Ensemble Monte Carlo dropout based uncertainty quantification in automated classification of spinal bone metastasis using abdominal CT scans

Soo Ho Ahn

The Graduate School
Yonsei University
Department of Integrative Medicine

Ensemble Monte Carlo dropout based uncertainty quantification in automated classification of spinal bone metastasis using abdominal CT scans

A Master's Thesis Submitted
to the Department of Integrative medicine
and the Graduate School of Yonsei University
in partial fulfillment of the
requirements for the degree of
Master's of Engineering

Soo Ho Ahn

June 2024

This certifies that the Master's Thesis
of Soo Ho Ahn is approved.

Thesis Supervisor Young Han Lee

Thesis Committee Member Hwiyoung Kim

Thesis Committee Member Hyungjin Rhee

The Graduate School
Yonsei University
June 2024

Acknowledgement

Completing this master's thesis would not have been possible without the support and encouragement of many individuals. I want to express my sincere gratitude to all those who have assisted me throughout my master's studies and the writing of this thesis.

First and foremost, I would like to thank my supervisors, Prof. Young Han Lee and Prof. Hwiyoung Kim. Their expertise, guidance, and unwavering support have been invaluable to my research and the completion of this thesis. I am also deeply grateful to my committee member, Prof. Hyungjin Rhee, for his valuable insights and suggestions.

I extend my heartfelt thanks to the members of the Translational AI Lab for their collaboration and camaraderie. Working with such a dedicated and supportive team has made this journey both meaningful and rewarding.

On a personal note, I would like to thank Hyoeun Kim for her unconditional love, patience, and support.

Finally, I would like to express my deepest gratitude to my parents for their unwavering support and belief in me. Your sacrifices, love, and encouragement have been the foundation upon which I have built my academic and personal life. Thank you for always being there for me.

This thesis results from the collective efforts and support of all these wonderful people. Thank you all for helping me reach this milestone.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Ensemble Monte Carlo dropout based uncertainty quantification in automated classification of spinal bone metastasis using abdominal CT scans

Purpose: To enhance the automatic detection and classification of spinal bone metastases from abdominal computed tomography (CT) scans, this study aimed to address the challenges in diagnostic sensitivity and efficiency by integrating uncertainty quantification.

Methods: This retrospective study analyzed 11,468 abdominal CT images from 116 patients diagnosed with spinal bone metastases and included data from 11 healthy normal-control participants, contributing 957 images to the dataset. The images were annotated and classified into "normal," "disc," and "metastasis." We introduced a novel and efficient technique for uncertainty quantification called ensemble Monte Carlo dropout (EMCD). This technique leverages the DenseNet201 architecture with added dropout layers for uncertainty management and employs YOLOv5m for precise spine region detection, complemented by a weighted voting ensemble for classification. The uncertainty quantification was articulated through numerical values, predictive probability intervals, and Uncertainty-CAM visualizations. Our performance evaluations focused on assessing spine detection efficiency, metastasis classification accuracy, and the robustness of the model against both healthy controls and out-of-distribution data.

Results: The YOLOv5m model achieved a high mean average precision of 0.995 in spine

detection. The EMCD model showed superiority in multi-class classification with an area under the receiver-operating-characteristic curve (AUC) of 0.93, outperforming traditional and other uncertainty quantification models. At 50% data retention, the EMCD model reached an AUC of 0.96 and an accuracy of 96%. Moreover, it maintained a high accuracy of 90% on a normal-control dataset. Additionally, the model demonstrated excellent calibration with an Expected Calibration Error (ECE) of 0.09.

Conclusion: The EMCD model significantly advances the automated detection of spinal bone metastases, offering superior diagnostic accuracy and a novel approach for uncertainty quantification. This contributes to more informed clinical decision-making and highlights the potential of integrating advanced artificial intelligence methodologies to improve patient care.

Key words : **Spine; Computed Tomography; Metastasis; Deep learning; Uncertainty**

1. Introduction

The bone is the third most common site of metastasis after the lungs and liver, and it is associated with an advanced stage and poor prognosis[1-3]. Approximately two-thirds of cancer patients develop bone metastasis[4]. Almost all patients with cancer have metastasis to some part of the body[5]. Skeletal metastasis is clinically significant because of its associated symptoms and complications, including back pain, pathologic fractures, muscle weakness, and bowel and bladder incontinence[6]. Thus, accurate and early detection of bone metastases is important for proper treatment planning. Computed tomography (CT) is widely accessible and cost-effective. Therefore, in clinical practice. Therefore, in the clinical setting, CT is the predominant imaging modality employed for both initial cancer staging and subsequent follow-up evaluations[7-9]. However, detection of bone metastasis using CT is not sensitive[10]. This is partly due to the fact that early bone metastasis is only seen as subtle changes on CT[10], but also due to radiologist burnout. Thousands of chest or abdominal CT scans are acquired during the clinical follow-up management of patients with malignancies, such as lung, prostate, or breast cancer. Because traditional whole vertebral screening is time-consuming, efforts have been made to enhance bone metastasis detection using various algorithms to remove bony structures[11-13].

Recent advancements in artificial intelligence (AI), particularly in machine learning and deep learning (DL), have shown promise for improving the detection of bone metastases in spinal CT scans, potentially serving as a valuable tool in aiding early diagnosis and informing treatment decisions[14-17]. Despite their high predictive accuracy, DL models face criticism for their "black box" nature, which obfuscates their decision-making processes and could lead to challenges in clinical adoption[18]. The classification of vertebral bone metastases poses significant challenges owing to the complex structure of the spine and diversity of lesions. Recent studies have demonstrated the potential of DL to address these challenges. Koike et al. developed an AI-based computer-aided detection system utilizing DL for the classification of lytic vertebral bone metastases from 79 CT scans, achieving an accuracy of 0.872 and area under the receiver operating characteristic curve (AUC) of 0.941[16]. Noguchi et al. introduced a DL algorithm aimed at assisting radiologists in

detecting bone metastases on CT images by analyzing more than 269 CT scans, with improvements in the free-response receiver operating characteristic scores from 0.746 to 0.899[17].

The field of DL has explored various methodologies for uncertainty quantification (UQ), among which the Bayesian approach and deep ensembles (DE) are prominently utilized[19]. The Bayesian approach, as exemplified by Gal et al., introduces stochastic variations in the model's weights to assess prediction uncertainty, with Monte Carlo dropout (MCDO) being a notable technique[20]. However, MCDO can be challenging to integrate with all DL architectures, often underperforms compared to conventional convolutional neural networks, and may lack robustness against noisy data. DE proposed by Lakshminarayanan et al., employ multiple DL models for prediction and offer a spectrum of possible outcomes[21]. While generally performing well, they can suffer from shared biases among ensemble models if they are not sufficiently diverse. The ensemble Monte Carlo dropout (EMCD) technique, introduced in this study, is designed to overcome these limitations by integrating MCDO's stochastic evaluation with the diversified predictive capability of DE. The EMCD aims to enhance the reliability and interpretability of uncertainty quantification in DL models, providing clinicians with a more nuanced understanding of the model's confidence in its predictions, especially for complex diagnostic tasks such as vertebral bone metastasis classification.

This study introduces a two-step DL methodology aimed at the automated detection and classification of vertebral regions from abdominal CT images. Our approach involves the initial detection and extraction of vertebral areas, followed by multiclass classification to determine the status of the vertebrae as normal, disc, or metastasis alterations indicative of metastases. A key objective was to quantify the uncertainty of classification outcomes, thereby enhancing the trustworthiness and interpretability of the automated system for clinical practitioners. By providing uncertainty measurements in different forms, such as a single numerical value, probability interval, or visual map (Uncertainty-CAM[22]), we aimed to empower clinicians with a clearer understanding of the model's confidence in its predictions, facilitating better-informed clinical decisions, and potentially improving patient outcomes.

2. Materials and methods

2.1. Data description and preparation

The patient cohort for this study was identified using a hospital information system integrated with a picture archiving and communication system and an electronic medical record targeting individuals who had undergone abdominopelvic CT scans from January to June 2017. The inclusion criteria were primarily based on a pathologically or clinically confirmed diagnosis of bone metastasis.

Table 1. Demographic information of the patient datasets

	Training/validation	**Test**
Per patient		
Patients/images	104/10,356	12/1,112
Age, mean (SD)	59.72 (12.48)	59.33 (12.18)
Age, range	24-86	39-85
Male, total (%)	37.50	41.67
Per image		

	Normal	Disc	Mets	Normal	Disc	Mets
Images, n (%)	7,826	1,907	623	728	254	130
	(75.56)	(18.41)	(6.02)	(65.47)	(22.84)	(11.69)

Mets: metastasis.

A total of 116 patients with metastatic manifestations within the thoracic or lumbar spine were included. CT scans of these patients were reconstructed axially to form a comprehensive CT series database, which yielded 11,468 slices.

Patient histories included lung cancer (adenocarcinoma) (21), breast cancer (26), rectal cancer (17), colon cancer (22), stomach cancer (8), tongue cancer (3), malignant gastrointestinal stromal tumor (2), pancreatic cancer (2), appendiceal mucinous cystadenocarcinoma with pseudomyxoma peritonei (1), thymic carcinoma (1), adrenal cortical carcinoma (1), renal cell carcinoma (1), Klatskin's tumor (1), neuroendocrine carcinoma of tail of pancreas (1), ampulla of Vater cancer (1), anaplastic hemangiopericytoma, malignancy (1), esophagus cancer (1), anal cancer (1), nasopharyngeal cancer (1), malignant melanoma (1), hepatocellular carcinoma (1), undifferentiated sarcoma (1), and primary unknown (1). This retrospective study was approved by our institutional review board.

In terms of data preparation, out of the total 11,468 slices, 8,554 were labeled as "normal," 2,161 as "disc," and 753 as "mets." These labels were determined using radiologists' annotations based on the scan content. Subsequently, to train the spine region detection model, a radiologist manually annotated the bounding box around each sliced vertebra. To standardize the image depth, we down-sampled the images from 16-bit to 8-bit and resized them to 512 × 512 pixels to align them with the input requirements of the YOLOv5m model. After spinal region detection, the region of interests (ROIs) was resized to 150 × 150 pixels to comply with the input specifications of the DenseNet201 model. We also performed image normalization to ensure a consistent range of pixel intensities

across the dataset. The datasets were randomly divided in a 7:2:1 ratio for training, validation, and testing. Table 1 presents the demographic information of the patient datasets.

Table 2. Demographic information of the healthy normal-control dataset

	Healthy normal-control test		
Per patient			
Patients/images	11/957		
Age, mean (SD)	52.73 (17.05)		
Age, range	24-66		
Male, total (%)	45.45		
Per image			
	Normal	Disc	Mets
Images, n (%)	798 (83.39)	159 (16.61)	0 (0)

The healthy normal-control dataset comprised 957 slices from 11 individuals falling within the age range of 24–66 years. These control images were processed in the same manner as the patient images to maintain consistency across datasets. This dataset comprised data collected from January to February 2023. Table 2 presents the demographic information of the healthy normal-control dataset.

2.2. Proposed model framework

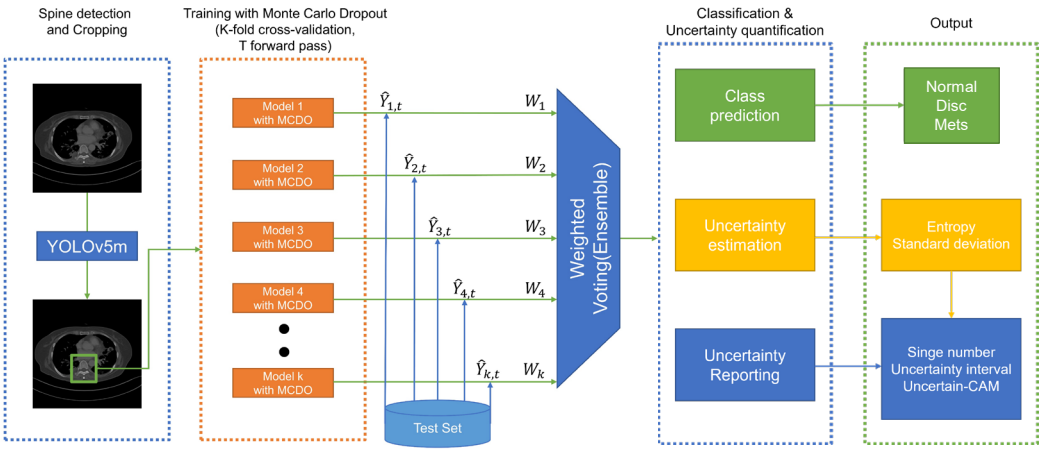


Figure 1. Overview of the proposed method.

The proposed bone metastasis prediction model comprised three main steps. First, we detected the spinal region on an abdominal CT scan and cropped the ROI. Second, we used the proposed EMCD method to classify the cropped CT image as normal, disc, or metastasis, and calculate the uncertainty based on the model's prediction probability. Finally, the calculated uncertainty was reported to medical experts in various ways.

a. Spine detection and cropping

In the first step of the proposed framework, we used the YOLOv5m[23,24] model pre-trained on ImageNet[25]. The YOLOv5m model is a state-of-the-art DL architecture for detecting objects in images. When applied to abdominal CT images, the model automatically identified spinal regions and isolated important ROIs for further analysis. The YOLOv5m model uses a CT image of size 512 × 512 as input and provides the width, height, and center position of the predicted bounding box with confidence.

b. Multi-class classification and uncertainty quantification

(1) EMCD

Next, we proposed an EMCD model to classify the extracted ROI into three classes. As the underlying model for the classification task, we used the DenseNet201[26] architecture, which is often used for image classification tasks. To quantify the MCDO uncertainty, we added a dropout[27] layer to the DenseNet201 network just before the final classification layer. We then developed K unique MCDO models to ensure the reliability and generalization of the model by utilizing a stratified K-fold cross-validation, where each MCDO model performed T probabilistic forward passes, resulting in predictive distributions. This method introduced variations into the activation pathways of the neural network, thereby generating different output spectra for a given input. The diversity of outputs over multiple passes reflects the prediction uncertainty of the model. A wider range in the output distribution indicates higher uncertainty, whereas a narrower range indicates higher confidence.

(2) Weighted voting ensemble

The core of the EMCD model is a weighted voting ensemble[28] technique that uses uncertainty estimation. The weighted voting ensemble technique assumes that some models in the ensemble perform better than others and gives them more weight in their predictions. Weighted voting ensembles are an evolution of regular voting ensembles, which assume that all models are equally capable and contribute proportionally to the ensemble prediction. Each model was assigned a specific weight multiplied by its predictions, and these weights were used to calculate the sum or average of the predictions. The difficulty associated with using such ensembles is finding an ensemble with equal model weights and a model weight that outperforms all the contributing models. The weighted voting ensemble using the uncertainty estimation presented in this study first trains the MCDO models using a stratified K-fold cross-validation dataset. Once trained, the MCDO models performed T iterations of predictions on the test set to infer $K * T$ prediction probabilities. From the K generated prediction probability distributions, the standard deviation or uncertainty was estimated and used as a weight to emphasize the reliability of the predictions of each model.

Finally, we computed the mean prediction probability $\mu_{pred,i}$ for each model i, which was the average of the predictions across all T stochastic forward passes: Modell's

$$\mu_{\text{pred},i} = \frac{1}{T}\sum_{t=1}^{T}\hat{Y}_{i,t} \tag{1}$$

where $\hat{Y}_{i,t}$ denotes the prediction probability of model i at forward pass t. Subsequently, we calculated the standard deviation, σ_i for the prediction probabilities of model i, to quantify the model's uncertainty:

$$\sigma_i = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(\hat{Y}_{i,t} - \mu_{\text{pred},i}\right)^2} \tag{2}$$

The voting-weight w_i for model i is then determined by the reciprocal of its standard deviation, to inversely correlate with the uncertainty:

$$w_i = \frac{1}{\sigma_i + \epsilon} \tag{3}$$

Here, ϵ is a small positive constant introduced to prevent division by zero.

The weights were normalized to sum to one across all models, forming the normalized weights w_i':

$$w_i^{'} = \frac{w_i}{\sum_{j=1}^{K} w_j} \tag{4}$$

Finally, we computed the final prediction $\hat{Y}_{\text{final}}$ as the weighted sum of the mean predictions from all K models:

$$\hat{Y}_{\text{final}} = \sum_{i=1}^{K} w_i^{'} \cdot \mu_{\text{pred},i} \tag{5}$$

This weighted voting ensemble allows a final prediction that is not only a consensus across multiple models but also adjusts for the confidence level of each model's prediction.

c. Uncertainty quantification

In our model, the uncertainty related to the predictions was quantified using two statistical measurements: the average entropy[29] (ET) of the prediction probabilities and the average standard deviation[20] (STD). These measures provide a dual perspective on the confidence of the model's classifications, with entropy capturing the average uncertainty inherent in predictions, and the standard deviation reflecting the variability of prediction probabilities. Entropy is a statistical measure of randomness commonly used to characterize the uncertainty associated with a set of probabilities. In the context of our model, this is analogous to the concept of entropy in the

information theory, which quantifies the amount of information and uncertainty. The ET across all classes and predictions was calculated using the following equation:

$$\mathrm{ET} = -\frac{1}{KT}\sum_{i=1}^{K}\sum_{t=1}^{T}\sum_{c=1}^{C} p_{i,t}(c)\log\left(p_{i,t}(c)\right) \tag{6}$$

where $p_{i,t}(c)$ represents the probability of class c being predicted by the i_{th} model at the t_{th} forward pass, and C is the total number of classes. A higher ET value indicated greater uncertainty and lower confidence in the predictions. The standard deviation provides insight into the dispersion of the prediction probabilities around their mean values. This is an important indicator of prediction reliability, because predictions with a high standard deviation are less reliable. The STD is calculated as follows:

$$\mathrm{STD} = \sqrt{\frac{1}{KTC-1}\sum_{i=1}^{K}\sum_{t=1}^{T}\sum_{c=1}^{C}\left(p_{i,t}(c)-\mu_c\right)^2} \tag{7}$$

where μ_c is the mean prediction probability of class c across all models and forward passes. By incorporating these two metrics, the predictive uncertainty of the model can be comprehensively assessed. This assessment is not only crucial for the reliability of the medical diagnostic process but also provides valuable insights that can be used to guide decision-making under uncertainty.

d. Uncertainty reporting

To effectively communicate the predictive uncertainty of our model to medical professionals, we developed a multimodal reporting system that conveys uncertainty in several comprehensible formats:

(1) Single number representation

For a quick and straightforward interpretation, the uncertainty was first represented as a single numerical value. This was achieved using previously calculated ET and STD measures. These measures provide an immediate sense of confidence associated with the model's predictions.

(2) Predictive probability interval

The second reporting format is a predictive probability interval, that visualizes the range within which a model's predictions fall with a 95% level of confidence. This interval offers a visual representation of prediction certainty, providing practitioners with an intuitive grasp of the possible variability in diagnosis.

(3) Uncertainty-CAM

The Uncertainty-CAM utilizes Grad-CAM[30] visualizations from multiple forward passes by applying a weighted fusion based on the uncertainty of each pass. Specifically, forward passes that yield a higher entropy, indicating less certainty in the predictions, have a reduced impact on the final visualization. This process aims to create a composite heat map that delineates the areas in which the model confidently identifies the features of interest. This is a visualization tool that provides an aggregate picture of a model's focus across various prediction iterations.

2.3. Experimental design

The experimental design of this study was strategically organized to comprehensively evaluate various aspects of the automatic classification of vertebral bone metastases using abdominal CT scans. We focused on the efficiency of spine detection, performance of bone metastasis classification, accuracy of uncertainty quantification, and robustness of the model to healthy normal-control data and out-of-distribution data.

Spine detection efficiency is critical for the accurate extraction of the ROI, which is essential for subsequent classification tasks. To validate this, we applied a methodology widely adopted in previous research to evaluate the performance of our spine detection model, focusing on its accuracy and detection power.

One of the key objectives of this study was to outperform current state-of-the-art methods in classifying bone metastases in ROIs. We established a benchmark by evaluating the performance of our model against a baseline model. In particular, to compare the effectiveness of the EMCD technique proposed in this study, we also compared its performance with other uncertainty quantification methodologies, such as the MCDO and DE approaches. We used a retained data validation approach to validate the accuracy of our model's uncertainty estimation[19]. We sought to understand the correlation between uncertainty and classification accuracy by systematically evaluating the classification performance of a subset of data ranked by uncertainty. This approach not only tests the reliability of uncertainty quantification, but also explores its practical utility in improving classification performance.

Evaluating the performance of the classification model on a healthy normal-control dataset allowed us to assess its specificity and robustness. This step ensured that the model maintained high accuracy and low false positives when analyzing data from individuals without bone metastases. The ability of the model to recognize and accurately classify out-of-distribution data indicates its reliability and generalizability. We conducted experiments to evaluate how well the model with uncertainty quantification identifies data that deviate significantly from the training distribution[31]. This evaluation was critical for understanding the potential of the model in real-world clinical applications, where unseen variables are common. To further validate the EMCD model's robustness, we evaluated its performance on both the healthy normal-control and patient test datasets. For this

evaluation, if a CT slice from a healthy normal dataset was incorrectly predicted as metastasis, it was considered a misclassification for that patient. Conversely, if all slices of a patient with metastasis were incorrectly predicted as normal or disc, the patient was classified as normal.

Through this comprehensive experimental design, we aimed to not only improve state-of-the-art spinal bone metastasis classification but also provide meaningful insights into the effectiveness of uncertainty quantification in medical image analysis.

3. Results

3.1. Experimental setup

For single-class object detection, the YOLOv5m model was trained over 100 epochs with a learning rate of 0.01 and a batch size of 32, employing the SGD optimizer. The initial weights were sourced from a pretrained ImageNet dataset.

For the multiclass classification task, we utilized the Adam optimizer with a learning rate of 1e-5 to minimize categorical cross-entropy loss over 100 epochs and a batch size of 256. Similar to the detection model, the initial weights were adopted from a pretrained ImageNet dataset.

For the uncertainty quantification comparison, we incorporated an ensemble of K = 5 MCDO models, with each model subjected to T = 200 stochastic forward passes to gather predictive outcomes and uncertainty measures[31]. In parallel, the MCDO configuration was calibrated to reflect the EMCD setup, in which 1000 stochastic forward passes were executed. This number was deliberately chosen to ensure that the aggregate number of predictions matched that of the EMCD, thereby providing a balanced data foundation for our comparative analysis. To compare these approaches, we implement a DE technique with K = 5 models, creating a direct comparison with the architecture and evaluation metrics of the EMCD model.

3.2. Experimental results

3.2.1 Spine detection

In the first stage of the study, the YOLOv5m model exhibited robust performance for vertebral region detection, reflected by a high mean average precision score of 0.995 at an intersection over the union threshold of 0.5, on the test dataset.

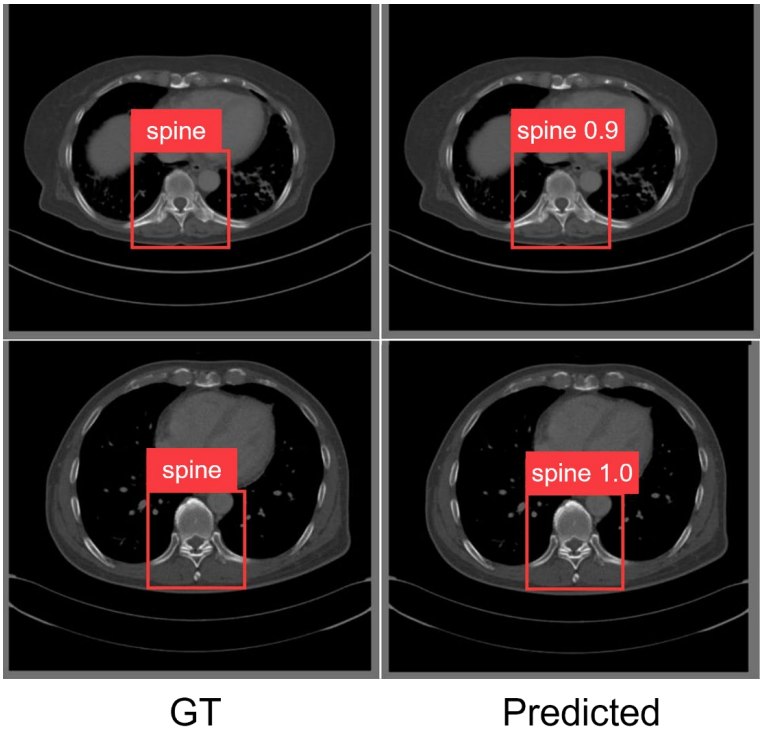


Figure 2. Spine detection results using the YOLOv5m model: representative examples of predicted bounding boxes along with the prediction probabilities. GT, ground truth; Predicted, predicted result.

Figure 2 illustrates the comparative results of the spinal region detection using abdominal CT scans. The left column shows the actual spinal areas outlined by red boxes representing the ground truth. The column on the right shows the predicted results. The consistent overlap between the

ground truth and predicted boxes demonstrates the high precision of the model in localizing spinal regions.

3.2.2 Multi-class classification

For the multiclass classification task, the accuracy, precision, recall, F1-score, AUC, and expected calibration error (ECE) were calculated to evaluate the performance of the model.

3.2.2.1 Baseline performance

To evaluate the classification performance of bone metastasis, four models were investigated: a baseline DL model (DenseNet201) without UQ, MCDO with T = 1000, a DE model with K = 5, and the proposed EMCD model with T = 200 and K = 5.

Table 3. Performance comparison of different models for classifying test datasets

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**	**AUC**	**ECE**
DenseNet201 without UQ	0.82	0.80	0.73	0.76	0.90	0.66
MCDO	0.82	0.84	0.72	0.77	0.91	0.30
DE	0.83	0.82	0.75	0.78	0.91	0.75
EMCD	0.86	0.87	0.77	0.82	0.93	0.09

AUC, macro-average area under the curve, ECE: expected calibration error, UQ: uncertainty quantification, MCDO: Monte Carlo dropout; DE: deep ensemble, EMCD: ensemble Monte Carlo dropout.

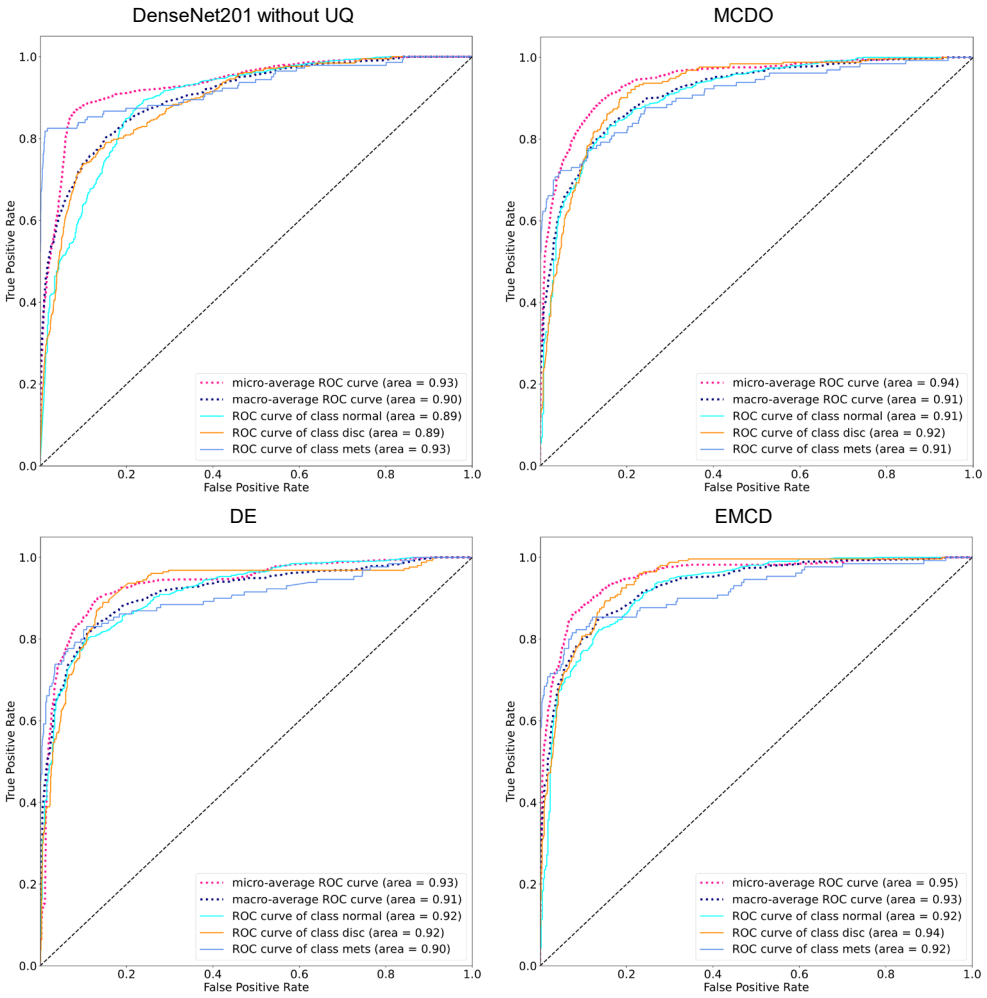


Figure 3. Receiver operating characteristic (ROC) curves obtained for the four considered deep learning models for the test datasets. UQ, uncertainty quantification; MCDO, Monte Carlo dropout; DE, deep ensemble; EMCD, ensemble Monte Carlo dropout; Mets, metastasis.

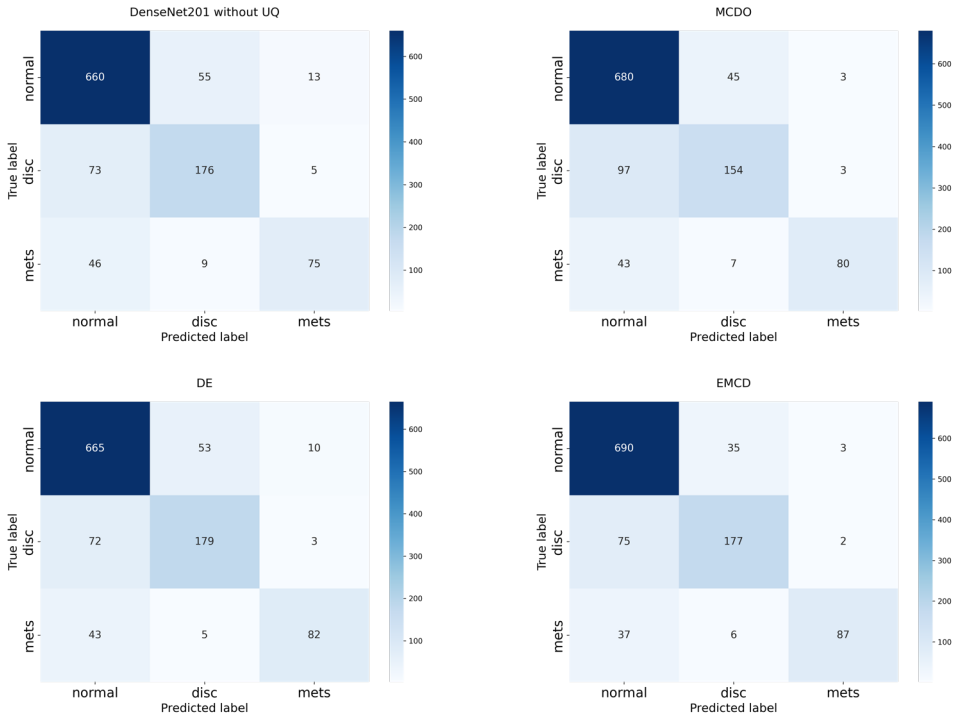


Figure 4. Confusion matrices obtained for the four considered deep learning models for the test datasets.

The obtained statistics, ROC curves, and the confusion matrices for the four competing DL models are presented in Table 3, Figure 3, and Figure 4, respectively.

Our proposed EMCD model, with T = 200 forward passes and K = 5, outperformed all the other models, demonstrating the highest accuracy, precision, recall, F1-score, AUC, and lowest ECE. Notably, the EMCD model attained an AUC of 0.93, which underscores its superior discriminative ability in bone metastasis classification. The enhanced AUC is a clear indicator that the EMCD model is particularly effective in distinguishing between the positive and negative classes, which is crucial for clinical decision-making. The precision of EMCD in managing the uncertainties inherent in medical imaging results in a significant gain in performance, particularly when compared to traditional DL models without UQ and even against other UQ methods such as MCDO and DE. In addition to these performance metrics, the EMCD model exhibited a low ECE of 0.09, indicating

that the predicted probabilities were well-calibrated and reflective of the true likelihood of correct classifications. These results underline the advantage of incorporating uncertainty quantification through the EMCD, as evidenced by the improved performance metrics across the board.

3.2.2.2 Healthy normal-control dataset evaluation

The effectiveness of the EMCD model was also validated using a healthy normal-control dataset, which was critical for assessing the specificity and ensuring the precision of the model in correctly classifying the absence of a disease.

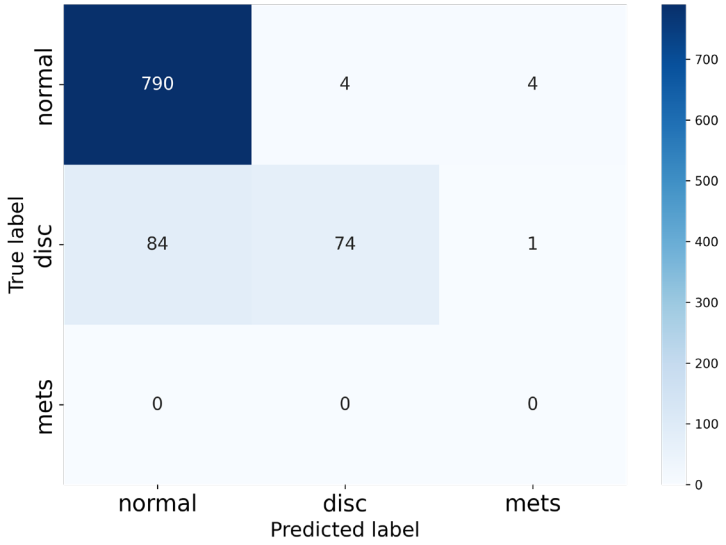


Figure 5. Confusion matrix or the proposed ensemble Monte Carlo dropout model for the healthy normal-control dataset.

Table 4. Performance of the proposed ensemble Monte Carlo dropout model on healthy normal-control dataset

Model	Accuracy	Precision	Recall	F1-score

EMCD	0.90	0.91	0.90	0.89

As shown in Table 4, our EMCD model demonstrated a strong performance accuracy (90%) in classifying the healthy normal-control dataset. These metrics underscore the robustness of the model and confirm its enhanced capability to classify healthy individuals accurately and minimize the risk of false diagnoses. Figure 5 illustrates the confusion matrix for the EMCD model, which graphically represents the classification results. The matrix contains a substantial number of true negatives, with few cases in which "normal" was incorrectly identified. This low misclassification rate, particularly for "normal" to "disc" or "metastasis," highlights the model's precision in avoiding false alarms, which is a critical aspect in clinical settings.

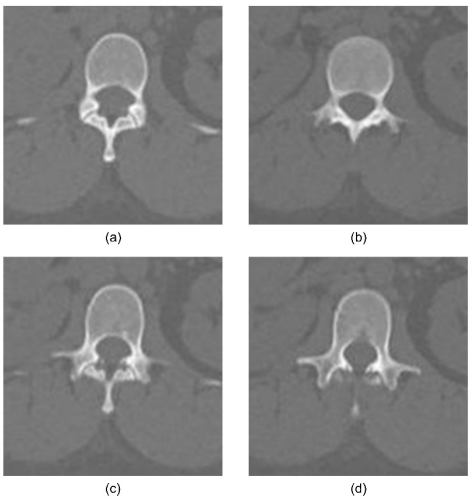


Figure 6. Representative cases incorrectly predicted by the ensemble Monte Carlo dropout model in the healthy normal-control dataset: (a) normal case predicted as metastasis, (b) normal case predicted as metastasis, (c) normal case predicted as metastasis with venous plexus, (d) normal case predicted as metastasis with Schmorl's nodes.

Figure 6 showcases four representative CT image examples from the healthy normal-control test dataset that were incorrectly predicted by the EMCD model, that is, normal slices were misclassified as pathological conditions. Notably, despite the images shown in cases (c) and (d) being non-pathological, they possess features that appear lesion-like. These characteristics were identified by radiologists as either a venous plexus[32] or Schmorl's nodes[33]. This suggests that even in instances of incorrect predictions, the EMCD model is capable of recognizing abnormal patterns that are similar to specific lesions.

To further validate the EMCD model's robustness, particularly in clinical settings, we expanded our evaluation to include patient-level experimental results in addition to the slice-level analysis. The results, summarized in Table 5, indicate that the EMCD model achieved a recall of 100% and an accuracy of 91% on the patient test dataset, confirming its efficacy in distinguishing between normal and bone metastasis patients at the patient level. The high recall value demonstrates the model's ability to correctly identify all positive cases of metastasis, while the overall accuracy underscores its reliability and robustness in clinical settings.

Table 5. Performance of the proposed ensemble Monte Carlo dropout model on healthy normal-control and patient test datasets

Model	TP	FN	TN	FP	Accuracy	Precision	Recall	F1-score
EMCD	12	0	9	2	0.91	0.86	1.00	0.92

TP, true-positive; FN, false-negative; TN, true-negative; FP, false-positive.

3.2.3 Uncertainty quantification

3.2.3.1 Retained data evaluation

In evaluating the quality of uncertainty quantification, one measure is the model's ability to maintain high performance because less certain predictions are systematically excluded or referred

to based on uncertainty estimates. This approach reflects a model's ability to identify and prioritize its most confident predictions.

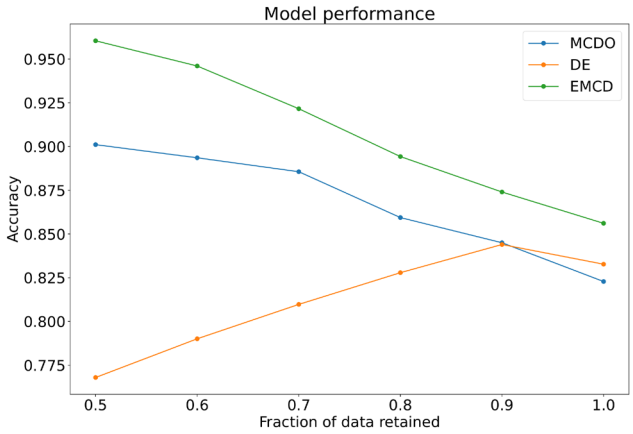


Figure 7. Visualizes the model accuracy in relation to the fraction of data retained, defined by uncertainty thresholds ranging from 0.5 to 1.0.

Table 6. Comparison of the performance of the proposed ensemble Monte Carlo dropout model with two uncertainty quantification models as a fraction of retained data

Model	**50% data retained**		**70% data retained**		**90% data retained**	
	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
MCDO	0.94	0.89	0.93	0.88	0.91	0.84
DE	0.88	0.77	0.90	0.81	0.91	0.84
EMCD	0.96	0.96	0.95	0.92	0.93	0.87

The MCDO model, while not surpassing the EMCD model in terms of overall performance, exhibited improved accuracy when data with higher uncertainty were removed, as indicated by the slope in Figure 7. This behavior suggests that MCDO provides reliable uncertainty estimates that

effectively identify less certain predictions. In contrast, the DE model, despite starting with a better performance than MCDO, showed a decline in accuracy as more uncertain data were excluded, hinting at less reliable uncertainty estimates. In particular, the EMCD model exhibits a steep slope, which indicates a more effective uncertainty estimation. Models that achieve steeper slopes in such evaluations are considered to produce better uncertainty estimates, because they can systematically exclude less reliable predictions. The EMCD model consistently outperformed the other models at all data retention levels. Notably, at 50% data retention, the EMCD model achieved an AUC of 0.96 and an impressive accuracy of 96%, confirming the robustness of the model and the efficacy of the EMCD technique in prioritizing the most reliable predictions. Even with 90% data retention, the EMCD model maintained a high level of accuracy, affirming its capability to provide dependable diagnostic predictions across varying levels of uncertainty.

3.2.3.2 Detection of out-of-distribution data

Figure 8. MedMNIST (https://medmnist.com) sample images and corresponding labels fed to deep learning models as out-of-distribution dataset. CXR, chest X-ray; Hand, hand X-ray; BreastMRI, breast magnetic resonance imaging.

This evaluation aimed to investigate the model's proficiency in quantifying uncertainty when presented with out-of-distribution data. To this end, the MedMNIST[34] dataset (Figure 8) containing image categories not observed during training was utilized as a benchmark for out-of-distribution (OOD) data. One hundred images from each class (CXR, Hand, and BreastMRI) were selected

randomly. Model performance was examined by measuring the STD and ET, which are metrics indicative of uncertainty in the model's predictions. The results for the MedMNIST[34] dataset are listed in Table 6.

Table 7. Comparison of proposed ensemble Monte Carlo dropout model results with two uncertainty quantification models for detecting out-of-distribution data in MedMNIST dataset

Method	**Class**							
	CXR		**Hand**		**BreastMRI**		**In-distribution**	
	STD ↑	ET ↑	STD ↑	ET ↑	STD ↑	ET ↑	STD ↓	ET ↓
MCDO	0.153	4.077	0.287	5.906	0.234	5.770	0.056	0.363
DE	0.166	0.929	0.141	1.302	0.147	1.427	0.080	1.496
EMCD	0.360	5.587	0.315	5.969	0.258	5.896	0.136	1.463

CXR, chest X-ray; Hand, hand X-ray, BreastMRI: breast magnetic resonance imaging; STD, standard deviation; ET, entropy.

The data indicate that the EMCD model consistently registered higher STD and ET values across OOD labels than the in-distribution test data. This enhanced uncertainty signaling by the EMCD suggests its potential for greater reliability in real-world applications, where distinguishing between familiar and unfamiliar inputs is crucial. The MCDO model also reflected an increase in uncertainty for the OOD data, albeit to a lesser degree than the EMCD, whereas the DE model demonstrated the lowest effectiveness in uncertainty estimation among the evaluated models. In contrast, all models reported a lower uncertainty for the in-distribution (abdominal CT) test data, as expected. Besides the above results, we also classified the OOD data concerning the In-distribution data based on the uncertainty metrics of the EMCD model, i.e., standard deviation and entropy.

Table 8. Accuracy of classifying out-of-distribution data using uncertainty measurements with the ensemble Monte Carlo dropout model

Uncertainty measurements	**Class**

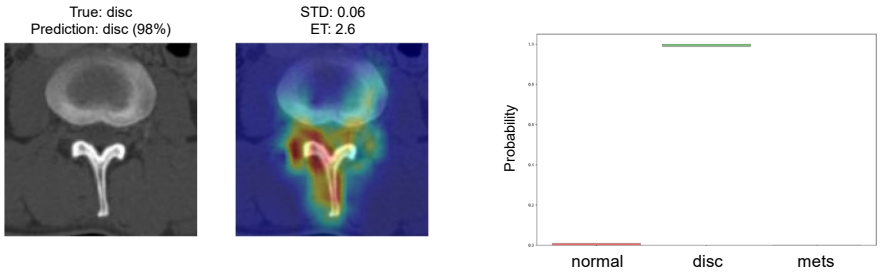
	CXR	Hand	BreastMRI	Total
STD	0.93	0.88	0.56	0.79
ET	1.00	1.00	1.00	1.00

CXR, chest X-ray; Hand, hand X-ray, BreastMRI: breast magnetic resonance imaging; STD, standard deviation; ET, entropy.

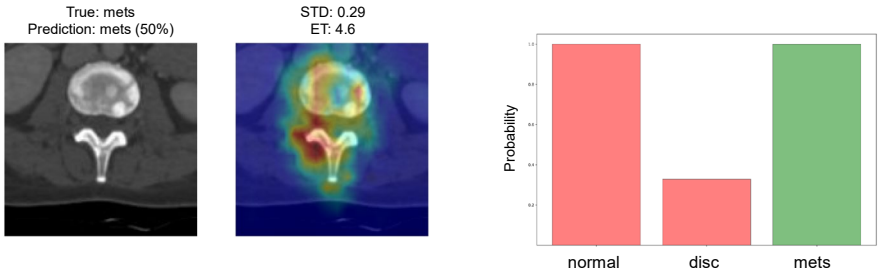
We classified the OOD data from the in-distribution data based on the STD and ET, which are the uncertainty metrics of the EMCD model. Using the maximum STD of 0.25 from the in-distribution data as a threshold, the result is 79%. When the maximum value of ET measured from the in-distribution data, 2.18, is used as a threshold, the result is 100%.

3.2.3.3 Uncertainty reporting

In this section, we introduce the results of the four types of uncertainty reporting (single number, probability interval, and Uncertainty-CAM) that will be presented to medical professionals.



(a) correct and certain prediction



(b) correct and uncertain prediction

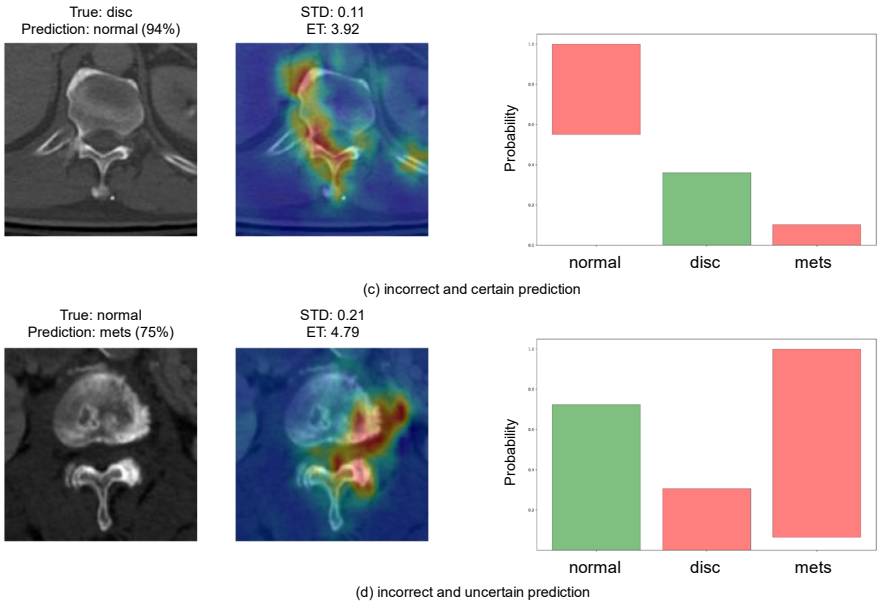


Figure 9. Examples of uncertainty reporting by the ensemble Monte Carlo dropout model: (a) correct and certain prediction, (b) correct and uncertain prediction, (c) incorrect and certain prediction, (d) incorrect and uncertain prediction, where the bar chart's green color represents the actual truth label. STD, standard deviation; ET, entropy.

Figure 9 showcases examples of uncertainty reporting obtained from CT slice images using the EMCD method. The uncertainty and Uncertainty-CAM, obtained through 1,000 predictions, are displayed along with the final prediction value. In the case depicted in (a), the prediction outcome is normal, and the model is confident in its correct decision (evidenced by a low uncertainty estimate [STD = 0.06, ET = 2.6]). Additionally, through the probability interval bar chart, the low uncertainty is easily identified by the short vertical length of the bar. In contrast, the case shown in (b) demonstrates that while the model predicts normalcy, it indicates high uncertainty (evidenced by a high uncertainty estimate [STD = 0.29, ET = 4.6]). The tall vertical length of the probability interval bar chart signifies low confidence in this prediction, akin to the model stating "I do not know." Case (c) in Figure 9 illustrates an incorrect prediction by the model. The uncertainty indicator and the vertical length of the bar chart convey that the model is confident in its decision, which clearly

indicates a mistake. In case (d), although the model makes an incorrect prediction, it shows that the model is uncertain about its decision. Because the model lacks confidence in its prediction, it suggests that medical professionals might seek a second opinion on the image[35].

4. Discussion

In this study, we introduce the EMCD model, which represents a significant advancement in medical imaging for the detection and classification of spinal bone metastases. Our findings demonstrate that EMCD not only enhances the accuracy of detecting bone metastases but also substantially improves the quantification of predictive uncertainty compared to existing methodologies such as the MCDO[20] and DE[21] approaches. This dual achievement is critical in the context of clinical diagnostics, where the precision of detection and confidence in diagnostic predictions can significantly influence patient management and treatment outcomes. The superior performance of the EMCD model, as evidenced by rigorous testing on both the MedMNIST[34] dataset for OOD data and a healthy normal-control dataset, highlights its potential in a clinical setting. Notably, the ability of the model to accurately estimate uncertainty offers a clearer understanding of its predictions, thereby facilitating more informed clinical decisions. This advancement addresses the significant challenge in deploying AI in healthcare by bridging the gap between AI prediction and clinical interpretability. A unified review by Lambert et al. discussed the challenges of the low acceptance of DL models in clinical practice, owing to the lack of transparency in decision-making processes. They emphasized that uncertainty quantification can significantly improve the interpretability and acceptability of DL predictions in medical image analysis, thereby fostering trust among end-users[36]. Xue et al. introduced a Bayesian convolutional neural network framework that quantified the uncertainty in DL predictions, providing surrogate estimates of the true error from the network model and measurement itself. This approach is crucial to ensure the reliability of medical diagnoses derived from imaging data[37].

A particularly compelling aspect of our findings was observed during the retained data evaluation experiment, underscoring the effectiveness of the EMCD model in handling uncertain predictions. By systematically excluding uncertain predictions based on predefined thresholds, the EMCD model demonstrated its potential to prioritize high-confidence predictions that are instrumental in clinical

scenarios requiring high diagnostic accuracy. This approach to managing and quantifying uncertainty could revolutionize how clinicians interpret and trust AI-generated diagnoses, particularly in ambiguous or borderline cases. Our study also incorporated ECE to assess the calibration of the EMCD model predictions. A low ECE value indicates that our model's predicted probabilities are well-calibrated, meaning that they accurately reflect the likelihood of correct classifications. This is a crucial aspect for clinical applicability as it ensures that the model's predictions can be trusted by clinicians. Although we have not yet conducted experiments on a per-patient basis, our findings suggest that the EMCD model is robust and reliable for slice-based analyses, with the potential for future expansion to patient-level evaluations.

This study represents a pioneering exploration of the uncertainty in bone metastasis predictions provided by DL models. The manner in which uncertainty is reported to radiologists through a single number, uncertainty intervals, and Uncertainty-CAM represents a significant advancement in clinical diagnostics. This multifaceted approach enables radiologists to interpret AI-generated predictions with a nuanced understanding of their reliability. Prior to this study, two recent studies used DL to detect bone metastases using CT images[16,17]. In contrast to previous research that largely focused on the accuracy of DL models in detecting spinal bone metastases, our study emphasized the importance of uncertainty quantification. The novel approach of the EMCD model for uncertainty estimation provides an essential tool for clinicians, offering insight into the model's confidence level for each prediction. This aspect is particularly beneficial for managing cases in which the model identifies potential metastases with high uncertainty, highlighting the need for further clinical evaluation. The categorization of CT slices into three distinct classes, normal, disc, and metastasis, was based on our observation of disc levels on axial-plane CT images resembling osteolytic lesion shapes. This insight illuminates the intricacies involved in accurately classifying spinal structures and pathologies, necessitating a refined annotation approach capable of distinguishing these crucial differences. This discernment further accentuates the advanced capabilities of our proposed model in navigating the complex landscape of spinal imaging.

However, our study had some limitations. The slice-wise labeling approach adopted in this study may not perfectly capture the complexity of certain cases, particularly those in which both the disc and the vertebral body are present in a single slice. This limitation highlights the need for more

nuanced labeling and analysis methods that can accurately reflect the multifaceted nature of spinal anatomy and pathology. Moreover, as this was a retrospective study, the real-world clinical applicability of our model requires further validation. Future research should include prospective multicenter studies to evaluate the performance of the model across diverse clinical settings and patient populations. Furthermore, although our proposed method for uncertainty quantification provides a mathematical measure of prediction confidence, it may not always fully align with the clinical interpretations of the uncertainty faced by radiologists. This misalignment highlights an area for future refinement: enhancing uncertainty quantification techniques to more closely mirror the complexities and realities of clinical decision-making.

In addition to enhancing our model for volumetric analysis, integrating patient-level research is essential. Our model analyzes CT slices individually, which may not provide a holistic view of the patient's condition. Future research should develop methods that consider all patient cases and encompass all relevant imaging and clinical data to provide a more comprehensive assessment. This approach allows for a better understanding of disease progression and variability across patients, further aligning the use of the model with clinical workflows.

Future directions for this research include addressing the limitations by exploring alternative labeling strategies, conducting prospective studies to validate the model's applicability in real-world settings, and refining the uncertainty quantification method to align better with clinical expectations. Further efforts will focus on developing and testing explainable AI features that meet the specific requirements of clinical practice.

5. Conclusion

The proposed EMCD model represents a significant advancement in the fully automated detection of spinal bone metastases using abdominal CT, achieving an AUC of 0.93. This improvement not only enhances diagnostic accuracy but also introduces an innovative approach for quantifying uncertainty. By providing clinicians with predictions and associated uncertainty assessments, the EMCD model offers insights that were not previously available in existing models, thereby facilitating more informed clinical decision-making. This development holds promise for aiding radiologists in the diagnosis of spinal bone metastases using CT scans, potentially improving patient care and quality of life.

References

1. Boland PJ, Lane JM, Sundaresan N. Metastatic disease of the spine. Clin Orthop Relat Res 1982:95-102.
2. Guillevin R, Vallee JN, Lafitte F, Menuel C, Duverneuil NM, Chiras J. Spine metastasis imaging: review of the literature. J Neuroradiol 2007;34:311-21.
3. Yin JJ, Pollock CB, Kelly K. Mechanisms of cancer metastasis to the bone. Cell Res 2005;15:57-62.
4. Shaw B, Mansfield FL, Borges L. One-stage posterolateral decompression and stabilization for primary and metastatic vertebral tumors in the thoracic and lumbar spine. J Neurosurg 1989;70:405-10.
5. Mundy GR. Metastasis to bone: causes, consequences and therapeutic opportunities. Nat Rev Cancer 2002;2:584-93.
6. Bach F, Larsen BH, Rohde K, Børgesen SE, Gjerris F, Bøge-Rasmussen T, et al. Metastatic spinal cord compression. Occurrence, symptoms, clinical presentations and prognosis in 398 patients with spinal cord compression. Acta Neurochir (Wien) 1990;107:37-43.
7. Heindel W, Gübitz R, Vieth V, Weckesser M, Schober O, Schäfers M. The diagnostic imaging of bone metastases. Deutsches Ärzteblatt International 2014;111:741.
8. Chmelik J, Jakubicek R, Walek P, Jan J, Ourednicek P, Lambert L, et al. Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data. Medical image analysis 2018;49:76-88.
9. Hammon M, Dankerl P, Tsymbal A, Wels M, Kelm M, May M, et al. Automatic detection of lytic and blastic thoracolumbar spine metastases on computed tomography. European radiology 2013;23:1862-70.
10. Rybak LD, Rosenthal DI. Radiological imaging for the diagnosis of bone metastases. Q J Nucl Med 2001;45:53-64.
11. Lee YH, Kim S, Lim D, Suh JS, Song HT. Spectral parametric segmentation of contrast-enhanced dual-energy CT to detect bone metastasis: feasibility sensitivity study using whole-body bone scintigraphy. Acta Radiol 2015;56:458-64.
12. Pache G, Krauss B, Strohm P, Saueressig U, Blanke P, Bulla S, et al. Dual-energy CT virtual noncalcium technique: detecting posttraumatic bone marrow lesions--feasibility study. Radiology 2010;256:617-24.
13. Sommer WH, Johnson TR, Becker CR, Arnoldi E, Kramer H, Reiser MF, et al. The value of dual-energy bone removal in maximum intensity projections of lower extremity

computed tomography angiography. Invest Radiol 2009;44:285-92.

14. Burns JE, Yao J, Wiese TS, Muñoz HE, Jones EC, Summers RM. Automated detection of sclerotic metastases in the thoracolumbar spine at CT. Radiology 2013;268:69-78.
15. Chang CY, Buckless C, Yeh KJ, Torriani M. Automated detection and segmentation of sclerotic spinal lesions on body CTs using a deep convolutional neural network. Skeletal Radiol 2022;51:391-9.
16. Koike Y, Yui M, Nakamura S, Yoshida A, Takegawa H, Anetai Y, et al. Artificial intelligence-aided lytic spinal bone metastasis classification on CT scans. Int J Comput Assist Radiol Surg 2023;18:1867-74.
17. Noguchi S, Nishio M, Sakamoto R, Yakami M, Fujimoto K, Emoto Y, et al. Deep learning-based algorithm improved radiologists' performance in bone metastases detection on CT. Eur Radiol 2022;32:7976-87.
18. Rasheed K, Qayyum A, Ghaly M, Al-Fuqaha A, Razi A, Qadir J. Explainable, trustworthy, and ethical machine learning for healthcare: A survey. Comput Biol Med 2022;149:106043.
19. Filos A, Farquhar S, Gomez A, Rudner T, Kenton Z, Smith L, et al. A Systematic Comparison of Bayesian Deep Learning Robustness in Diabetic Retinopathy Tasks; 2019.
20. Gal Y, Ghahramani Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In: Maria Florina B, Kilian QW, editors. Proceedings of The 33rd International Conference on Machine Learning. Proceedings of Machine Learning Research: PMLR; 2016. p.1050--9.
21. Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California, USA: Curran Associates Inc.; 2017. p.6405–16.
22. Aldhahi W, Sull S. Uncertain-CAM: Uncertainty-Based Ensemble Machine Voting for Improved COVID-19 CXR Classification and Explainability. Diagnostics 2023;13:441.
23. Zhu X, Lyu S, Wang X, Zhao Q. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW): IEEE Computer Society; 2021. p.2778-88.
24. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR): IEEE Computer Society; 2016. p.779-88.
25. Deng J, Dong W, Socher R, Li LJ, Kai L, Li F-F. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009. p.248-55.

26. Huang G, Liu Z, Maaten LVD, Weinberger KQ. Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR): IEEE Computer Society; 2017. p.2261-9.
27. Labach A, Salehinejad H, Valaee S. Survey of dropout methods for deep neural networks. arXiv preprint arXiv:1904.13310 2019.
28. Dogan A, Birant D. A Weighted Majority Voting Ensemble Approach for Classification. 2019 4th International Conference on Computer Science and Engineering (UBMK); 2019. p.1-6.
29. Gal Y. Uncertainty in deep learning. 2016.
30. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 IEEE International Conference on Computer Vision (ICCV); 2017. p.618-26.
31. Abdar M, Salari S, Qahremani S, Lam H-K, Karray F, Hussain S, et al. UncertaintyFuseNet: Robust uncertainty-aware hierarchical feature fusion model with Ensemble Monte Carlo Dropout for COVID-19 detection. Information Fusion 2023;90:364-81.
32. Eckenhoff JE. The vertebral venous plexus. Can Anaesth Soc J 1971;18:487-95.
33. Kyere KA, Than KD, Wang AC, Rahman SU, Valdivia-Valdivia JM, La Marca F, et al. Schmorl's nodes. Eur Spine J 2012;21:2115-21.
34. Yang J, Shi R, Ni B. MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis. 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI); 2021. p.191-5.
35. Raghu M, Blumer K, Sayres R, Obermeyer Z, Kleinberg RD, Mullainathan S, et al. Direct Uncertainty Prediction for Medical Second Opinions. International Conference on Machine Learning; 2018.
36. Lambert B, Forbes F, Doyle S, Dehaene H, Dojat M. Trustworthy clinical AI solutions: A unified review of uncertainty quantification in Deep Learning models for medical image analysis. Artif Intell Med 2024;150:102830.
37. Xue Y, Cheng S, Li Y, Tian L. Reliable deep-learning-based phase imaging with uncertainty quantification. Optica 2019;6:618-29.

Abstract in Korean

앙상블 몬테 카를로 드롭아웃을 활용한 복부 CT 를 통한 척추골 전이 자동 분류의 불확실성 정량화

복부 CT 스캔을 이용해 척추의 골 전이를 자동으로 탐지하고 분류하는 것을 개선하는 것을 목표로 한다. 특히, 정량화된 불확실성을 도입함으로써 진단의 민감도와 효율성에 대한 문제를 해결하고자 한다.

이 후향적 연구에서는 척추 골 전이 진단을 받은 116 명의 환자로부터 얻은 11,468 개의 복부 CT 이미지를 분석했으며, 11 명의 건강한 대조군으로부터 얻은 957 개의 이미지를 데이터셋에 포함시켰다. 이미지는 '정상', '디스크', '전이'으로 분류되어 주석 처리되었다. 불확실성 추정을 위해 DenseNet201 구조에 dropout 층을 추가하였고, 정밀한 척추 영역 탐지를 위해 YOLOv5m 을 사용하며, 불확실성 가중치 투표 앙상블을 통해 새롭고 효과적인 ensemble Monte Carlo dropout (EMCD) 모델을 도입하였다. 계산된 불확실성은 수치 값, 예측 확률 간격, 그리고 Uncertainty-CAM 를 통해 표현되었다. 성능 평가는 척추 탐지의 효율성, 전이 분류의 정확도, 그리고 건강한 대조군 및 분포 외 데이터에 대한 모델의 견고성에 중점을 두었다.

YOLOv5m 모델은 척추 탐지에서 mean average precision 0.995 를 달성했다. EMCD 모델은 다중 클래스 분류에서 area under the receiver operating characteristic curve (AUC) 0.93 으로 우수한 성능을 보여, 기존 및 기타 불확실성 정량화 모델들을 능가했다. 50%의 데이터를 유지할 때, EMCD 모델은 AUC 0.96 과 96%의 정확도를 달성했다. 건강한 대조군 데이터셋에서는 EMCD 모델이 90%의 높은 정확도를 유지했다.

EMCD 모델은 척추 골 전이의 자동 탐지 및 분류에서 현저한 향상을 제공하며, 우수한 정확도와 더불어 예측과 함께 불확실성 측정을 동시에 제공하는 새로운 접근법을 도입함으로써, 임상의가 기존 딥러닝 모델에서는 볼 수 없었던 새로운 정보에 기반한 의사 결정을 내릴 수 있을 것으로 보인다. 이는 정보에 기반한 보다 정확한 임상 의사결정을 가능하게 하여, 환자의 삶의 질에 긍정적인 영향을 미칠 수 있다.

핵심되는 말 : 척추; 컴퓨터 단층 촬영; 전이; 딥러닝; 불확실성