Medical image segmentation, the process of delineating a medical image into distinct anatomical structures, is an essential technology in diagnostic imaging that enables clinicians to precisely analyze complex structures. In human medicine, deep-learning-based segmentation techniques have greatly improved clinical workflows. However, in veterinary medicine, segmentation techniques are still in their early stages of development. Current practice relies primarily on manual segmentation, which is labor-intensive and time-consuming. To the best of our knowledge, no previous studies have applied deep learning-based automated segmentation specifically targeting the canine liver, despite its clinical relevance in veterinary medicine. Accurate liver segmentation is beneficial because the precise quantification of liver volume aids in assessing conditions such as hepatomegaly or microhepatica. Against this background, this study investigates deep-learning-based liver segmentation on canine CT across three complementary levels.

The first part of this research aimed to develop and validate a deep-learning-based model specifically designed for canine liver segmentation using a 3D U-Net architecture. To verify clinical applicability, model performance was evaluated in two distinct settings: a baseline assessment using livers without focal lesions, and a comprehensive assessment including hepatic masses to test robustness in clinically heterogeneous conditions. The dataset consisted of 220 post-contrast CT scans from 205 dogs collected at Hokkaido University Veterinary Teaching Hospital. Experiment 1 used 158 scans without hepatic masses, and Experiment 2 used the combined dataset including 62 cases with hepatic masses. A 3D U-Net architecture was employed, and the model was trained using patches of 96×96×96 voxels.

The model trained on cases without hepatic masses in Experiment 1 demonstrated strong performance, with a mean Dice similarity coefficient (DSC) of 0.925 and an Intersection over Union (IoU) of 0.863. The model trained on the combined dataset in Experiment 2 achieved a slightly higher segmentation performance, with a DSC of 0.928 and an IoU of 0.868. However, a subset analysis revealed that the segmentation performance was higher for cases without hepatic masses (DSC = 0.930) than for those with hepatic masses (DSC = 0.924). These results indicate that the model reliably segments the liver in cases without significant anatomical complexities. Nevertheless, the model exhibited increased segmentation variability and a slight decrease in performance in cases involving hepatic masses, likely because of the limited number of such cases in the dataset. The current findings clearly demonstrate that deep learning-based segmentation models can effectively overcome these complexities and achieve competitive performance comparable to human medical studies.

While deep-learning-based segmentation models are commonly assessed using geometric overlap metrics, these do not fully capture clinical applicability or workflow efficiency. Therefore, the second part of this study aimed to characterize and compare the segmentation outputs of the deep learning model with those produced by a clinical expert and a non-expert across a multi-axis framework of quantitative and qualitative traits. Additionally, a semi-automatic workflow, in which a human operator edits the deep-learning-generated masks, was evaluated. Ten post-contrast CT cases were selected. Efficiency was assessed using Active Task Time (ATT) and Total Turnaround Time (TTT). Qualitative traits, including anatomical completeness, anatomical plausibility, and topological integrity, were evaluated by a veterinary radiology specialist.

Time analysis showed that the deep-learning-based workflow was much faster, with a median TTT of 9.4 seconds and ATT of 4.0 seconds, compared to manual workflows which required over 800 seconds. The semi-automatic workflow, with a median TTT of 240.9 seconds, reduced the total time to approximately one-third of the expert manual workflow (865.5 seconds). Qualitatively, the deep-learning-based masks displayed a distinct error pattern different from human operators. Some areas were missing as relatively sharp, rectangular blocks along the contour, known as tiling artifacts. However, in the semi-automatic workflow, manual editing eliminated all major inclusions and reduced the range of included organs to a level comparable with the expert masks. Consequently, all semi-automatic masks met the clinical usability criterion. These findings suggest that a minimal human-in-the-loop correction can effectively address both anatomical and topological errors, yielding segmentation masks suitable for clinical application.

Assessing liver size is clinically important, but radiographic measurements such as the gastric-axis angle are indirect estimates. CT volumetry is the reference standard but is limited by time constraints. Building on the previous results, the third part of this study adopts a semi-automatic approach to perform CT liver volumetry efficiently. It evaluates the association between actual CT-derived liver volume and radiographic indices, specifically the gastric-axis angle and the radiographic liver length-to-T11 ratio. A total of 57 CT-radiograph pairs were included. Normalized liver volume (NLV) was modeled against the gastric-axis angle with vertebral reference (GAA-VRT), gastric-axis angle with rib-space reference (GAA-RIB), and radiographic liver length-to-T11 ratio (RLL:T11).

All radiographic indices showed a significant positive association with NLV. Specifically, as the gastric axis tilted caudally, the liver tended to be larger. However, the coefficients of determination were modest, with R-squared values ranging from 0.10 to 0.17, indicating that the indices captured the overall trend but left most of the volumetric variation unexplained. When grouped into the conventional three radiographic categories, agreement with CT-based liver-size classification was only fair, with an overall accuracy of 49%. The most marked finding was the extremely low positive predictive value for microhepatica at 4.3%. This primarily reflected the very low prevalence of true microhepatica in this sample. Only one of the 23 dogs labeled radiographically as having a small liver actually had reduced NLV on CT. Consequently, most radiographic impressions of a small liver represented false positives, consistent with the general concern that radiographic assessment is influenced by body conformation. These findings indicate that commonly used radiographic indices function mainly as rough directional cues rather than precise volumetric surrogates. The semi-automatic CT workflow provides a practical means of obtaining reliable liver volumes and a scalable framework for future multi-breed studies.

The aim of this study was to make CT-based liver volumetry more accessible in routine practice by developing a deep-learning-based liver segmentation model, examining how it behaves in realistic workflows, and using it to relate everyday radiographic indices to actual liver volume. The study demonstrated that accurate automatic delineation of the canine liver is achievable using a 3D U-Net, even in a population with diverse breeds. Furthermore, it showed that a human-in-the-loop, semi-automatic workflow is a realistic way to apply deep learning in clinical practice, offering a good balance between speed and reliability. Finally, the application of this workflow revealed that radiographic indices are limited as precise volumetric surrogates and that advanced volumetric imaging is necessary for accurate assessment. In conclusion, deep-learning-based liver segmentation can provide accurate and efficient CT volumetry in dogs. Future studies should aim to validate the model across multiple centers and investigate how deep-learning-assisted volumetry can be integrated into routine decision-making for a broader range of hepatobiliary diseases.