



OPEN ACCESS

EDITED BY
Blaz Cugmas,
University of Latvia, Latvia

REVIEWED BY
Vahideh Ghobadi,
Universiti Putra Malaysia, Malaysia
Mohammed Sallah,
Mansoura University, Mansoura, Egypt

*CORRESPONDENCE
Mitsuyoshi Takiguchi
✉ mtaki@vetmed.hokudai.ac.jp

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 07 August 2025
ACCEPTED 03 October 2025
PUBLISHED 21 October 2025

CITATION
Lee S, Shimbo G, Yokoyama N, Nakamura K, Togo R, Ogawa T, Haseyama M and Takiguchi M (2025) Deep-learning-based automatic liver segmentation using computed tomography images in dogs.
Front. Vet. Sci. 12:1681820.
doi: 10.3389/fvets.2025.1681820

COPYRIGHT
© 2025 Lee, Shimbo, Yokoyama, Nakamura, Togo, Ogawa, Haseyama and Takiguchi. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Deep-learning-based automatic liver segmentation using computed tomography images in dogs

Seungyeon Lee^{1†}, Genya Shimbo^{2†}, Nozomu Yokoyama¹, Kensuke Nakamura¹, Ren Togo³, Takahiro Ogawa³, Miki Haseyama³ and Mitsuyoshi Takiguchi^{1*}

¹Laboratory of Veterinary Internal Medicine, Department of Veterinary Clinical Science, Graduate School of Veterinary Medicine, Hokkaido University, Sapporo, Japan, ²Veterinary Teaching Hospital, Graduate School of Veterinary Medicine, Hokkaido University, Sapporo, Japan, ³Faculty of Information Science and Technology, Hokkaido University, Sapporo, Japan

Introduction: Deep learning-based automated segmentation has significantly improved the efficiency and accuracy of human medicine applications. However, veterinary applications, particularly canine liver segmentation, remain limited. This study aimed to develop and validate a deep learning model based on a 3D U-Net architecture for automated liver segmentation in canine abdominal computed tomography (CT) scans.

Methods: A total of 221 canine abdominal CT scans were analyzed, comprising 159 cases without hepatic masses and 62 cases with hepatic masses. The model was trained and evaluated using two separate datasets: one containing cases without hepatic masses (Experiment 1) and the other combining cases with and without hepatic masses (Experiment 2).

Results: Both experiments demonstrated high segmentation performance, achieving mean Dice similarity coefficients of 0.926 (Experiment 1) and 0.929 (Experiment 2).

Discussion: The manual and predicted liver volumes showed excellent agreement, highlighting the potential clinical applicability of this approach.

KEYWORDS

artificial intelligence, deep learning, automatic segmentation, liver, canine, dog, computed tomography

1 Introduction

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 by enhancing diagnostic accuracy and efficiency (1–3). These advancements support clinical applications, such as surgical planning, radiation therapy, disease monitoring, and medical education (4). Among the various organs, the liver is a key target for segmentation because of its complex anatomy and role in numerous diseases. Accurate segmentation of liver structures facilitates quantification tasks such as liver volumetry, which are essential for clinical decision-making, including surgical resections and disease management (5–7). Recent deep learning-based segmentation models have demonstrated the capability to accurately segment liver structures from computed tomography (CT) images, achieving performance comparable to that of expert manual segmentation (8, 9).

However, in veterinary medicine, segmentation techniques are still in their early stages of development (10, 11). Current practice relies primarily on manual segmentation, which is labor-intensive and time-consuming, hindering the consistency and efficiency of diagnostic workflows. To date, only a few studies have explored automated segmentation in veterinary imaging. Specifically, deep learning has proven effective for the volumetric analysis of the canine kidneys and adrenal glands, the identification of abnormalities like kidney calculi, and the segmentation of complex structures such as the medial retropharyngeal lymph nodes, demonstrating feasibility in dogs (12–15). Most of these reports employed U-Net or transformer-derived models. Overall, favorable performance was reported for kidney and adrenal gland volumetry and for renal calculi detection, whereas lower performance was observed for medial retropharyngeal lymph-node segmentation, which the authors attributed to the nodes' irregular morphology and their close proximity to surrounding soft tissues, and the limited size of their single-center dataset.

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, which are important indicators of underlying diseases (16, 17). Moreover, precise volumetric measurements have proven useful for monitoring postoperative liver regeneration, such as in dogs undergoing surgical correction of portosystemic shunts (PSS) (18–20). Given these clinical implications, efforts have been made in veterinary medicine to optimize liver segmentation procedures through simplified manual methods, such as reducing the total number of CT slices and reorienting the imaging plane (21, 22). Although these approaches partially reduce the workload, they rely on manual segmentation.

Given these challenges, there is a clear need to develop automated and reliable segmentation methods for canine liver CT imaging. Therefore, we aimed to develop a deep learning-based automated segmentation model to reduce the manual effort and time required while maintaining a segmentation accuracy comparable to that of expert manual delineation.

2 Materials and methods

2.1 Dataset for CT scans

This retrospective study utilized post-contrast CT scans of dogs collected at Hokkaido University Veterinary Teaching Hospital (HUVTH), Japan. Approval from the Animal Care and Use Committee of our institution was not required because of the retrospective nature of the study. A total of 221 CT scans from 206 dogs were included, with some dogs having undergone multiple CT scans for various clinical reasons. All medical records used in this study were fully anonymized. Owing to the limited number of samples, the evaluation was performed at the individual image level rather than at the individual dog level. Therefore, the separation of scans from the same dog across the training, validation, and test sets was not guaranteed. Comprehensive medical records

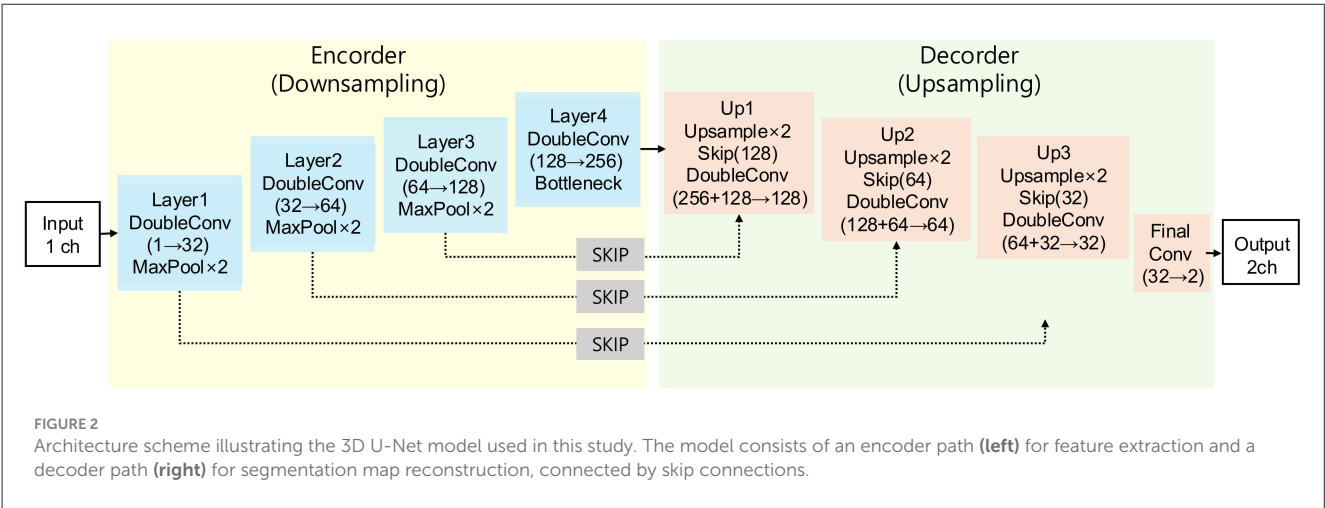
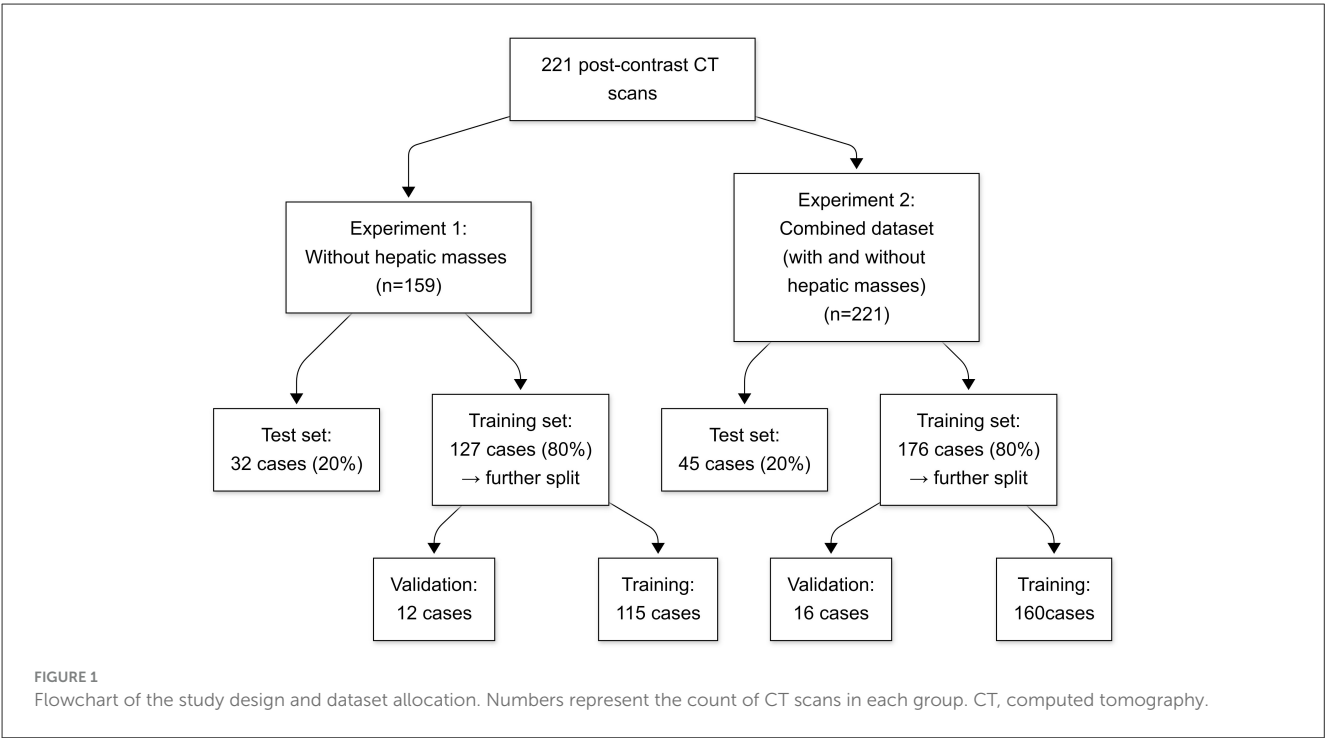
including breed and body weight were reviewed to provide a detailed characterization of the study population.

The data were structured into two distinct datasets to evaluate the model performance under various liver conditions. The first dataset consisted of 159 CT scans without hepatic masses, obtained from 147 dogs. As an initial step toward automatic liver segmentation, this dataset specifically included scans that demonstrated normal hepatic anatomy. The second dataset combined cases with and without hepatic masses, resulting in 221 CT scans from 206 dogs. This dataset included all the scans from the first dataset and an additional 62 scans from 60 dogs with hepatic masses. Because many abdominal CT scans performed at HUVTH are primarily intended for surgical planning, hepatic masses were included in the second dataset. Cases that could alter hepatic morphology owing to factors other than hepatic masses, such as cirrhosis or PSS, as well as those with a large amount of ascites were excluded because of their limited number, which could adversely affect the robustness of the deep learning model. In the case of PSS, the timing of imaging acquisition at HUVTH differed from the standard imaging protocol used for other abdominal cases, potentially introducing inconsistencies in the training data. Both datasets were randomly shuffled and initially split into training and test sets in an 80:20 ratio. The training set was further subdivided by allocating 10% of the data for validation. For the second dataset, particular attention was paid to ensure a balanced representation of cases with and without masses across the training, validation, and test sets. This stratified splitting aimed to minimize bias during model training and evaluation, thereby providing a robust assessment of the generalizability of the model. The overall study design and the detailed allocation of CT scans into the datasets are summarized in Figure 1.

CT scans were acquired using an 80-row multidetector CT scanner (Aquilion PRIME; Toshiba Medical Systems, Tochigi, Japan) with standardized imaging protocols. The scanning parameters included tube voltage settings of 80–120 kVp, Auto Exposure Control (Sure Exposure 3D, Toshiba Medical Systems, Tochigi, Japan), a 512×512 matrix, slice thickness of 0.5 mm, helical pitch of 0.813, and a 0.5-s rotation time. All dogs were anesthetized with propofol induction and maintained using isoflurane inhalation during the CT procedures, positioned in either the dorsal or sternal recumbency position. The contrast agents iohexol (Omnipaque 300, GE Healthcare Co., Ltd., Tokyo, Japan) and iopamidol (Iopamidol 150, Fuji Pharma Co., Ltd., Tokyo, Japan) were administered at a dose of 600 mgI/kg via the cephalic vein using a power injector (Dual Shot, Nemoto Kyorindo Co., Ltd., Tokyo, Japan), with an injection duration of 20 s. Segmentation was conducted using portal venous phase images, although the exact timing of portal venous phase acquisition after contrast injection varied slightly depending on the case.

2.2 Manual segmentation

To obtain an accurate and consistent ground truth, a single-annotator segmentation protocol was implemented. A highly skilled veterinary radiology technologist with 15 years of experience performed all manual segmentation on a 3D image



analysis workstation (Synapse Vincent, Fujifilm Corporation, Tokyo, Japan). The intrahepatic vessels were included within the operator-defined region of interest. The segmentation results were subsequently reviewed by a veterinary radiology specialist with >10 years of experience (G.S.) to ensure accuracy.

The CT scans were preprocessed to standardize the dataset and ensure compatibility with the deep learning model. All preprocessing steps were implemented using Python, PyTorch and TorchIO library. Binary masks were generated from the manually segmented liver labels by assigning a voxel value of 1 to the liver and 0 to all other areas. This conversion enabled the model to clearly distinguish the liver from the surrounding structures during training. Subsequently, CT images were normalized by scaling their voxel intensities to a range of $[-1, 1]$ using TorchIO's RescaleIntensity transform, thereby enhancing the model's training efficiency. To ensure a uniform input size for the network, all volumes were resized to $256 \times 256 \times 200$ voxels using TorchIO's

Resize transform, with linear interpolation applied to the images and nearest-neighbor interpolation to the masks.

2.3 Model architecture

In this study, we employed a three-dimensional (3D) U-Net architecture for liver segmentation of CT images (Figure 2). The network architecture was based on the original U-Net and was modified to support volumetric data by replacing 2D operations with their corresponding 3D counterparts, including 3D convolution, 3D max-pooling, and trilinear upsampling methods (23). The 3D U-Net model uses an encoder-decoder format designed to effectively capture spatial features from 3D CT data. The encoder path includes 3D convolutions and pooling layers to progressively extract higher-level spatial features, whereas the decoder mirrors this structure using transposed convolutions to

restore spatial resolution and facilitate precise localization of the liver regions. Skip connections between the encoder and decoder allow the integration of both high- and low-level features, thereby preserving the spatial details and enhancing the segmentation accuracy. These characteristics make 3D U-Net highly suitable for volumetric data and for addressing the anatomical variability of canine livers. In this study, preprocessed CT images and the corresponding binary label masks were used as inputs to the model.

2.4 Network parameters

The model was trained for 100 epochs using an NVIDIA RTX 4090 GPU. A stochastic gradient descent optimizer was utilized, configured with a batch size of four, learning rate of 0.01, momentum of 0.9, and weight decay set to 1×10^{-4} . The learning rate remained constant throughout the training process.

The input data comprised grayscale CT images with a single input channel, and the network produced a two-channel output that represented the background and liver regions. The primary loss function employed was cross-entropy loss, which is a standard and widely adopted choice for multiclass segmentation tasks in medical imaging. To efficiently manage large 3D volumes, the training data were divided into smaller patches of size $96 \times 96 \times 96$ pixels, with each patch overlapping by eight pixels. These patches were systematically extracted using the TorchIO GridSampler, whereas predictions from the model were aggregated into full-volume segmentations during the evaluation using the TorchIO GridAggregator. This patch-based approach optimizes memory usage while maintaining spatial coherence in the segmentation results.

Model performance was monitored during training through periodic evaluations of the validation set. The validation accuracy served as the primary metric for checkpointing, and the top 10 epochs were saved based on this metric to ensure optimal model performance during testing.

2.5 Evaluation metrics

To evaluate the model performance, several quantitative metrics were employed to assess the segmentation accuracy between the predicted liver masks and ground truth masks on the test set. The metrics were calculated for each test sample and averaged across the entire dataset to obtain the overall performance indicators. These were computed on the reassembled full-volume predictions aggregated from patch-based evaluations, as described in the Implementation Details section.

The Dice similarity coefficient (DSC) was used to measure the overlap between the predicted and ground truth segmentation masks. This is defined as the ratio of twice the area of intersection to the total area of both the predicted and ground truth masks. The formula used is as follows:

$$DSC = \frac{2 \times |P \cap G|}{|P| + |G|} \quad (1)$$

where P and G denote the predicted and ground truth masks, respectively.

Sensitivity was calculated to evaluate the ability of the model to correctly identify the positive pixels in the segmentation mask. This is defined as the ratio of true-positive pixels to the sum of true-positive and false-negative pixels:

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

Specificity measures the ability of the model to correctly identify negative pixels (background) and is calculated as the ratio of true-negative pixels to the sum of true-negative and false-positive pixels:

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \quad (3)$$

Precision quantifies the proportion of true-positive pixels among all pixels classified as positive by the model:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (4)$$

The overall accuracy of the model was calculated as the ratio of correctly classified pixels (positive and negative) to the total number of pixels in the segmentation mask:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Pixels}} \quad (5)$$

The intersection-over-union (IoU), also known as the Jaccard Index, measures the overlap between the predicted and ground truth masks by dividing the intersection area by the union area:

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|} \quad (6)$$

The volume ratio was computed to assess the volumetric accuracy of the segmentation. This is defined as the ratio of the predicted liver volume to the ground truth liver volume. Each liver volume was calculated by multiplying the number of positive voxels in the respective segmentation mask by the voxel volume derived from the original CT scan as follows:

$$\text{Volume ratio} = \frac{\text{Number of positive Voxels in Predicted Mask} \times \text{Voxel volume}}{\text{Number of positive Voxels in Ground Truth Mask} \times \text{Voxel volume}} \quad (7)$$

This patch-based evaluation approach, combined with metrics computed from aggregated full-volume predictions, ensured consistency and reliability in assessing segmentation accuracy. All metrics were implemented using Python and PyTorch and calculated voxel-wise across the entire 3D volume. In addition, Pearson correlation coefficient (r) was calculated to assess the linear relationship between the predicted and manually measured liver volumes.

3 Results

3.1 Evaluation of liver segmentation on the dataset without hepatic masses (Experiment 1)

This experiment used a dataset of 159 CT scans without hepatic masses, divided into 115, 12, and 32 scans for training,

TABLE 1 Characteristics of the study datasets, including breed distribution, median body weight, and median liver volume.

Group	Number of cases	Median body weight (kg) [IQR]	Median liver volume (mL) [IQR]	Breeds included (number of dogs)
Without hepatic masses	159	9.05 [6–18]	297 [192–511]	Mixed (22), Miniature Dachshund (21), Shiba (12), Toy Poodle (12), Labrador Retriever (9), French Bulldog (8), Golden Retriever (8), Miniature Schnauzer (6), Pomeranian (6), Jack Russell Terrier (5), Shetland Sheepdog (5), Bernese Mountain Dog (4), Border Collie (4), Bulldog (4), Chihuahua (4), Samoyed (4), Shih Tzu (4), Welsh Corgi (4), American Cocker Spaniel (2), Bichon Frise (2), Hokkaido (2), Italian Greyhound (2), Pekingese (2), Boston Terrier (1), Cavalier King Charles Spaniel (1), English Cocker Spaniel (1), Flat-coated Retriever (1), Miniature Pinscher (1), Scottish Terrier (1), Standard Poodle (1)
With hepatic masses	62	7.55 [5.03–10.95]	427.5 [245.5–721.5]	Miniature Dachshund (11), Mixed (11), Toy Poodle (9), Shiba (4), Shih Tzu (4), Border Collie (3), Samoyed (3), Yorkshire Terrier (3), Miniature Schnauzer (2), Welsh Corgi (2), Chihuahua (1), French Bulldog (1), Jack Russell Terrier (1), Labrador Retriever (1), Papillon (1), Scottish Terrier (1), Shetland Sheepdog (1), Standard Poodle (1), West Highland White Terrier (1), Wire Fox Terrier (1)
Overall total	221	8.4 [5.53–14.88]	318 [204–563]	Mixed (33), Miniature Dachshund (32), Toy Poodle (21), Shiba (16), Labrador Retriever (10), French Bulldog (9), Miniature Schnauzer (8), Shih Tzu (8), Golden Retriever (8), Border Collie (7), Samoyed (7), Jack Russell Terrier (6), Pomeranian (6), Shetland Sheepdog (6), Welsh Corgi (6), Chihuahua (5), Bernese Mountain Dog (4), Bulldog (4), Yorkshire Terrier (3), American Cocker Spaniel (2), Bichon Frise (2), Hokkaido (2), Italian Greyhound (2), Pekingese (2), Scottish Terrier (2), Standard Poodle (2), Boston Terrier (1), Cavalier King Charles Spaniel (1), English Cocker Spaniel (1), Flat-coated Retriever (1), Miniature Pinscher (1), Papillon (1), West Highland White Terrier (1), Wire Fox Terrier (1)

Values are presented as median [IQR]. IQR, Interquartile range.

TABLE 2 Quantitative segmentation performance metrics of the two experiments on the test set.

Experiment	DSC	IoU	Sensitivity	Specificity	Precision	Accuracy	Volume ratio
Exp 1: Without hepatic masses	0.926	0.865	0.946	0.995	0.910	0.992	1.042
Exp 2: Combined dataset	0.929	0.868	0.931	0.995	0.928	0.991	1.006
Without hepatic masses subset	0.931	0.872	0.942	0.995	0.921	0.992	1.025
With hepatic masses subset	0.924	0.861	0.906	0.994	0.945	0.987	0.959

DSC, dice similarity coefficient; IoU, intersection-over-union.

validation, and testing, respectively. [Table 1](#) provides an overview of these scans, including the number of dogs, median body weight, and median liver volume. Regarding the training performance, the model showed stable convergence, with mean training and validation losses of 0.124 and 0.125, respectively ([Figure 3A](#)). For the test set, the model achieved an accuracy of 0.981 and a loss of 0.05. The detailed segmentation metrics are summarized in [Table 2](#), revealing a mean DSC of 0.926, IoU of 0.865, and volume ratio of 1.042, indicating strong volumetric agreement between the predicted and ground truth masks. The additional metrics included sensitivity (0.946), specificity (0.995), precision (0.910), and accuracy (0.992). [Figures 4A–D](#) illustrates a representative normal liver slice and its segmentation overlay, demonstrating the precise delineation of the liver boundaries.

3.2 Evaluation of liver segmentation on the combined dataset of cases with and without hepatic masses (Experiment 2)

In this experiment, the original 159 CT scans without hepatic masses were supplemented with 62 cases exhibiting hepatic masses, resulting in 221 CT scans, which were divided into 160 for training,

16 for validation, and 45 for testing. [Table 1](#) summarizes the characteristics of both datasets, including the breed distribution, median body weight, and median liver volume. The model trained on this more heterogeneous dataset exhibited a mean training loss of 0.170 and validation loss of 0.132 ([Figure 3B](#)). For the test set, the model achieved an accuracy of 0.980 and a loss of 0.057. [Table 2](#) shows a mean DSC of 0.929, an IoU of 0.868, and volume ratio of 1.006, which is slightly higher than that in Experiment 1. Sensitivity (0.931), specificity (0.995), precision (0.928), and accuracy (0.991) remained high. However, a subset analysis of the test set, which involved dividing it into cases with and without hepatic masses, revealed that the segmentation performance was higher for cases without hepatic masses (DSC = 0.931) than for those with hepatic masses (DSC = 0.924). [Figures 4E–H](#) shows a sample slice from the liver with a hepatic mass, demonstrating greater morphological variability owing to the presence of the hepatic mass.

3.3 Correlation analysis between predicted and manually measured liver volumes

[Figure 5](#) illustrates the correlation between the manually measured and predicted liver volumes in both experiments.

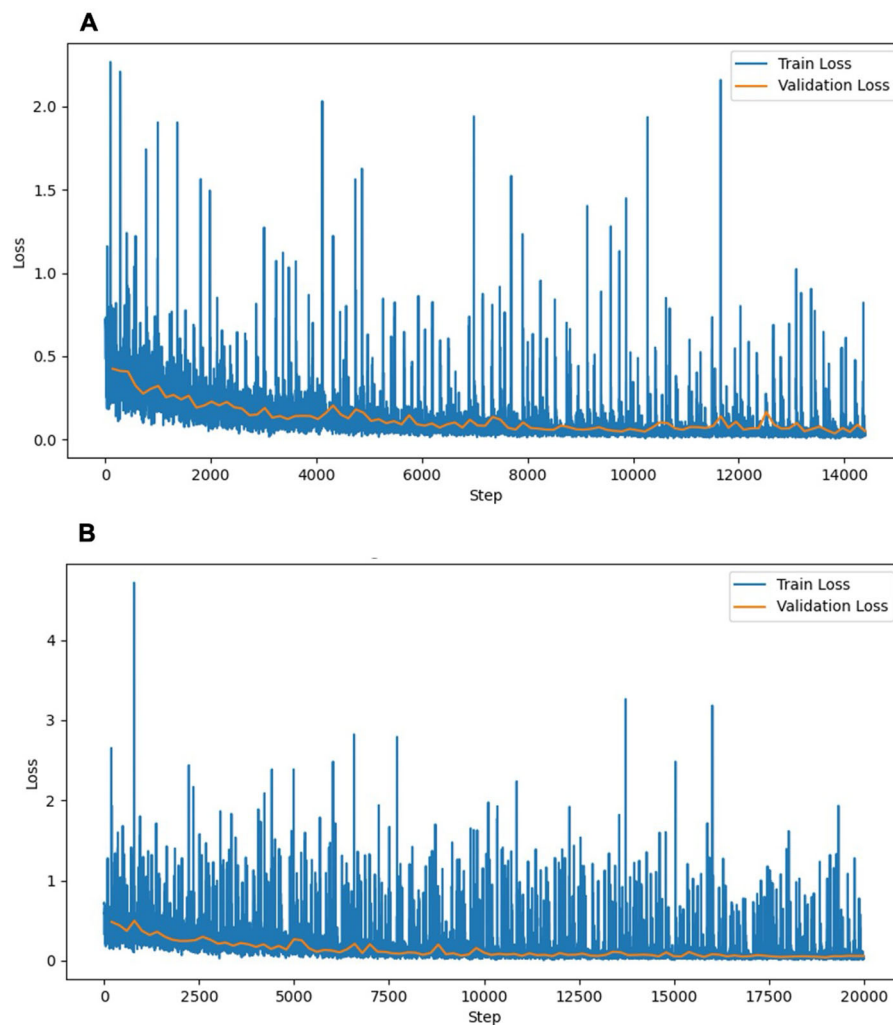


FIGURE 3

Training and validation loss curves of the liver segmentation models. (A) Experiment 1 (without hepatic masses). (B) Experiment 2 (combined dataset with and without hepatic masses). The training loss (blue) and validation loss (orange) decreased steadily over the training steps, indicating stable convergence without notable overfitting. The training loss curve in Experiment 2 exhibited higher fluctuations, reflecting the increased complexity associated with the diverse morphological features of hepatic masses. The validation loss curves remained similarly stable across both experiments, indicating consistent generalization.

Experiment 1 showed a strong correlation ($r = 0.995$, $R^2 = 0.990$), and Experiment 2, which was conducted using a combined dataset of cases with and without hepatic masses, demonstrated slightly higher agreement ($r = 0.997$, $R^2 = 0.993$). Most predictions were closely aligned with the ideal correlation line, although slightly increased variability was observed in Experiment 2.

4 Discussion

This study is the first to develop a deep learning-based model specifically for canine liver segmentation in CT images, addressing the existing gap in veterinary medicine. The model trained exclusively on cases without hepatic masses (Experiment 1) demonstrated strong performance, with a mean DSC of 0.926 and an IoU of 0.865 (Table 2). These results indicate that the model reliably segments the liver

in cases without significant anatomical complexities such as hepatic masses. Establishing such a performance baseline is essential because it provides a clear reference for interpreting segmentation performance in anatomically more complex cases involving hepatic masses. This baseline helps clarify whether decreased segmentation performance results primarily from anatomical variations caused by lesions, leading to distributional shifts, or from the intrinsic limitations of the segmentation model itself. A previous study highlighted the influence of distributional shifts on model performance when encountering previously unseen anatomical characteristics (24). Thus, the results of Experiment 1 serve as a robust benchmark for subsequent evaluations involving more diverse and complex hepatic conditions.

In contrast, the model trained on a combined dataset, which included cases with and without hepatic masses (Experiment 2), achieved a slightly higher segmentation performance, with a DSC

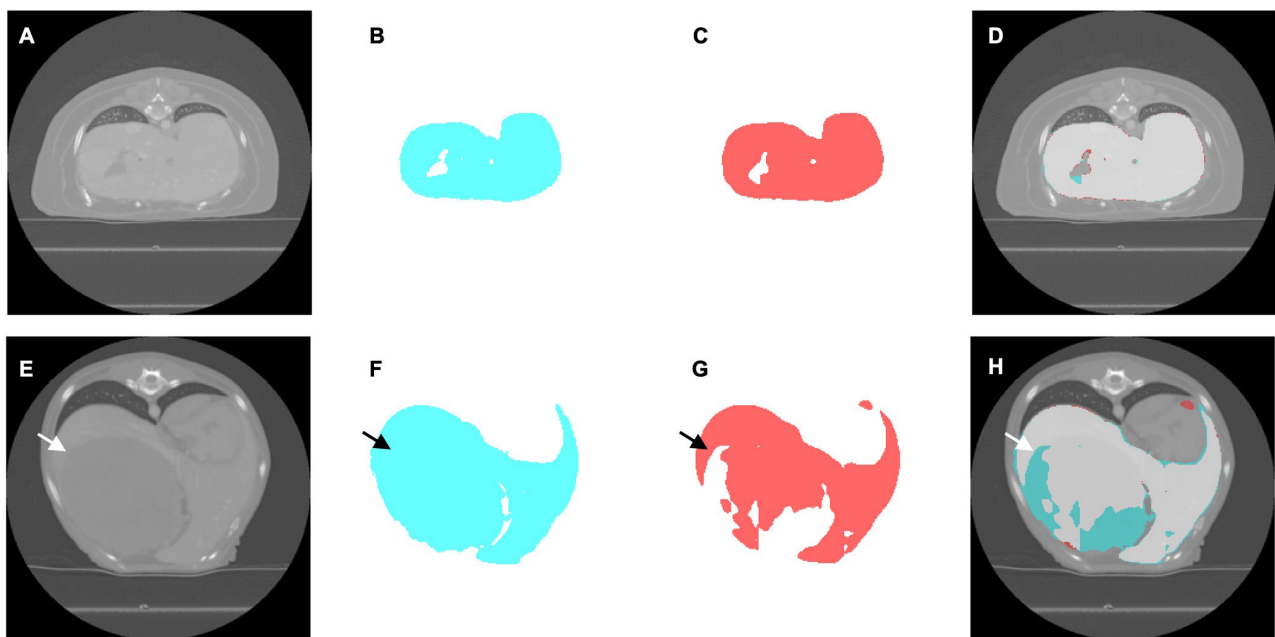


FIGURE 4

Representative computed tomography (CT) slices illustrating liver segmentation in cases without (A–D) and with (E–H) hepatic masses. (A, E) Original CT image. (B, F) Ground truth segmentation masks generated by expert radiologists (cyan). (C, G) Segmentation masks predicted using the deep learning model (red). (D, H) Overlay images demonstrating segmentation accuracy, indicating overlapping regions between the ground truth and model predictions (light gray), regions segmented only by experts (cyan), and regions segmented only by the model (red). The arrows indicate the location of the hepatic mass.

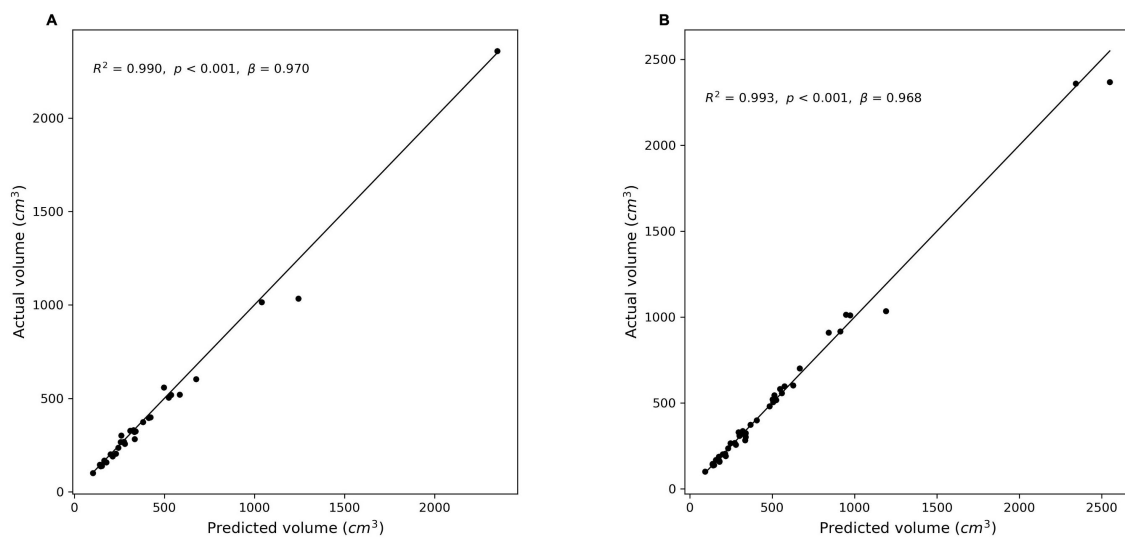


FIGURE 5

Correlation between predicted and manually measured liver volumes. (A) Experiment 1 (without hepatic masses). (B) Experiment 2 (combined dataset with and without hepatic masses). Scatter plots illustrate the relationship between the actual (manually segmented) liver volumes and the volumes predicted by the segmentation models. The diagonal line ($y = x$) represents perfect prediction. (A): $r = 0.995$, $R^2 = 0.990$, $p < 0.001$; (B): $r = 0.997$, $R^2 = 0.993$, $p < 0.001$.

of 0.929 and an IoU of 0.868 (Table 2), suggesting that the increased dataset size and diversity may contribute to improved model robustness. Human medical research has previously demonstrated that training on diverse datasets covering a broad age range and

multiple abnormalities such as tumors, inflammation, and vascular disorders can enhance model generalizability (25). Similarly, our study showed improved general segmentation performance when the training included cases with and without hepatic

masses. 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 our dataset. This limitation may have restricted the ability of the model to fully learn the morphological variability associated with hepatic masses.

Further insights were gained through the analysis of the volume ratio metric, which is defined as the ratio between the predicted liver volume and the volume derived from the ground truth masks. This metric demonstrated a strong agreement between the predicted and ground truth volumes. However, because it does not directly assess spatial accuracy, it should be complemented with spatially sensitive methods. Relying solely on the volume ratio may overlook spatial discrepancies caused by complex or irregular morphologies. Thus, it is important to use a combination of different evaluation metrics to assess distinct aspects and provide complementary information (26). Spatially sensitive metrics such as DSC and IoU, which evaluate spatial overlap, are valuable for assessing anatomically irregular cases (27). In this study, we comprehensively evaluated segmentation performance by incorporating multiple metrics.

The performance of the model closely aligns with that of the leading liver segmentation methods used in medical imaging. For instance, top-ranking algorithms in the Liver Tumor Segmentation Challenge consistently report DSCs between 0.93 and 0.96, which is comparable to expert manual segmentation (28). This comparison highlights the effectiveness of our approach, particularly given the inherent challenges in veterinary imaging, such as significant inter-breed variations in liver shape and size, and relatively limited sample sizes compared to human studies. Considering these challenges, achieving DSC and IoU values that closely match those of human studies further demonstrates that advanced segmentation techniques can be effectively adapted to veterinary imaging, despite these dataset limitations.

Veterinary imaging presents unique complexities distinct from those in human medicine, primarily due to breed-related anatomical variations and diverse clinical presentations. Similar challenges have been reported in canine kidney segmentation studies (12). Such factors complicate the segmentation task, particularly in livers with hepatic masses, which typically exhibit greater morphological variability than human livers, as demonstrated by the decreased segmentation performance and increased segmentation variability observed in Experiment 2. Furthermore, although the dataset in this study, consisting of 221 CT scans from 206 dogs, is large compared with several previous veterinary imaging studies, such as those involving 40 (14) and 76 dogs (13), it remains relatively modest compared with human imaging databases (29). Nonetheless, the current findings clearly demonstrate that deep learning-based segmentation models can effectively overcome these complexities, achieve competitive performance. This highlights their clinical applicability, in providing rapid and accurate liver volumetry for diagnosing conditions like hepatomegaly or microhepatia, aiding pre-operative planning for mass resection, and enabling objective monitoring of postoperative liver regeneration in veterinary medicine.

This study has some limitations. First, the dataset size was derived from a single institution and was small compared with studies in human medicine, with a limited representation of various breeds and clinical conditions. A larger and more diverse dataset could enhance the robustness and adaptability of the model. Although some studies in human medicine have demonstrated good segmentation performance using comparable or smaller datasets, the inherent variability in veterinary medicine owing to breed differences suggests that similar-sized datasets may not sufficiently represent this diversity (28, 30). The model's generalizability could be further enhanced by employing data augmentation to compensate for the limited dataset size. The evaluation in this study was performed at the image level rather than at the individual dog level, without ensuring the separation of dogs across the training, validation, and test sets, potentially influencing model generalizability. Additionally, although various dog sizes, ranging from small to large breeds, were included, the breed size distribution across subsets was not controlled, further limiting the representativeness of our dataset.

To address these limitations, several future directions are planned. Multi-institutional validation is essential to assess the model's robustness across diverse imaging conditions. This issue is particularly relevant in veterinary medicine, where there is no clear scientific consensus regarding optimal patient positioning, leading to substantial variations among institutions (31). Transfer learning could also be explored, using models pre-trained on large human liver datasets (32). Finally, future work could explore the use of other loss functions, such as focal loss, to potentially enhance segmentation accuracy, particularly for challenging cases involving hepatic masses.

5 Conclusion

This study presents the first deep learning-based liver segmentation model specifically developed for veterinary medicine to achieve reliable segmentation results. Including hepatic masses improved the generalizability but slightly increased the segmentation variability, whereas simpler anatomies without lesions resulted in a more consistent performance. Despite challenges such as breed-specific anatomical variations, limited sample sizes, and non-standardized imaging protocols, this approach demonstrates considerable potential for enhancing diagnostic precision, reducing clinician workload, and expediting decision-making in veterinary practice. Future research should prioritize expanding the dataset to include a broader variety of clinical cases, performing external validations, and exploring hybrid training strategies to further enhance the reliability and clinical applicability of the model.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SL: Writing – original draft, Conceptualization, Data curation, Funding acquisition, Investigation, Project administration, Software, Visualization, Methodology, Formal analysis. GS: Writing – review & editing, Conceptualization, Data curation, Funding acquisition, Methodology, Supervision, Investigation, Project administration. NY: Writing – review & editing. KN: Writing – review & editing. RT: Writing – review & editing, Methodology, Software, Supervision. TO: Writing – review & editing. MH: Writing – review & editing. MT: Writing – review & editing, Project administration, Supervision, Conceptualization.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the Japan Society for the Promotion of Science KAKENHI (grant number 23K1407403) and JST SPRING (grant number JPMJSP2119).

Acknowledgments

The authors thank Yoshihisa Yamane (Veterinary Teaching Hospital, Faculty of Veterinary Medicine, Hokkaido University,

Sapporo, Japan) for performing the manual segmentation of the CT dataset.

Conflict of interest

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.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. ChatGPT-4.5 was used to proof write the article.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
- Thakur GK, Thakur A, Kulkarni S, Khan N, Khan S. Deep learning approaches for medical image analysis and diagnosis. *Cureus.* (2024) 16:e59507. doi: 10.7759/cureus.59507
- Moghbel M, Mashohor S, Mahmud R, Saripan MI. Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography. *Artif Intell Rev.* (2018) 50:497–537. doi: 10.1007/s10462-017-9550-x
- Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. *Bioengineering.* (2023) 10:1435. doi: 10.3390/bioengineering10121435
- Gotra A, Sivakumaran L, Chartrand G, Vu KN, Vandenbroucke-Menu F, Kauffmann C, et al. Liver segmentation: indications, techniques and future directions. *Insights Imaging.* (2017) 8:377–92. doi: 10.1007/s13244-017-0558-1
- Kutaiba N, Chung W, Goodwin M, Testro A, Egan G, Lim R. The impact of hepatic and splenic volumetric assessment in imaging for chronic liver disease: a narrative review. *Insights Imaging.* (2024) 15:146. doi: 10.1186/s13244-024-01727-3
- Buijk MS, Dijkshoorn M, Dwarkasing RS, Chorley AC, Minnee RC, Boehnert MU. Accuracy of preoperative liver volumetry in living donor liver transplantation—a systematic review and meta-analysis. *J Liver Transpl.* (2023) 10:100150. doi: 10.1016/j.liver.2023.100150
- Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans Med Imaging.* (2018) 37:2663–74. doi: 10.1109/TMI.2018.2845918
- Gul S, Khan MS, Bibi A, Khandakar A, Ayari MA, Chowdhury MEH. Deep learning techniques for liver and liver tumor segmentation: a review. *Comput Biol Med.* (2022) 147:105620. doi: 10.1016/j.compbiomed.2022.105620
- Appleby RB, Basran PS. Artificial intelligence in diagnostic imaging. *Adv Small Anim Care.* (2024) 5:67–77. doi: 10.1016/j.yasa.2024.06.005
- Xiao S, Dhand NK, Wang Z, Hu K, Thomson PC, House JK, et al. Review of applications of deep learning in veterinary diagnostics and animal health. *Front Vet Sci.* (2025) 12:1511522. doi: 10.3389/fvets.2025.1511522
- Ji Y, Cho H, Seon S, Lee K, Yoon H. A deep learning model for CT-based kidney volume determination in dogs and normal reference definition. *Front Vet Sci.* (2022) 9:1011804. doi: 10.3389/fvets.2022.1011804
- Ji Y, Hwang G, Lee SJ, Lee K, Yoon H. A deep learning model for automated kidney calculi detection on non-contrast computed tomography scans in dogs. *Front Vet Sci.* (2023) 10:1236579. doi: 10.3389/fvets.2023.1236579
- Schmid D, Scholz VB, Kircher KR, Lautenschlaeger IE. Employing deep convolutional neural networks for segmenting the medial retropharyngeal lymph nodes in CT studies of dogs. *Vet Radiol Ultrasound.* (2022) 63:763–70. doi: 10.1111/vru.13132
- Park SH, Cho H, Lee K, Yoon H. Deep learning model for CT-based adrenal gland volume determination and normal reference definition in dogs. *Pak Vet J.* (2025) 45:320–7. doi: 10.29261/pakvetj/2025.138
- Wrigley RH. Radiographic and ultrasonographic diagnosis of liver diseases in dogs and cats. *Vet Clin North Am Small Anim Pract.* (1985) 15:21–38. doi: 10.1016/S0195-5616(85)50002-9
- Suter PF. Radiographic diagnosis of liver disease in dogs and cats. *Vet Clin North Am Small Anim Pract.* (1982) 12:153–73. doi: 10.1016/S0195-5616(82)50026-5
- Kummeling A, Vrakking DJE, Rothuizen J, Gerritsen KM, van Sluijs FJ. Hepatic volume measurements in dogs with extrahepatic congenital portosystemic shunts before and after surgical attenuation. *J Vet Intern Med.* (2010) 24:114–9. doi: 10.1111/j.1939-1676.2009.0439.x
- Stieger SM, Zwingerberger A, Pollard RE, Kyles AE, Wisner ER. Hepatic volume estimation using quantitative computed tomography

- in dogs with portosystemic shunts. *Vet Radiol Ultrasound*. (2007) 48:409–13. doi: 10.1111/j.1740-8261.2007.00268.x
20. Zwingenberger AL, Daniel L, Steffey MA, Mayhew PD, Mayhew KN, Culp WTN, et al. Correlation between liver volume, portal vascular anatomy, and hepatic perfusion in dogs with congenital portosystemic shunt before and after placement of ameroid constrictors. *Vet Surg*. (2014) 43:926–34. doi: 10.1111/j.1532-950X.2014.12193.x
21. Israeliantz N, Lodzinska J, Woods G, Pontes J, Parys M, Schwarz T, et al. simplified CT-volumetry method for the canine liver. *Vet Radiol Ultrasound*. (2022) 63:47–53. doi: 10.1111/vru.13018
22. Kinoshita K, Kurihara H, Moore GE, Murakami M. Pilot study: the effects of slice parameters and the interobserver measurement variability in computed tomographic hepatic volumetry in dogs without hepatic disease. *Vet Sci*. (2023) 10:177. doi: 10.3390/vetsci10030177
23. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Lect Notes Comput Sci*. (2015) 9351:234–41. doi: 10.1007/978-3-319-24574-4_28
24. Fernando T, Gammulle H, Denman S, Sridharan S, Fookes C. Deep learning for medical anomaly detection – a survey. *ACM Comput Surv*. (2022) 54:141. doi: 10.1145/3464423
25. Wasserthal J, Breit HC, Meyer MT, Pradella M, Hinck D, Sauter AW, et al. TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. *Radiol Artif Intell*. (2023) 5:e230024. doi: 10.1148/ryai.230024
26. Nai YH, Teo BW, Tan NL, O'Doherty S, Stephenson MC, Thian YL, et al. Comparison of metrics for the evaluation of medical segmentations using prostate MRI dataset. *Comput Biol Med*. (2021) 134:104497. doi: 10.1016/j.compbiomed.2021.104497
27. Müller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res Notes*. (2022) 15:210. doi: 10.1186/s13104-022-06096-y
28. Bilic P, Christ P, Li HB, Vorontsov E, Ben-Cohen A, Kaissis G, et al. The liver tumor segmentation benchmark (LiTS). *Med Image Anal*. (2023) 84:102680. doi: 10.1016/j.media.2022.102680
29. Zhang Z, Li G, Wang Z, Xia F, Zhao N, Nie H, et al. Deep-learning segmentation to select liver parenchyma for categorizing hepatic steatosis on multinational chest CT. *Sci Rep*. (2024) 14:11987. doi: 10.1038/s41598-024-62887-2
30. Kavur AE, Gezer NS, Bariş M, Aslan S, Conze PH, Groza V, et al. CHAOS Challenge – combined (CT-MR) healthy abdominal organ segmentation. *Med Image Anal*. (2021) 69:101950. doi: 10.1016/j.media.2020.101950
31. Oliveira CR, Henzler MA, Johnson RA, Drees R. Assessment of respiration-induced displacement of canine abdominal organs in dorsal and ventral recumbency using multislice computed tomography. *Vet Radiol Ultrasound*. (2015) 56:133–43. doi: 10.1111/vru.12198
32. Ghobadi V, Ismail LI, Wan Hasan WZ, Ahmad H, Ramli HR, Norsahperi NMH, et al. Challenges and solutions of deep learning-based automated liver segmentation: a systematic review. *Comput Biol Med*. (2025) 185:109459. doi: 10.1016/j.compbiomed.2024.109459