1018 cases sourced from seven academic centers and eight medical imaging companies. Meanwhile, DSB-17 includes data from 1595 patients, comprising 1397 for training purposes (362 cases of cancer and 1035 without) and 198 for testing or evaluation (57 cancer cases and 141 non-cancer). It is clear that these databases not only contain more cases but also exhibit less inherent variability due to the fewer number of different CT scanners used for data acquisition compared to what is used in this study. Nevertheless, the DSC reported by the authors (when mentioned) shows considerable variation: for instance, it ranges from 0.52 (Shimazaki et al. 2022) on the LIDC-IDRI dataset to 0.96, as reported by (Said et al. 2023) on the Decathlon lung dataset (http://medicaldecathlon.com/dataaws/).

In this research, despite utilizing a relatively small dataset comprising 217 instances (with 187 for training-validation and 30 for testing), we achieved an average Dice Similarity Coefficient (DSC) of 0.91 and an Average Symmetrical Surface Distance (ASSD) of 1.88 mm, which align with the current leading models in human medicine. As discussed in the Results section, pleural effusion and intralesional mineralization detrimentally impacted segmentation accuracy; pleural effusion was incorrectly segmented in 2/4 cases in the test set as a mass, and intralesional mineralizations were omitted from automated segmentation. We omitted cases with post-obstructive atelectasis since these are known to decrease model accuracy in human studies (Kashyap et al. 2025) but consolidation was present in different areas of the lungs in some instances. To our knowledge, the impacts of intralesional mineralizations and pleural effusion on segmentation accuracy have not been detailed in the context of human medicine.

Automated segmentation software offers several advantages over manual segmentation, particularly in time-sensitive and high-volume clinical environments. First, it significantly accelerates the segmentation process, reducing labor-intensive tasks for clinicians and allowing for faster turnaround times. Second, automated methods exhibit improved consistency and reproducibility, overcoming the well-documented inter- and intra-observer variability inherent in manual segmentations (Hiraman et al. 2024). This enhanced reliability is essential for longitudinal studies or when comparing outcomes across large patient populations. Third, automated approaches are more readily scalable, enabling analysis of extensive datasets without the prohibitive time and cost constraints typically associated with manual annotation. Finally, these advantages collectively improve overall diagnostic confidence and help streamline clinical workflows, freeing up valuable expert time for higher-level decision-making and patient care.

It is important to acknowledge that this is a multicentric study involving two university-based veterinary hospitals (University of Padua and University of Pisa) and a teleradiology service (Antech Imaging Services). Therefore, the final database had a significant variability in terms of CT scanners, imaging settings, and overall imaging quality among cases. This considerable variability in image quality and imaging settings, in our view, impacts the results of this study in two distinct ways. Firstly, this wide range of variability might reduce the accuracy of segmentations, as the network must manage not only a complex task but also the diversity among the CT scanners [Zech2018]. Conversely, this diverse database more accurately mimics the variability commonly encountered in everyday clinical activity, making the results reported in this paper more stable among different systems.

The databases of several human studies on the development of AI tools for the automatic segmentation of lung masses were developed starting from selected facilities; for example (Kashyap et al. 2025), used cases from patients undergoing radiotherapy in two different institutions. The dataset generated for

this study contained generic cases with lung masses, and therefore, other thoracic lesions, such as enlarged lymph nodes, thoracic wall masses, lung nodules, and pleural effusion, were sometimes present. The impact of other lesions on the overall accuracy of the network is difficult to assess, as different lesions might impact the accuracy of the network differently.

A key limitation of this study is that it only included masses with the largest axial dimensions of at least 2 cm, thereby excluding smaller nodules, which are far more prevalent than masses, from the dataset. The study aimed solely at developing an automatic lung mass segmentation tool and did not attempt to identify smaller lesions. The results demonstrated that the UNet consistently detected the pertinent lung lesions, as indicated by positive DSC scores. In human studies, dimensional cut-offs are rarely used as inclusion or exclusion criteria (Hunter et al. 2022). Additional research, potentially involving a greater sample size, is essential for the creation and evaluation of an AI-driven tool aimed at automatically identifying and segmenting lung nodules of any size.

5. Conclusions

In this study, we successfully developed and tested an AI-based tool for automatic segmentation of canine pulmonary masses on CT scans. Despite the relatively small dataset compared to large-scale human imaging repositories, our model demonstrated high overlap with expert annotations(DSC of 0.91) and robust performance across diverse scanning protocols, institutions, and equipment. Our findings highlight the practical benefits of automated segmentation, including time savings, reproducibility, and scalability for busy clinical or teleradiology settings. Nonetheless, challenges remain regarding segmenting smaller nodules, managing heterogeneous attenuation patterns, and standardizing ground truth annotations. Future directions will include prospective validation on additional scanners and institutions, expanded coverage to sub-2 cm nodules, and exploration of advanced image analysis approaches (e.g. radiomics) to refine diagnosis and therapy planning further. Ultimately, this research represents a step toward more efficient, reliable, and data-driven oncology workflows in veterinary medicine.

Acknowledgments

We gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Center: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2025/018194.

Author contributions

AJ and MW developed the U-Net algorithms and drafted the manuscript, TB, SB, CP, GBC, SC, GP, NM, MB, DW performed the segmentatiotion of the images, drafted the manuscript. TB, AJ, and MW.

Generative artificial intelligence (AI)

Generative Artificial Intelligence (AI) tools, namely https://chatgpt.com/ (ChatGPT o1) were used to proofread the paper.

Disclosure statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Open Access funding provided by Università degli Studi di Padova—University of Padua, Open Science Committee. The present paperwas part of a project funded by a research grant from the Department of Animal Medicine, Production and Health – MAPS,University of Padua, Italy: SID- Banzato et al. 2023 (C20.000).

Data availability statement

The data that support the findings of this study are available from the corresponding author, TB, upon reasonable request.

References

Adrien-Maxence H et al. 2022. Comparison of error rates between four pretrained DenseNet convolutional neural network models and 13 board-certified veterinary radiologists when evaluating 15 labels of canine thoracic radiographs. Vet Radiol Ultrasound. 63(4):456–468. https://doi.org/10.1111/vru.13069

Ansari MY et al. 2022. Practical utility of liver segmentation methods in clinical surgeries and interventions. BMC Med Imaging. 22(1):141. https://doi.org/10.1186/s12880-022-00825-2

Banzato T et al. 2017. Texture analysis of magnetic resonance images to predict histologic grade of meningiomas in dogs. Am J Vet Res. 78(10):1156–1162. url: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85030106300&doi=10.2460%2Fajvr.78.10.1156&partnerID=40&md5=9388ac93103c3cac51bb1b64c53f464c. https://doi.org/10.2460/ajvr.78.10.1156

Banzato T et al. 2018. A methodological approach for deep learning to distin- guish between meningiomas and gliomas on canine MR-images. BMC Vet Res. 14(1):317. https://doi.org/10.1186/s12917-018-1638-2

Banzato T et al. 2018. Development of a deep convolutional neural network to predict grading of canine meningiomas from magnetic resonance images. Vet J. 235:90–92. https://doi.org/10.1016/j.tvjl.2018.04.001

Banzato T et al. 2023. An AI-based algorithm for the automatic evaluation of image quality in canine thoracic radiographs. Sci Rep. 13(1):17024. https://doi.org/10.1038/s41598-023-44089-4

Begum Ergun G, Guney S. 2021. Classification of canine maturity and bone fracture time based on X-ray images of long bones. IEEE Access. 9:109004–109011. https://doi.org/10.1109/ACCESS.2021.3101040

Boissady E et al. 2020. Artificial intelligence evaluating primary thoracic lesions has an overall lower error rate compared to veterinarians or veterinarians in conjunction with the artificial intelligence. Vet Radiol Ultrasound. 616:619–627. https://doi.org/10.1111/vru.12912

Burti S et al. 2020. Use of deep learning to detect cardiomegaly on thoracic radiographs in dogs. The Veterinary Journal. 2622020:105505. In: (), 105505issn: URL: https://doi.org/10.1016/j.tvjl.2020.105505. https://doi.org/10.1016/j.tvjl.2020.105505

Burti S et al. 2024. Artificial intelligence in veterinary diagnostic imaging: perspectives and limitations. Res Vet Sci. 175:105317. https://doi.org/10.1016/j.rvsc.2024.105317

Hiraman A, Viriri S, Gwetu M. 2024. Lung tumor segmentation: a review of the state of the art. Front Comput Sci. 6. https://doi.org/10.3389/fcomp.2024.1423693

Hunter B et al. 2022. A radiomics-based decision support tool improves lung cancer diagnosis in combination with the Herder score in large lung nodules. EBioMedicine. 86(Dec. 2022):104344. https://doi.org/10.1016/j.ebiom.2022.104344

Isensee F et al. 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 18(2):203–211. DOI: https://doi.org/10.1038/s41592-020-01008-z

Isensee F et al. 2024. nnU-Net revisited: a call for rigorous validation in 3D medical image segmentation. In: Linguraru MG, et al., editors. Medical image computing and computer assisted intervention – MICCAI 2024. Switzerland: Springer Nature. p. 488–498. https://doi.org/10.1007/978-3-031-72114-4_47

Ji Y et al. 2022. A deep learning model for CT-based kidney volume determination in dogs and normal reference definition. Front Vet Sci. 9:1011804. https://doi.org/10.3389/fvets.2022.1011804

Kashyap M et al. 2025. Automated deep learning-based detection and segmentation of lung tumors at CT. Radiology. 314(1):e233029. https://doi.org/10.1148/radiol.233029

Marcinowska A et al. 2024. Canine lung carcinoma—a descriptive review. Front Vet Sci. 11. https://doi.org/10.3389/fvets.2024.1464659

McEvoy FJ et al. 2021. Deep transfer learning can be used for the detection of hip joints in pelvis radiographs and the classification of their hip dysplasia status. Vet Radiol Ultrasound. 624:387–393. https://doi.org/10.1111/vru.12968

Müller V, Solano M, Tsunemi MH. 2022. Accuracy of artificial intelligence software for the detection of confirmed pleural effusion in thoracic radiographs in dogs. Vet Radiol Ultrasound. 63: 573–579. https://doi.org/10.1111/vru.13089

Nguyen SM et al. 2015. Response evaluation criteria for solid tumours in dogs (v1.0): a Veterinary Cooperative Oncology Group (VCOG) consensus document. Vet Comp Oncol. 13(3):176–183. https://onlinelibrary.wiley.com/doi/10.1111/vco.12032. https://doi.org/10.1111/vco.12032

Park J et al. 2021. Deep-learning-based automatic segmentation of head and neck organs for radiation therapy in dogs. Front Vet Sci. 8:721612. https://doi.org/10.3389/fvets.2021.721612

Ronneberger O, Fischer P, Brox T 2015. U-Net: convolutional networks for biomedical image segmentation. In: Navab N et al., editors. Medical image computing and computer-assisted intervention – MICCAI 2015. Springer International Publishing. p. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Said Y et al. 2023. Medical images segmentation for lung cancer diagnosis based on deep learning architectures. Diagnostics. 13(3):546. https://doi.org/10.3390/diagnostics13030546

Schmid D et al. 2022. Employing deep convolutional neural networks for segmenting the medial retropharyngeal lymph nodes in CT studies of dogs. Vet Radiol Ultrasound. 63:763–770. https://doi.org/10.1111/vru.13132

Shimazaki A et al. 2022. Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method. Sci Rep. 12(1):727. https://doi.org/10.1038/s41598-021-04667-w

Tahghighi P et al. 2024. Classification of the quality of canine and feline ventrodorsal and dorsoventral thoracic radiographs through machine learning. IEEE Access. 417–427. https://doi.org/10.1111/vru.13373https://onlinelibrary.wiley.com/doi/10.1111/vru.13373

Vail DM, Thamm DH, Liptak JM. 2020. 24 - Tumors of the respiratory system. In: Vail DM, Thamm DH, and Liptak JM. 6th ed. W.B. Saunders. p. 492–523. https://doi.org/10.1016/B978-0-323-59496-7.00024-4

Valente C et al. 2023. Development of an artificial intelligence-based method for the diagnosis of the severity of myxomatous mitral valve disease from canine chest radiographs. Front Vet Sci. 10:1227009. https://doi.org/10.3389/fvets.2023.1227009