

DHAN: Dilated Heatmap Attention Network for Omental Metastasis Detection in CT

Seokho Woo

The Catholic University of Korea
Department of Data Science
Seoul, South Korea
seokhowoo@catholic.ac.kr

Hanhong Lee*

Seoul St. Mary's Hospital,
The Catholic University of Korea
Division of Gastrointestinal Surgery,
Department of Surgery, College of Medicine
Seoul, South Korea
painkiller9@catholic.ac.kr

Bohyun Kim*

Seoul St. Mary's Hospital,
The Catholic University of Korea
Department of Radiology
Seoul, South Korea
kbh@catholic.ac.kr

Youjin Shin*

The Catholic University of Korea
Department of Data Science
Seoul, South Korea
yjshin@catholic.ac.kr

Abstract

Peritoneal metastasis (PM) is a serious complication of abdominal malignancies with extremely poor prognosis, where untreated patients survive approximately 3-6 months. This study aims to develop a deep learning-based model to accurately predict omental metastasis (OM), one of the most common and clinically important forms of PM, from preoperative computed tomography (CT) images to support more objective clinical decision-making.

To this end, this study proposes a Dilated Heatmap Attention Network (DHAN), a deep learning-based classification model integrating three key innovations. First, an omentum filter(OMF) is introduced to automatically remove clinically insignificant slices based on positional statistics of OM lesions. Second, to address computational constraints in clinical environments, EfficientNet-B0 is selected as the lightweight backbone (8.4 times smaller and 6.1 times faster than ResNet-50) with dilated convolution to capture fine-grained lesion features while preserving feature map resolution. Third, an heatmap attention (HA) is proposed to focus on anatomically relevant regions based on lesion positional information.

At the slice level, DHAN achieved a sensitivity of 0.74, a specificity of 0.97, an F1-score of 0.76, and an area under the curve (AUC) of 0.94, demonstrating strong performance compared with existing image-based diagnostic models. These results suggest the feasibility of CT-based deep learning for OM detection and provide a basis for further development toward clinical decision support.

*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '26, Jeju, South Korea

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/2026/08
<https://doi.org/XXXXXXX.XXXXXXX>

CCS Concepts

• Computing methodologies → Computer vision.

Keywords

Peritoneal Metastasis, Deep Learning, CT Imaging, Dilated Convolution, EfficientNet, CNN, Computer-Aided Diagnosis, Medical Image Classification

ACM Reference Format:

Seokho Woo, Bohyun Kim, Hanhong Lee, and Youjin Shin. 2026. DHAN: Dilated Heatmap Attention Network for Omental Metastasis Detection in CT. In *Proceedings of Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '26)*. ACM, New York, NY, USA, 7 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 Introduction

Cancer is one of the leading causes of death worldwide. Approximately 20 million new cancer cases were reported in 2022, with 9.7 million deaths [2]. Peritoneal metastasis (PM) is a common form of metastasis in abdominal malignancies such as gastric, colorectal, and ovarian cancers, occurring in up to 40% of patients with metastatic gastric cancer [10]. Early diagnosis of PM is remarkably low, with most cases detected at advanced stages [17]. Therefore, early detection and objective diagnostic support can significantly improve patient prognosis and treatment outcomes, making it clinically crucial.

PM is known as a challenging disease to diagnose, presenting low sensitivity (~50%) on computed tomography (CT) imaging [10]. This is attributed to the anatomical characteristics and imaging features of the peritoneum. The peritoneum is a thin membrane surrounding visceral organs, making it difficult to distinguish from surrounding adipose tissue on CT images [11]. Furthermore, lesions appear in diverse locations and morphologies, with inflammatory changes and malignant tumors often presenting similar imaging findings. Due to these factors, inter-reader agreement among clinicians is low, creating a persistent demand for objective and consistent diagnostic assistance systems [5]. To address these limitations, statistical approaches combining manually extracted quantitative

CT features such as tumor size and peritoneal thickness with logistic regression and nomograms have been proposed [3, 8]. However, these methods rely on manual feature extraction, introducing inter-observer variability and limiting generalizability.

To overcome these diagnostic challenges, deep learning-based diagnostic assistance systems have recently been proposed. Deep learning demonstrates excellent performance in medical image analysis [1], and several artificial intelligence models for PM diagnosis have been developed. Wang et al. presented a model combining DenseNet121 and SVM on CT images from 168 PM patients, where peritoneal ROIs were manually delineated across three abdominal regions [15]. Zhang et al. proposed a radiomics-based deep learning model utilizing PET/CT multimodal data with manual ROI delineation [16]. Jiang et al. developed PMetNet, a deep learning model for predicting occult peritoneal metastasis in gastric cancer from CT images alone, but similarly relied on manual tumor region delineation [9]. These studies depend on manual ROI delineation, introducing inter-observer variability. Furthermore, models relying on additional modality data such as PET require data beyond CT imaging alone. Motivated by these limitations, this study proposes a deep learning-based classification model, Dilated Heatmap Attention Network (DHAN), which integrates automated preprocessing, a model relying solely on CT images with a lighter and faster backbone, and an anatomically informed attention mechanism. The four key innovations of this study are as follows:

- An **omentum filter (OMF)** based on positional statistics of OM lesions from training data is incorporated to automatically remove clinically insignificant slices, thereby enhancing model learning and improving diagnostic accuracy by eliminating unnecessary information.
- **Dilated convolution** with an expanded receptive field is employed to simultaneously capture fine-grained lesion features and broad contextual information while preserving feature map resolution. To address deployment constraints in clinical environments, EfficientNet-B0 is selected as the backbone (8.4 times smaller and 6.1 times faster in inference compared to ResNet-50) [14], enabling rapid deployment in resource-limited settings.
- A **heatmap attention mechanism (HA)** directs the model to focus on anatomically relevant regions by generating a heatmap based on positional information of OM lesions and integrating it into the feature map, thereby prioritizing regions where lesions are likely to exist.
- Using **real clinical CT data** from 288 patients at Seoul St. Mary's Hospital, we demonstrate that DHAN achieves an AUC of 0.94, sensitivity of 0.74, and specificity of 0.97 for omentum detection, confirming its clinical significance for real-world deployment.

Through automated preprocessing and anatomically informed attention, DHAN enables PM diagnosis using only CT images, without requiring manual ROI delineation or additional imaging modalities.

2 Method

The proposed method, DHAN, consists of three components: preprocessing for meaningful slice selection, dilated convolution with

an expanded receptive field, and a heatmap attention mechanism focused on anatomically relevant regions. The overall architecture is illustrated in Figure 1.

2.1 Preprocessing

Because OM is confined to the abdominal region, it is unnecessary to use all CT slices for training. In practice, including irrelevant slices introduces noise that interferes with model training. The upper and lower ends of a full CT scan contain thoracic and lower-body regions, providing no meaningful information for abdominal analysis. Specifically, the slice positions containing OM were analyzed across 100 positive patients in the training data. Because the total number of slices differs across patients, positions were normalized to a 0–100% scale. OM lesions were distributed within the range of 5.7% to 93.9% of the total slices, and all positive OM cases fell within this range, confirming that no clinically relevant lesions were excluded. The distribution of these lesion positions is illustrated in Figure 3.

Rather than relying on fixed anatomical landmarks, this boundary was derived statistically from the actual distribution of lesion positions in the training data. Visual inspection of this empirically-derived boundary (5.7%–93.9%) confirmed that it closely corresponded to the actual anatomical boundary between the thoracic and lower-body regions, indicating that the proposed boundary is not only automatically derived from data but also anatomically valid. As shown in Figure 2, these irrelevant regions can therefore be automatically excluded during preprocessing using the omentum filter (OMF), while clinically relevant slices are retained for training. Slices outside this range were removed from both the training and test data, enabling the model to focus on meaningful diagnostic information. The OMF preprocessing step, demonstrated in Figure 1 (a), eliminates unnecessary information by including only clinically relevant slices during training, thereby enhancing the model's learning efficiency and reducing training noise.

2.2 Dilated Convolution

Standard convolution has limitations in expanding the receptive field without increasing the number of parameters. To address this, blocks four and six of EfficientNet-B0 were replaced with dilated convolution, allowing the receptive field to be expanded while preserving spatial resolution. As shown in Figure 1 (b), this architectural modification enables the model to capture fine-grained lesion features while maintaining high-resolution feature maps.

In the original EfficientNet-B0, the feature map resolution is reduced to 16×16 after passing through these blocks. In contrast, the proposed model maintains a resolution of 64×64 , which is 4 times larger in each spatial dimension. Consequently, while each feature map cell in EfficientNet-B0 covers a 32×32 pixel region, the proposed model observes an 8×8 pixel region in detail. This high-resolution feature map is advantageous for detecting subtle lesion differences and small lesions.

Dilated convolution is defined as follows:

$$y(i, j) = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} x(i + d \cdot m, j + d \cdot n) \cdot w(m, n), \quad (1)$$

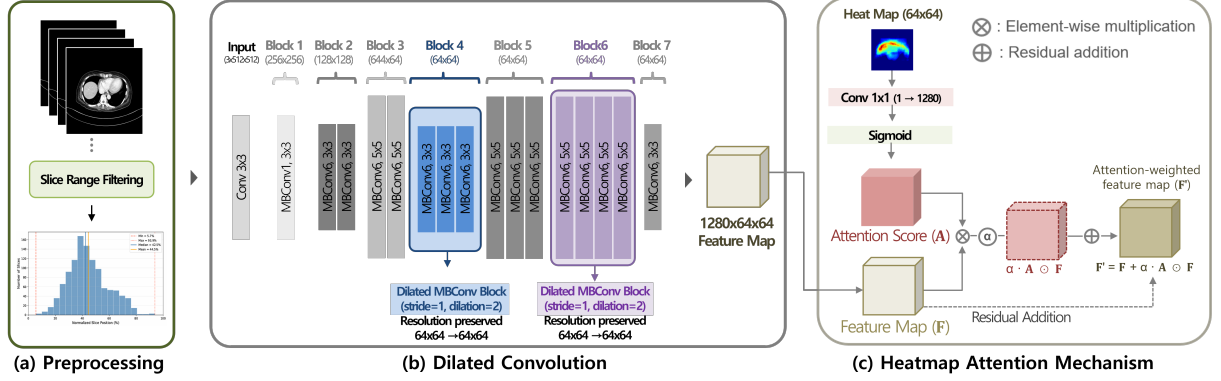


Figure 1: Overall architecture of the proposed DHAN model. The backbone is EfficientNet-B0, combined with dilated convolution blocks and an HA.



Figure 2: CT slices demonstrating the OMF range. (Left) Upper-body CT, excluded from both training and evaluation. (Middle) Lower-body CT, excluded from both training and evaluation. (Right) Middle region slice containing OM lesions, included in both training and evaluation. Blue highlighting indicates the actual OM lesions.

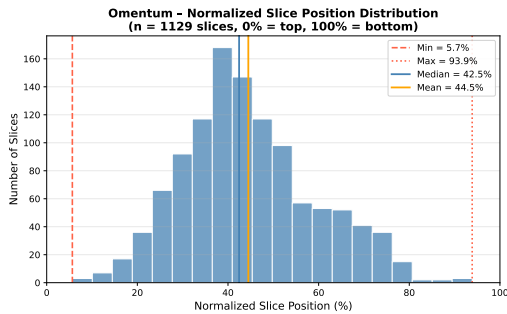


Figure 3: Distribution of OM lesion positions across 100 patients, normalized to a 0–100% scale. Lesions are concentrated between 5.7% and 93.9%.

where i and j are the position indices of the output feature map, m and n are the position indices within the kernel, d is the dilation rate, and K is the kernel size. The effective kernel size is defined as follows:

$$K_{\text{eff}} = K + (K - 1)(d - 1). \quad (2)$$

This enables the receptive field to be expanded without increasing the number of parameters, allowing more effective capture of fine anatomical structures and small lesions.

2.3 Heatmap Attention Mechanism

As noted above, PM lesions are often difficult to distinguish from surrounding adipose tissue or inflammatory changes based on imaging features alone. Given this limitation of relying solely on imaging findings, information regarding the anatomical locations where lesions are more likely to occur can serve as a valuable additional cue for diagnosis. Motivated by this, we propose a heatmap attention mechanism (HA) that explicitly integrates the spatial distribution of lesion positions into the high-resolution feature maps obtained from dilated convolution.

Positional information of OM lesions was collected from 100 positive patients comprising 1,129 lesion-bearing slices in the training data, and a heatmap was constructed to reflect regions where lesions frequently appear. The heatmap visualizes the anatomical distribution of OM lesions in axial CT views, as illustrated in Figure 4. The resulting heatmap demonstrates that lesions are predominantly concentrated along the peritoneum, which aligns with clinical observations that peritoneal metastases frequently occur in peritoneal-adjacent regions. By integrating this spatial distribution into the heatmap attention mechanism, the model learns to focus on peritoneal regions where lesions are most likely to be present – compensating for the limited discriminative power of imaging features alone – as demonstrated in Figure 1 (c).

The generated heatmap is passed through a 1×1 convolution applied to the high-resolution feature map to adjust the number of channels, followed by a sigmoid function to produce attention scores in the range of 0–1. These scores act as spatial weights that emphasize lesion-relevant regions.

The final feature map is defined as follows:

$$F' = F + \alpha \cdot A \odot F, \quad (3)$$

where F is the original feature map, A is the attention score, $\odot$ denotes element-wise multiplication, and α is a learnable scaling parameter. The residual connection allows the model to adaptively

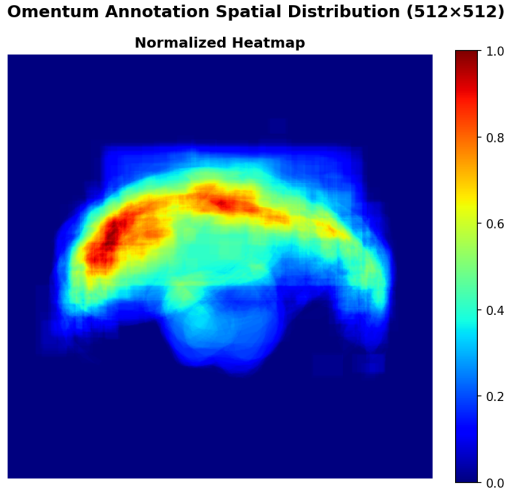


Figure 4: Axial CT view-based heatmap visualization of OM lesion distribution. The heatmap was generated by aggregating positive slices from 100 patients with peritoneal metastasis. Warmer colors (red/yellow) indicate regions with a higher frequency of OM occurrence, which are primarily concentrated along the peritoneum. This aligns with clinical observations that peritoneal metastases are frequently found along the lateral abdominal regions.

control attention intensity during training, ensuring stable optimization and improved representational flexibility.

3 Experiments

3.1 Dataset

A CT dataset for detecting OM lesions was constructed using data from patients who visited Seoul St. Mary’s Hospital. A total of 288 patients were retrospectively collected and analyzed, and the data collection and utilization process was conducted with approval from the institutional review board (IRB No. KC25RISI0813).

The training and validation sets were constructed using an 80:20 patient-level split, comprising 23,432 slices from 250 patients. Among these, positive samples consisted of 1,129 slices from 100 patients, and negative samples consisted of 22,303 slices from 150 patients. The test set was composed of completely independent data from a non-overlapping patient cohort and time period, comprising 3,241 slices from 38 patients, of which 373 slices from 19 patients were positive and 2,868 slices from 19 patients were negative. Negative slices were filtered as described in Section 2.1 to remove clinically meaningless slices, resulting in a slice imbalance ratio of approximately 1 : 19 (positive:negative).

For positive patients, only slices containing confirmed OM lesions were included, whereas for negative patients, all negative slices within the filter range were included. The same criteria were applied consistently to both the training and test sets, ensuring that the model was trained and evaluated on positive slices from positive patients and negative slices from negative patients.

3.2 Experimental Settings

To objectively evaluate the performance of the proposed model, experiments were conducted comparing against five baseline models: four convolutional neural network (CNN)-based models (ResNet-50 [6], DenseNet-121 [7], Inception-V3 [13], and MobileNetV2 [12]), and one transformer-based model (Vision Transformer Tiny (ViT-Tiny)) [4]. All baseline models were initialized with ImageNet pre-trained weights and trained under the same conditions as the proposed model, including the same optimizer, learning rate schedule, batch size, data augmentation, OMF-based preprocessing, and threshold selection strategy. The specific training configuration is described as follows.

The AdamW optimizer was used with layer-wise differential learning rates. The backbone learning rate was set to 1×10^{-5} to preserve pretrained features, while the learning rates for the classification head and attention module were set to 1×10^{-4} . The learnable scaling parameter α was configured with a learning rate of 5×10^{-4} and weight decay of 1×10^{-2} . Cosine annealing was applied as the learning rate scheduler, with early stopping (patience of 5) triggered when the validation loss did not improve for 5 consecutive epochs, and training was conducted for a maximum of 30 epochs. To address the class imbalance (approximately 1 : 19), mini-batches were constructed using WeightedRandomSampler, and focal loss ($\alpha = 0.5$, $\gamma = 2.0$) was employed as the loss function. The batch size was set to 32 and the input image resolution was 512×512 pixels. Data augmentation was applied only during training, including random horizontal flipping, rotation of ± 10 degrees, and color jitter (brightness and contrast ± 0.1).

Conventional CT-based diagnosis for OM achieves a sensitivity of only approximately 50% [10]. While higher sensitivity is desirable, increasing sensitivity inevitably reduces specificity due to the inherent trade-off between the two metrics. Therefore, a minimum sensitivity of 0.6 was set as a constraint for threshold selection, as it sufficiently exceeds conventional CT performance while maintaining clinically acceptable specificity. The optimal slice-level threshold was determined as the value that maximized the F1-score on the validation set subject to this constraint, and the same threshold was applied to the test set.

We additionally evaluated all models at the patient level. For each patient, the maximum predicted probability across all slices within the OMF range was used as the patient-level score, and a separate patient-level threshold was determined on the validation set using the same criterion and applied unchanged to the test set. The same procedure was applied to all baseline models.

4 Results

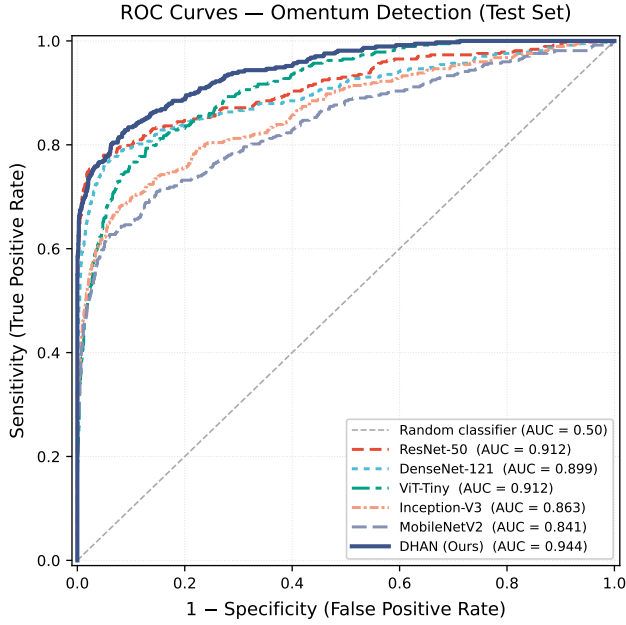
4.1 Slice-Level OM Detection Performance

To evaluate the OM detection performance of our proposed model, we compared it against several baseline algorithms. The performance metrics are presented in Table 1, where bold indicates the best performance, underline indicates the second-best performance, and improvement represents the percentage gain of the best performance over the second-best.

DHAN achieved the best performance across all metrics except specificity, recording an F1-score of 0.76 and an AUC of 0.94, outperforming the second-best model ResNet-50 by 3.48% in F1-score

Table 1: Slice-level performance comparison with baseline models.

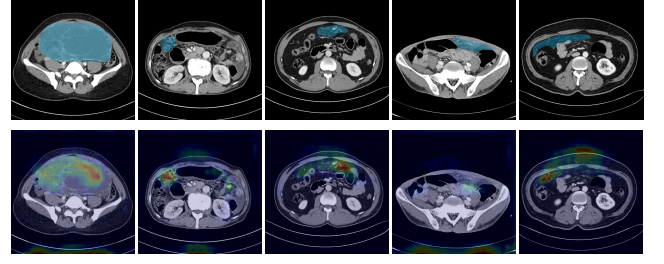
	Sensitivity	Specificity	F1 Score	AUC
ResNet-50	0.5979	<u>0.9972</u>	<u>0.7384</u>	<u>0.9214</u>
DenseNet-121	<u>0.6649</u>	0.9791	0.7283	0.8993
Inception-V3	0.4850	0.9850	0.6090	0.8532
MobileNetV2	0.4210	0.9990	0.5030	0.8406
ViT-Tiny	0.4799	0.9843	0.5997	0.9122
DHAN (Ours)	0.7426	0.9738	0.7641	0.9443
Improvement	+11.69%	-	+3.48%	+2.49%

**Figure 5: Slice-level ROC curves on the test set. DHAN (AUC = 0.94) outperforms all baseline models.**

and 2.49% in AUC. Although MobileNetV2 achieved the highest specificity of 0.99, it suffers from critically low sensitivity of 0.42, failing to detect positive lesions effectively. Similarly, Inception-V3 achieves high specificity of 0.98 but has low sensitivity of 0.48. In contrast, DHAN achieves a clinically balanced performance with sensitivity of 0.74 and specificity of 0.97, successfully avoiding both missing lesions and generating excessive false positives.

The superior balance between sensitivity and specificity achieved by DHAN is clearly illustrated in the ROC curve presented in Figure 5. The dark blue solid line representing DHAN is positioned closest to the upper-left corner, indicating optimal performance across both metrics.

Grad-CAM visualizations demonstrate that DHAN not only achieves superior quantitative performance but also provides strong interpretability. Grad-CAM activation maps highlight the image regions that contribute most to the model’s lesion detection predictions, clearly revealing the basis for the model’s decisions. Warmer

**Figure 6: Grad-CAM visualization of DHAN demonstrating model interpretability on five test cases. (Top) CT images with ground truth OM lesion locations marked in blue. (Bottom) Grad-CAM heatmaps highlighting the image regions that contribute most to the model’s lesion detection predictions.****Table 2: Slice-level ablation study results. Dilated convolution is fixed across all models, while OMF and HA are progressively added.**

Model	Sensitivity	Specificity	F1 Score	AUC
EfficientNet-B0	0.4794	0.9842	0.5027	0.8728
+ OMF	0.8097	0.9268	0.6825	0.9161
+ HA	0.6020	0.9708	0.5204	0.8945
+ OMF + HA (DHAN, Ours)	<u>0.7426</u>	<u>0.9738</u>	0.7641	0.9443
Improvement	+11.69%	-	+11.96%	+3.07%

colors (red/yellow) indicate regions showing high activation in lesion detection, while cooler colors (blue) indicate regions showing low activation. As shown in Figure 6, the regions of high activation in the Grad-CAM maps align well with the ground truth OM lesion locations, confirming that the model focuses on clinically relevant anatomical structures. This visualization, distinct from the preprocessing heatmap that identifies anatomically relevant regions, enhances clinical trustworthiness and provides practical value as a diagnostic assistance tool.

4.2 Ablation Study

To evaluate the effectiveness of each proposed component, additional experiments were conducted. Dilated convolution was maintained as a fixed component across all models, while the OMF and HA were progressively added to EfficientNet-B0 to isolate the effect of each innovation. The Ablation Study results are presented in Table 2.

DHAN achieved an F1-score of 0.76 and an AUC of 0.94, demonstrating that the combination of all proposed components yields superior performance. Compared to the second-best configuration (EfficientNet-B0 + OMF), DHAN achieved a relative improvement of 11.96% in F1-score and 3.07% in AUC.

In terms of sensitivity and specificity, DHAN demonstrates a clinically balanced performance. EfficientNet-B0 + OMF achieves the highest sensitivity of 0.81 but suffers from reduced specificity of 0.93, resulting in increased false positives.

Conversely, the baseline model (EfficientNet-B0) achieves the highest specificity of 0.98 but has critically low sensitivity of 0.48, failing to detect positive lesions effectively. DHAN ranks second

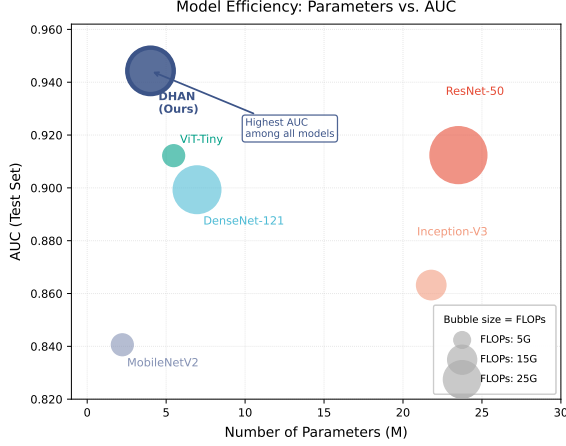


Figure 7: Model efficiency: AUC performance versus number of parameters. Bubble size represents FLOPs. DHAN achieves the highest AUC with competitive parameter efficiency.

in both metrics with a sensitivity of 0.74 and a specificity of 0.97, achieving an optimal balance that avoids missing lesions while minimizing unnecessary clinical interventions due to false positives. Therefore, DHAN achieves the best F1-score and AUC while maintaining clinically acceptable sensitivity and specificity. The ablation study validates that both OMF and HA are essential components, and their combination with dilated convolution leads to the overall performance improvement observed in DHAN.

4.3 Model Efficiency Analysis

To evaluate the parameter efficiency and computational performance of DHAN, we compared the AUC and parameter count across all baseline models. The efficiency analysis results are illustrated in Figure 7.

DHAN achieved the highest AUC of 0.94 with approximately 5 million parameters. In contrast, ResNet-50, despite approximately 25 million parameters, achieved an AUC of only 0.92. Other lightweight models such as MobileNetV2 and ViT-Tiny, despite having comparable or even fewer parameters, achieved significantly lower AUC performance (0.84 and 0.91, respectively), demonstrating the challenge of balancing parameter efficiency with diagnostic performance. DHAN uniquely achieves this optimal balance, simultaneously attaining superior AUC performance with competitive parameter efficiency. These results present a practical solution that can be deployed even in resource-constrained clinical environments without compromising diagnostic accuracy.

4.4 Patient-Level OM Detection Performance

To further assess performance at the level of clinical decision-making, we additionally evaluated all models on a per-patient basis. The results are presented in Table 3, following the same slice inclusion criterion as the slice-level setting, in which only lesion-bearing slices were used for positive patients. DHAN achieved the highest patient-level AUC of 0.875, while maintaining a balanced sensitivity of 0.737 and specificity of 0.895.

Table 3: Patient-level performance comparison with baseline models.

	Sensitivity	Specificity	F1 Score	AUC
ResNet-50	<u>0.737</u>	0.947	0.824	<u>0.859</u>
DenseNet-121	0.842	0.579	0.744	0.812
Inception-V3	0.632	0.842	0.706	0.729
MobileNetV2	<u>0.737</u>	0.632	0.700	0.751
ViT-Tiny	0.526	<u>0.895</u>	0.645	0.798
DHAN (Ours)	<u>0.737</u>	<u>0.895</u>	<u>0.800</u>	0.875

5 Conclusion

This study proposed DHAN, a lightweight deep learning model for OM detection from preoperative CT images, integrating three key components: omentum filter (OMF), dilated convolution, and heatmap attention (HA).

OMF quantified the clinically meaningful slice range (5.7% to 93.9%) through statistical analysis of 100 patients, reducing learning noise caused by irrelevant information. Dilated convolution expanded the receptive field without increasing the number of parameters, improving feature map resolution by a factor of 4 and enabling the capture of fine-grained lesion features. HA directly integrated anatomical positional information of lesions into the model, guiding it to focus on diagnostically relevant regions. These three components work synergistically to enable accurate and efficient OM detection from CT images alone, without relying on manual annotation or additional imaging modalities.

The proposed model achieved a sensitivity of 0.74, specificity of 0.97, F1-score of 0.76, and AUC of 0.94, outperforming all baseline models. Compared to ResNet-50, sensitivity improved by a relative 11.69% and AUC by 2.49%, and the ablation study confirmed the necessity of all proposed components. In particular, achieving competitive performance with approximately 5 million parameters suggests practical potential for deployment in resource-constrained clinical environments. These results demonstrate the feasibility of CT-based deep learning for OM detection and provide a foundation for extending the approach toward clinical deployment.

6 Discussion and Future Work

This study was designed as the first stage of a staged approach, establishing the model’s ability to discriminate lesion-bearing slices before extending to the full-volume setting. Accordingly, several limitations must be acknowledged to guide future development.

The training data consists exclusively of positive slices from positive patients and negative slices from negative patients. As a result, the model does not encounter negative slices within positive patients during training. In clinical practice, however, all slices from positive patients must be reviewed, including both lesion-bearing and lesion-free regions within the same patient. This discrepancy between the training distribution and clinical evaluation scenarios may affect the model’s ability to generalize to diverse anatomical contexts within positive patient cohorts.

As a preliminary investigation toward the full-volume setting, we fine-tuned the model after incorporating lesion-free slices from positive patients, and observed a decrease in performance relative

to the controlled setting. Because the model architecture and training were established under the setting that uses only lesion-bearing slices from positive patients, the subsequent addition of a large number of lesion-free slices aggravated the class imbalance and shifted the data distribution away from the one the model was optimized for. These results indicate that simple fine-tuning is insufficient and that dedicated training strategies are required to address this setting.

The present study therefore focuses on establishing feasibility under this controlled setting, and extending the evaluation to the full-volume scenario is left as the central objective of ongoing follow-up work.

Additionally, the optimal threshold identified on the validation set may require fine-tuning when applied to different patient populations with varying imaging protocols. Furthermore, since this study was conducted using data from a single institution, the generalizability of the proposed model to diverse patient populations and scanner configurations has not been fully validated. These limitations suggest that future work should focus on training strategies and external validation that better reflect real-world clinical deployment scenarios.

To address these limitations, two primary future research directions are proposed. First, explicitly incorporating negative slices from positive patients into training would allow the model to learn the overall anatomical characteristics of positive patients, enabling training under conditions more consistent with real clinical settings. In particular, applying techniques such as hard negative mining or contrastive learning to prioritize difficult samples with lesion-like features would substantially improve model robustness and generalization ability. Second, external validation on datasets from multiple institutions with varying imaging protocols is necessary to assess the generalizability of the proposed model. Validating performance across diverse patient populations and scanner configurations would be an essential step toward clinical deployment.

References

- [1] Diego Ardila, Atilla P Kiraly, Sujeeth Bharadwaj, Bokyoung Choi, Joshua J Reicher, Lily Peng, Daniel Tse, Mozziyar Etemadi, Wenxing Ye, Greg Corrado, et al. 2019. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine* 25, 6 (2019), 954–961.
- [2] Freddie Bray, Mathieu Laversanne, Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Isabelle Soerjomataram, and Ahmedin Jemal. 2024. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 74, 3 (2024), 229–263.
- [3] Shuxiang Chen, Huijuan Zhang, Hong Wei, Yongxiu Tong, and Xiaofang Chen. 2022. Practical nomogram based on comprehensive CT texture analysis to preoperatively predict peritoneal occult metastasis of gastric cancer patients. *Frontiers in Oncology* 12 (2022), 882584.
- [4] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xi-aohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
- [5] Malek Elhaddad and Sara Hamam. 2024. AI-driven clinical decision support systems: an ongoing pursuit of potential. *Cureus* 16, 4 (2024), e57728.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [7] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.
- [8] Weicai Huang, Kangneng Zhou, Yuming Jiang, Chuanli Chen, Qingyu Yuan, Zhen Han, Jingjing Xie, Shitong Yu, Zepang Sun, Yanfeng Hu, et al. 2020. Radiomics nomogram for prediction of peritoneal metastasis in patients with gastric cancer. *Frontiers in Oncology* 10 (2020), 1416.
- [9] Yuming Jiang, Xiaokun Liang, Wei Wang, Chuanli Chen, Qingyu Yuan, Xiaodong Zhang, Na Li, Hao Chen, Jiang Yu, Yaoqin Xie, et al. 2021. Noninvasive prediction of occult peritoneal metastasis in gastric cancer using deep learning. *JAMA network open* 4, 1 (2021), e2032269.
- [10] Su Jin Kim, Hyung-Ho Kim, Young Hoon Kim, Sun Hwi Hwang, Hye Seung Lee, Do Joong Park, So Yeon Kim, and Kyoung Ho Lee. 2009. Peritoneal metastasis: detection with 16-or 64-detector row CT in patients undergoing surgery for gastric cancer. *Radiology* 253, 2 (2009), 407–415.
- [11] Michela Polici, Federica Palmeri, Erica Golia, Emanuela Parri, Guido Gentiloni Silveri, Francesco Puglisi, Domenico De Santis, Marta Zerunian, Francesco Pucciarelli, Tiziano Polidori, et al. 2025. CT and MR in peritoneal malignancies: pearls and pitfalls at preoperative examination. *Insights into Imaging* 16, 1 (2025), 171.
- [12] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4510–4520.
- [13] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2818–2826.
- [14] Mingxing Tan and Quoc Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*. PMLR, 6105–6114.
- [15] Jipeng Wang, Yuannan Hu, Hao Xiong, Tiantian Song, Shuyi Wang, Haibo Xu, and Bin Xiong. 2023. CT-based deep learning model: a novel approach to the preoperative staging in patients with peritoneal metastasis. *Clinical & Experimental Metastasis* 40, 6 (2023), 493–504.
- [16] Ding Zhang, BingShu Zheng, LiuWei Xu, YiCong Wu, Chen Shen, ShanLei Bao, ZhongHua Tan, and ChunFeng Sun. 2024. A radiomics-boosted deep-learning for risk assessment of synchronous peritoneal metastasis in colorectal cancer. *Insights into Imaging* 15, 1 (2024), 150.
- [17] Jimmeng Zhang, Shilong Dai, Jing Zhou, QingSong Zhang, Mingjun Wang, and Hao Wang. 2026. The survival and peritoneal metastasis of hyperthermic intraperitoneal chemotherapy in advanced gastric cancer: a systematic review and meta-analysis of randomized controlled trials. *International Journal of Surgery* 112, 2 (2026), 5101–5110.