Among the most important diagnostic tools in small animal medicine is radiography, which is a non-invasive and cost-effective method for diagnosing and monitoring diseases in animals. Over the years, the advancement of digital imaging has allowed for ease in image acquisition, thereby making radiography a routine diagnostic in many practices. This has led to increased demand for radiology services, leading to some challenges surrounding technical and human errors. Technical errors, including motion artifacts, inadequate exposure, improper patient positioning, processing errors, and inappropriate use of collimation and grids, negatively affect the diagnostic quality of radiographs. In both human and veterinary medicine, thoracic radiographs are routinely taken for cardiopulmonary evaluation and metastasis screening.

While various errors can occur during radiograph acquisition, one commonly identified error is the inappropriate collimation of images. That is, the field of view for the study is centered to collimate either the cranial or caudal thorax out of view. This type of error is typically a human error associated with the positioning of the patient on the X-ray table. A light guide is used to center the patient over the radiograph plate, the patient is still frequently positioned incorrectly concerning the X-ray beam in the authors’ experience.

OBJECTIVES
To determine the feasibility of machine learning algorithms for the classification of appropriate collimation of the cranial and caudal borders in ventrodorsal and dorsoventral thoracic radiographs.

SAMPLES
900 ventrodorsal and dorsoventral canine and feline thoracic radiographs were retrospectively acquired from the Picture Archiving and Communication system (PACS) system of the Ontario Veterinary College.

PROCEDURES
Radiographs acquired from April 2020 to May 2021 were labeled by 1 radiologist in Summer of 2022 as either appropriately or inappropriately collimated for the cranial and caudal borders. A machine learning model was trained to identify the appropriate inclusion of the entire lung field at both the cranial and caudal borders. Both individual models and a combined overall inclusion model were assessed based on the combined results of both the cranial and caudal border assessments.

RESULTS
The combined overall inclusion model showed a precision of 91.21% (95% CI [91, 91.4]), accuracy of 83.17% (95% CI [83, 83.4]), and F1 score of 87% (95% CI [86.8, 87.2]) for classification when compared with the radiologist’s quality assessment. The model took on average 6 ± 1 second to run.

CLINICAL RELEVANCE
Deep learning-based methods can classify small animal thoracic radiographs as appropriately or inappropriately collimated. These methods could be deployed in a clinical setting to improve the diagnostic quality of thoracic radiographs in small animal practice.
Recently, artificial intelligence has been used to identify pathology (computer-aided diagnostics [CAD]) in the thorax and musculoskeletal systems of companion animals. At the time of writing, no publications are evaluating the role of AI in the quality control of radiographs in companion animal practice. However, in human radiography, AI has been used to make meaningful improvements in the quality of diagnostic images. For instance, Nousiainen et al. defined a set of thresholds on the distance of human lungs to the 4 sides of the radiograph border and specified the minimum number of ribs for proper inspiration to analyze radiographs. Most recently, Meng et al. used 10 quality assessment criteria to assess the quality of human radiograph layout and positioning. This study segmented and localized landmarks on each radiograph using U-Net and then used these results to measure different distances. Finally, they input these values into a multiple linear regression model to output the quality on a scale of 10.

Inappropriate collimation can negatively impact patient care. By not including portions of the thorax one may miss important findings, such as pulmonary infiltrates, nodules, or masses. While manual inspection of the acquired images by experienced personnel can avoid this mistake, a lack of appropriate training and time constraints for veterinary personnel may limit this in practice. Therefore, there is a potential role for artificial intelligence to be applied at the point of image acquisition to confirm appropriate collimation of patients before they leave the radiology suite. Such a system would improve the diagnostic quality of images and improve the confidence of the interpreting veterinarian or veterinary radiologist. Such a system also has the potential to reduce the need for repeat images, which may require additional cost or additional use of sedation if the patient has already left the radiology suite. In this work, we propose a fully automated method based on recent advancements in CNNs and machine learning to classify a given thoracic radiograph based on the inclusion of the cranial and caudal borders.

**Materials and Methods**

To classify cranial and caudal borders, we first utilized a deep-learning model to segment the spine, ribs, and abdomen. Then, we use these segmented regions in a specialized machine-learning model to classify cranial and caudal borders.

**Dataset**

A total of 900 ventrodorsal (VD) and dorsoventral (DV) radiographs of companion animals (canine and feline) were retrospectively extracted from the Picture Archiving and Communications System (PACS) at the Ontario Veterinary College (OVC) at the University of Guelph. The radiographs selected were provided by referring clinics from various computed radiography and direct digital radiography systems as part of the patient history during referral. Cases were selected by searching the PACs for studies classified as “imported” up to May 2021. The most recent studies which included thoracic images were selected until a dataset of 900 images was achieved. Images were not excluded for any abnormalities in positioning, exposure, and collimation. Images were either in Joint Photographic and Experts Group (JPEG) or Digital Imaging and Communications in Medicine (DICOM) format. All radiographs were labeled by a veterinary radiologist with more than 3 years of experience. The cranial and caudal borders were assessed individually in each image. Radiographs were deemed to be inappropriately collimated if the cranial lung lobes were collimated out of view or the caudal lung lobes were incompletely imaged as judged by the labeling radiologist (Figure 1). Images that included at least the entire C7 vertebra were deemed appropriate cranially. The caudal borders were deemed appropriate if the images included at least 2 vertebral lengths beyond the point of the angle formed at the junction of the diaphragm and thoracic wall. Instances, where too much anatomy was included in the image (eg, including too much of the neck cranially or abdomen caudally), were not excluded. Following this labeling process, the dataset included 110 radiographs where the caudal thorax was inappropriately collimated, 130 images in which the cranial thorax was inappropriately collimated, and 660 images deemed normal for collimation.

The overall dataset handling pipeline for training and evaluating the deep learning and the machine learning model is provided (Figure 2). Due to the

**Figure 1**—Representative ventrodorsal canine thoracic radiographs from among 900 ventrodorsal or dorsoventral canine and feline thoracic radiographs acquired between April 2020 and May 2021 and retrospectively evaluated in Summer of 2022 to determine the feasibility of machine learning algorithms to identify the appropriate inclusion of the entire lung field vs collimation that excluded the cranial (A) and caudal (B) aspects of the lungs from view.
lack of standard test data, stratified 5-fold cross-validation was used to evaluate our deep learning model to minimize the impact of selecting training and testing datasets for the evaluation process. We used 80% of the whole data for training and 20% of the whole data for evaluation on each fold. On each fold, we had 720 radiographs for training which included 88 radiographs with missing caudal borders and 104 radiographs with missing cranial borders. We evaluated the result on each fold using 180 radiographs, which included 22 radiographs with missing caudal borders and 26 radiographs with missing cranial borders. All of the reported results in this research were based on averaging the results across 5-folds that guaranteed each radiograph appeared exactly once in the test sets across all folds.

**Image processing**

All the radiographic images were exported in 8-bit JPEG format and manually anonymized. Before passing radiographs to the deep learning model, we resized them to the size 1,024 X 1,024 and performed adaptive histogram equalization\(^\text{16}\) to improve contrast. During the training of the deep learning model, we randomly flipped radiographs on the horizontal axis to increase generalization.

**Deep learning**

We implemented our deep learning models in Python and utilized Pytroch\(^\text{17}\) and Scikit-learn\(^\text{18}\) libraries. All the models were trained on a 24GB GPU (NVIDIA RTX3090) and a CPU @ 3.4 GHz X 16 cores (AMD Ryzen 9). We used UNETR\(^\text{19}\) for the segmentation model, which is a combination of CNNs and Visual Transformers (ViT).\(^\text{20}\) We chose this model over purely CNN segmentation models because these models generally exhibit limitations for modeling long-range dependencies. The training learning rate was automatically adjusted using an adaptive learning rate function called Adam.\(^\text{21}\) We used training data on each fold to train the UNETR model and to avoid overfitting, we used early stopping. After finishing the training on each fold, we evaluated the model using the test data. We used a combination of focal loss\(^\text{22}\) and dice coefficient as loss functions to train the model.

Utilizing the defined training scheme, we trained the UNETR model to segment ribs, spine, and abdomen. We then found the thorax region from the segmented ribs using the method described in our previous lab research.\(^\text{23}\) The segmented thorax, spine, and abdomen were used to classify inclusion on cranial and caudal borders as was described in the next section.

**Classification**

**Cranial edges**—To classify cranial boundaries, we utilized the ribs and thorax segmentations to extract the cervical region of the spine (Figure 3). To extract the cervical region of the spine the segmented thorax was subtracted from the spine, then features were extracted from this region including the area (in pixels) of the region and the ratio of the height of the cervical region to the total height of the spine. Then we passed these features to a Multi-Layer Perceptron (MLP) model for classification.

**Caudal edges**—To classify the inclusion of the caudal borders the abdomen was segmented using the segmentation method explained in section 2.3 (Figure 4). As the caudodorsal aspect of the lungs extends dorsal to the cranial abdomen, an appropriately collimated image must extend beyond the diaphragmatic margin. To ensure the inclusion of the entire lung field, the thorax cranial to the diaphragm was segmented by subtracting the included portion of the abdomen. The ribs were subtracted from the mask leaving only the inner margins of the thoracic cavity. Following this, a bounding box was fitted to the lungs cranial to the diaphragm. The bounding box allowed identification of the caudal aspects of the thorax where the diaphragm and thoracic wall meet.
The residual abdominal region was then isolated by subtracting the bounding box. This feature was used in an MLP model.

**Overall inclusion**—To combine the cranial inclusion and caudal inclusion models together to classify a given thoracic radiograph based on inclusion on both sides, we used the labels from both models. If both models predict that a given radiograph should be accepted for inclusion, then the radiograph is accepted. If either of the models predicts that the radiograph should be rejected based on inappropriate collimation, the radiograph is rejected.

**Statistical analysis**

We evaluated all of the developed models using precision, recall, F1, and Area Under Curve Receiver Operator Characteristic (AUC-ROC) curve. Precision, recall, and F1 can be derived from a confusion matrix for a binary classification task. The precision, recall, and F1 metrics were defined as:

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

\[
F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

Where a True Positive (TP) in our work indicates the number of wrongly collimated radiographs classified as wrongly collimated by our model. Similarly, True Negative (TN) indicates the number of correctly collimated radiographs predicted as correctly collimated by the model, False Positive (FP) indicates the number of wrongly collimated radiographs predicted as correctly collimated and False Negative (FN) indicates the number of correctly collimated radiographs predicted as wrongly collimated by the model.

A ROC curve is a graph showing the performance of a classification model at all classification thresholds. This curve plots 2 parameters, True Positive Rate (TPR) or recall and False Positive Rate (FPR) which are defined as:

\[
\text{FPR} = \frac{FP}{FP + TN}
\]

**Figure 3**—Cranial border assessment segmentation process for the dataset of 900 ventrodorsal or dorsoventral canine and feline thoracic radiographs described in Figure 1. An input radiograph (A, D) is segmented (B, E) to identify the spine (red) and thorax (green). Vertebrae cranial to the cranial most edge of the thoracic segmentation are considered the cervical vertebrae and are isolated from the image (C, F). In this example, the top row includes the C4-7 vertebrae and was labeled as accepted while the bottom row image includes only the caudal portion of the C7 vertebra and was labeled as rejected. In the bottom row, the cranial most tips of the lungs are collimated out of view.
A ROC curve plots TPR vs FPR at different classification thresholds. This plot can be utilized to tune the classification threshold. The AUC measures the entire 2D area underneath the entire ROC curve and provides an aggregate measure of performance across all possible classification thresholds. Its value can be used to compare the classification capability of different classifiers and the higher the value, the better the model is at making classification decisions.

**Results**

The results of the cranial inclusion, caudal inclusion, and overall inclusion are provided (Table 1) along with 95% CI values. The model took on average 6 ± 1 seconds to make the overall inclusion prediction. The cranial inclusion model had higher recall and F1 in comparison with caudal inclusion, while caudal inclusion was more precise in prediction than cranial inclusion. Combining these 2 models to predict the overall inclusion resulted in a model which its precision, recall, and F1 lie between the cranial inclusion and caudal inclusion model values. However, its AUC was lower than the 2 individual models.

Confusion matrices and ROC curves for 3 models are reported (Figure 5). The numbers in confusion matrices are the summation of predictions across all 5 folds, an ROC curve for each fold is provided. In general, our model had 40 FP and 16 FN. Additionally, the cranial inclusion model had the least FP (16 images), while the caudal inclusion model had the least FN (6 images).

![Figure 4](image)

**Table 1**—Precision, Recall, F1, and Area Under Curve (AUC) values for cranial inclusion, caudal inclusion, and overall inclusion machine learning models developed from a dataset of 900 canine and feline dorsoventral and ventrodorsal thoracic radiographs at 95% CIs.

<table>
<thead>
<tr>
<th>Metric model</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>F1 (%)</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cranial edges</td>
<td>90.65 (95% CI [89.7, 91.6])</td>
<td>87.92 (95% CI [87.2, 88.6])</td>
<td>89.06 (95% CI [88.5, 89.6])</td>
<td>98.95 (95% CI [98.9, 0.99])</td>
</tr>
<tr>
<td>Caudal edges</td>
<td>92.23 (95% CI [91.3, 93.2])</td>
<td>81.53 (95% CI [80.1, 83])</td>
<td>86.25 (95% CI [85.2, 87.3])</td>
<td>97.84 (95% CI [97.5, 98.1])</td>
</tr>
<tr>
<td>Overall inclusion</td>
<td>91.21 (95% CI [91.00, 91.4])</td>
<td>83.17 (95% CI [83.00, 83.4])</td>
<td>87.00 (95% CI [86.8, 87.2])</td>
<td>90.3 (95% CI [89.9, 90.7])</td>
</tr>
</tbody>
</table>
Discussion

This work demonstrates the role of artificial intelligence in improving companion animal thoracic radiograph quality. The proposed model used recent advancements in deep learning to segment the ribs, spine, and abdomen from a given radiograph. Then by isolating the cervical spine and abdomen, the cranial and caudal borders can be assessed for appropriate inclusion of the entire pulmonary parenchyma. The cranial and caudal borders were each analyzed separately. The overall inclusion model used the prediction output from both models and classified each given radiograph as “rejected” or “accepted.” Our model performed well and could be used as the basis for a clinically applicable technology to assist in veterinary radiograph acquisition.

One of the strengths of this study is the diversity of the dataset. That is, the referral images used are from different clinics and were acquired by different radiograph devices with different resolutions and overall different image qualities. The high classification accuracy across 5-folds (Table 1) suggests that the developed classification method could become a reliable, ready-to-use tool in clinical practice. Nonetheless, as stated by Zach et al.,

\[\text{Equation} \]

the performance of CNNs may vary significantly from one site to another, primarily related to the imaging device parameters and acquisition technique. Henceforth, the on-site accuracy should be tested before the introduction of this classification method for clinical practices. Additionally, despite the diversity of data, the dataset remains relatively small. The 5-fold cross-validation was used to minimize the effects of this small sample size; however, the reliability of the reported metrics could be improved by increasing the dataset size.

In deep learning, transfer learning is a method in which a pre-trained model is used for a specific task as a starting point for another task. The main benefit of using transfer learning is that the model can learn faster and achieve better results.\[25,26\] It becomes increasingly important to use pre-trained models when we are working with a small dataset, such as ours consisting of 900 radiographs. In these scenarios, transfer learning will help the model generalize better and achieve higher results.\[24\] Here, we utilized ImageNet\[27\] dataset for the pre-trained weights. Because lateral radiographs also include the spine, ribs, and abdomen, the ideas discussed in this research could be applied, after some modifications, to left or right lateral radiographs. While similar segmentation and isolation of the thorax and spine can be performed, it is imperative that the hyperparameters of the machine learning models need to be tuned and studied separately to ensure the best performance for lateral images.

In this study, we did not exclude poorly exposed radiographs from our dataset. As our lab has described previously,\[23\] the thorax segmentation method is robust to over-exposure and under-exposure. This ensures that the thorax and spine can be effectively segmented across varying image quality and will allow future models to provide clinical end users with collimation feedback even if exposure is poor. Before implementing the proposed method...
in clinical practice, future research should examine the effects of underexposure and overexposure for the segmentation of the abdomen.

Considering the prediction time for our proposed model, thorax segmentation took the highest amount of computational time with 6 ± 1 seconds. However, because we used thorax segmentation in both the cranial inclusion and caudal inclusion model, we only needed to run the thorax segmentation once. As for the rest of the prediction procedures such as segmentation prediction, feature extraction, and machine learning model prediction, the process took less than 1 second and is negligible.

Future research should focus on incorporating the results of this study into the development of a comprehensive thoracic radiograph quality control model that includes 1 of the components for classifying collimation quality. This comprehensive model could be studied clinically to determine if it supports improved diagnostic quality of radiographs. This comprehensive thoracic radiograph quality control model that includes 1 of the components for classifying cranial and caudal borders in thoracic canine and feline radiographs. The results of this study suggest that the proposed method based on deep learning for automatic classification of cranial and caudal inclusion of thoracic radiographs could be applicable in practice.

Acknowledgments
Funding for this study was provided by the CARE-AI seed fund and NSERC Alliance. The CLAIM guidelines were used in the preparation of this manuscript. The authors have no conflict of interest to declare.

References
