RESEARCH ARTICLE

Editorial Process: Submission:20/08/2025 Acceptance:17/08/2026 Published:20/08/2026

Differential Diagnosis of Ovarian Tumor Using FDG PET/CT, based on Convolutional Neural Networks

Ji Sun Park, Sun Seong Lee*

Abstract

Objective: This study aims to evaluate the utility of convolutional neural networks (CNNs) in the differential diagnosis of benign and malignant ovarian tumors using FDG PET/CT imaging. **Methods:** A total of 101 patients with pelvic masses who underwent F-18 FDG PET/CT before surgery between January 2020 and December 2023 were included. Patients were classified into benign and malignant tumor groups based on histopathological diagnosis. FDG PET/CT images were analyzed using two CNN models: a ResNet-18 model and a simple CNN architecture. The models were evaluated using accuracy, area under the curve (AUC), sensitivity, and specificity, with Gradient-weighted Class Activation Mapping (Grad-CAM) applied to visualize important regions within the images. **Results:** The ResNet-18 model achieved an accuracy of 0.882, an AUC of 0.938, sensitivity of 0.829, and specificity of 0.913. The simpler CNN model achieved an accuracy of 0.876, an AUC of 0.957, sensitivity of 0.761, and specificity of 0.942. Grad-CAM heatmaps identified regions of the PET/CT images that the models found most relevant for classification. **Conclusions:** This study demonstrates the potential of deep learning-based FDG PET/CT analysis for differentiating benign and malignant ovarian tumors. The CNN model showed high diagnostic accuracy, emphasizing the value of CNN-based models in PET-based tumor classification.

Keywords: ovarian tumor, CNN, F-18 FDG PET/CT

Asian Pac J Cancer Prev, **27 (8)**, 2847-2852

Introduction

In a clinical setting, ovarian tumors are commonly encountered and accurately distinguishing between benign and malignant ovarian tumors is very important. The differentiation is essential not only for determining the appropriate treatment plan but also for providing patients with accurate prognostic information. Malignant ovarian tumors require more aggressive management and often involve surgery, chemotherapy, targeted therapies, or other immunotherapies, while benign tumors may be managed conservatively or with minimal intervention [1-3]. Early and accurate diagnosis plays a critical role in improving patient outcomes, making the differentiation between benign and malignant ovarian tumors a key aspect of clinical practice [4].

In recent years, convolutional neural networks (CNNs) have been increasingly applied to various medical imaging fields, demonstrating promising results in areas such as radiology, pathology, and oncology. CNNs excel at identifying patterns and features in complex medical images, making them a valuable tool in the analysis of ultrasound, MRI, CT, and PET scans [5-7]. Studies have shown that CNN-based models can match or even surpass human expertise in tasks like tumor classification, lesion detection, and disease diagnosis [8-10]. Given these advancements, CNN techniques are now being explored as a powerful means to enhance the accuracy and efficiency of ovarian tumor differentiation, further improving clinical decision-making and patient outcomes [11, 12].

In FDG PET/CT, some cases can be challenging to interpret, especially when FDG uptake is intermediate, making it difficult to differentiate between malignant and benign tumors. A CNN-based model could offer valuable assistance in differential diagnosis. The application of CNN analysis using PET/CT is still relatively under-researched, and studies specifically focusing on the use of deep learning in PET/CT for diagnosing ovarian cancer are particularly scarce. The purpose of this study is to evaluate the utility of deep learning-based FDG PET/CT analysis in the differential diagnosis of ovarian tumors. By exploring this approach, we aim to assess whether PET/CT-based deep learning analysis can enhance the accuracy of ovarian cancer diagnosis.

Materials and Methods

Patients

A total of 101 patients with pelvic masses, who underwent F-18 FDG PET/CT before surgery between

*Department of Nuclear Medicine, Busan Paik Hospital, University of Inje College of Medicine, 75, Bokji-ro, Busanjin-gu, Busan, Republic of Korea. *For Correspondence: forcefullss@naver.com*

January 2020 and December 2023, were included in this retrospective study. All patients underwent resection of pelvic masses and were categorized into benign or ovarian cancer groups by histopathological diagnosis. The following patients were excluded from the study: 1) patients under 18 years of age.

The study was approved by the Institutional Review Board of Inje University Busan Paik Hospital (IRB No. 2024-10-043) and was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

FDG PET/CT imaging

All patients enrolled in the study fasted for at least 6 h before the PET/CT scan. The blood glucose levels of the patients were checked before the injection of F-18 FDG and did not exceed 200 mg/dl. A whole-body scan from head to upper thigh was acquired 60 min post intravenous injection of approximately 370 MBq of F-18 FDG. PET/CT examinations were performed using a PET/CT scanner (Discovery STE; GE Healthcare, Milwaukee, WI, USA). The CT images were obtained with a reference of 140 kV, 60–80 mA, and a section thickness of 3.75 mm. Then PET data were acquired for 2 min per bed position. PET images were reconstructed using an ordered-subset expectation maximization iterative reconstruction algorithm and then fused with the CT images.

Datasets

Out of a total of 101 patients, 34 had benign tumors, and 67 had malignant tumors. We used Python to analyze FDG PET/CT DICOM files, delineating a rectangular region of interest (ROI) on the transaxial images for each tumor. All ROIs were manually delineated by one experienced nuclear medicine physician to maintain consistency across the dataset. For malignant tumors, the ROI was drawn to include the hypermetabolic solid portion. In the case of benign tumors, the ROI was drawn on each slice around the area suspected to be the tumor. This process yielded a dataset comprising 980 labeled images for benign tumors and 1,406 labeled images for malignant tumors. The dataset was split into training and test sets. Patients included from June to December 2023 were assigned to the test dataset, while the remaining patients were allocated to the training dataset with 323 images (206 for benign tumors, 117 for malignant tumors) for test data, and 2063 images (774 for benign tumors, 1289 for malignant tumors) for training data. Data augmentation was applied to increase the diversity of the training dataset and improve model generalization. Each original image was augmented using flipping techniques: (1) horizontal flip, (2) vertical flip, and (3) both horizontal and vertical flips. This process resulted in a total of 2,322 augmented images for benign tumors and 3,867 augmented images for malignant tumors. Finally, 8,252 images (3,096 for benign tumors, 5,156 for malignant tumors) were assigned for the training data and 323 images for the test data. All images were resized to 112x112 pixels. To enhance model robustness, all PET images were normalized prior to input to the networks. Pixel intensities were rescaled to the range of 0 to 1 using min-max scaling.

CNN Model training and evaluation

Initially, we employed ResNet-18 architecture to train our model. Following this, a simpler CNN architecture was developed and trained. The model's performance was evaluated using accuracy and the area under the curve (AUC) as the primary metrics. Accuracy was used to assess the overall correctness of the model's predictions, while AUC provided a measure of the model's ability to distinguish between benign and malignant tumors, reflecting the quality of the classification. Additionally, sensitivity and specificity were considered to further evaluate the model's performance. To assess the uncertainty of model performance and compare the ResNet-18 and simple CNN models, we applied a bootstrap resampling approach on the independent test set. For each model, 1,000 bootstrap samples were generated by randomly sampling with replacement from the test set. Accuracy, AUC, sensitivity, and specificity were calculated for each bootstrap sample to obtain 95% confidence intervals. To evaluate whether the difference between the two models was statistically significant, the difference of each metric was computed for corresponding bootstrap samples, and the 95% confidence interval of the differences was examined. A difference was considered statistically significant if the confidence interval did not include zero.

The architecture of ResNet-18 was described in figure 1. The simple CNN model consists of three convolutional and fully connected layers designed for effective feature extraction and classification. The input layer accepts images of size 112×112×3. The model comprises three convolutional layers with rectified linear unit (ReLU) activation functions. The first convolutional layer applies 8 filters of size 3×3, followed by a max-pooling layer with a 2×2 window. The second convolutional layer increases the number of filters to 16 while maintaining the same kernel size and is also followed by a 2×2 max-pooling layer. The third convolutional layer further expands the feature representation using 32 filters, again with a 3×3 kernel and a subsequent 2×2 max-pooling operation. After feature extraction, the output is flattened and passed through a fully connected layer with 32 neurons and a ReLU activation function. Finally, a single-node output layer with a sigmoid activation function is used for binary classification (Figure 2). The model was trained for 10 epochs with a batch size of 32. We applied Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize which regions of the image the model assigns greater importance to.

Results

Characteristics of The Study Population

A total of 101 patients were included in this study (mean age, 54.9 ± 15.1). Among the patients, 34 had benign tumors, while 67 had ovarian cancer. The detailed histopathological results of pelvic masses are provided in Table 1.

Model Performance

The ResNet-18 model achieved a test accuracy of

Table 1. Histopathological Results of Pelvic Masses

Histopathological results	Number (%)
Epithelial ovarian cancer	67
Serous carcinoma	35 (52)
Endometrioid carcinoma	10 (15)
Mucinous carcinoma	6 (9)
Clear cell carcinoma	5 (7)
Carcinosarcoma	4 (6)
Seromucinous carcinoma	3 (5)
Other malignant tumors	4 (6)
Benign tumor	34
Endometriotic cyst	13 (38)
Mucinous cystadenoma	3 (9)
Fibrothecoma	3 (9)
Mature cystic teratoma	3 (9)
Sclerosing stromal tumor	2 (6)
Uterine myoma	2 (6)
Hemorrhagic cyst	2 (6)
Serous cystadenoma	1 (3)
Other benign tumors	5 (14)

0.882 (95% CI, 0.845–0.913), a test AUC of 0.938 (95% CI, 0.909–0.965), a sensitivity of 0.829 (95% CI, 0.758–0.893), and a specificity of 0.913 (95% CI, 0.871–0.950) on the independent test set. The simple CNN model achieved a test accuracy of 0.876 (95% CI, 0.836–0.913), a test AUC of 0.957 (95% CI, 0.936–0.975), a sensitivity of 0.761 (95% CI, 0.688–0.835), and a specificity of 0.942 (95% CI, 0.905–0.970). Bootstrap-based comparison of the two models indicated no statistically significant differences for any metric, as all p-values were greater than 0.05, suggesting comparable performance between ResNet-18 and the simpler CNN (Table 2). To further interpret the model's decision-making process, Grad-CAM was applied to visualize the regions of interest within the images. The heatmaps generated by Grad-CAM highlight the areas where the model assigns greater importance for classification (Figure 3 and 4).

Table 2. Model Performance

Metrics	ResNet-18 (95% CI)	Simple CNN (95% CI)	P-value
Accuracy	0.882 (0.845–0.913)	0.876 (0.836–0.913)	0.77
Sensitivity	0.829 (0.758–0.893)	0.761 (0.688–0.835)	0.072
Specificity	0.913 (0.871–0.950)	0.942 (0.905–0.970)	0.102
AUC	0.938 (0.909–0.965)	0.957 (0.936–0.975)	0.054

CNN, Convolutional Neural Network; AUC, Area Under the Curve; ResNet, Residual Network; CI, Confidence Interval

Discussion

In this study, we assessed the potential of CNNs for the differentiation of benign and malignant ovarian tumors using FDG PET/CT imaging. Both the ResNet-18 and simple CNN model were evaluated, and the results indicated no significant difference in diagnostic performance between the two models. Despite architectural differences, both models demonstrated comparable accuracy, AUC, sensitivity, and specificity, suggesting that the increased complexity of ResNet-18 did not confer a substantial advantage over the simpler CNN model. This observation may be attributed to the imaging characteristics of ovarian tumors in PET scans, where hypermetabolic solid components within cystic masses serve as a key indicator of malignancy. Unlike CT and MRI, which provide detailed anatomical and structural information, PET primarily visualizes metabolic activity, resulting in a relatively simpler image composition. This fundamental difference may explain the observed results, as the discriminative features of ovarian malignancies on PET images are less complex and more distinct, thereby reducing the necessity for deeper learning architectures such as ResNet-18 [13, 14].

These findings are consistent with previous studies that have demonstrated the utility of FDG PET/CT parameters,

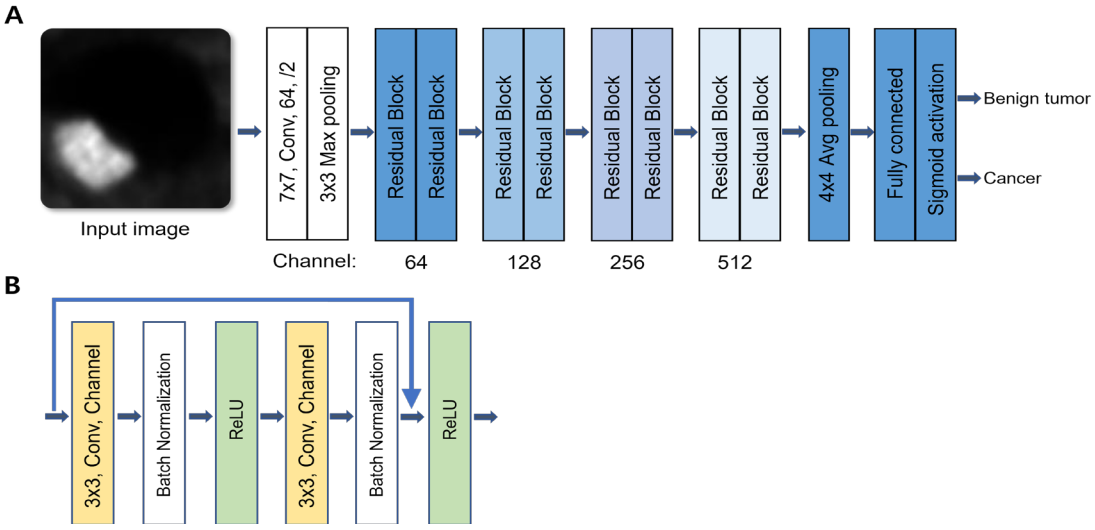

Figure 1. A, Architecture of ResNet-18. Conv, convolutional; Avg, average B, Architecture of Residual Block. ReLU, rectified linear unit

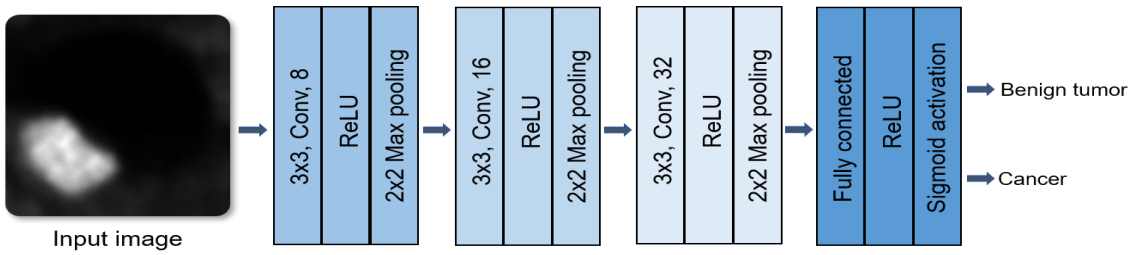

Figure 2. Architecture of Simple CNN. CNN, convolutional neural network; Conv, convolutional; ReLU, rectified linear unit

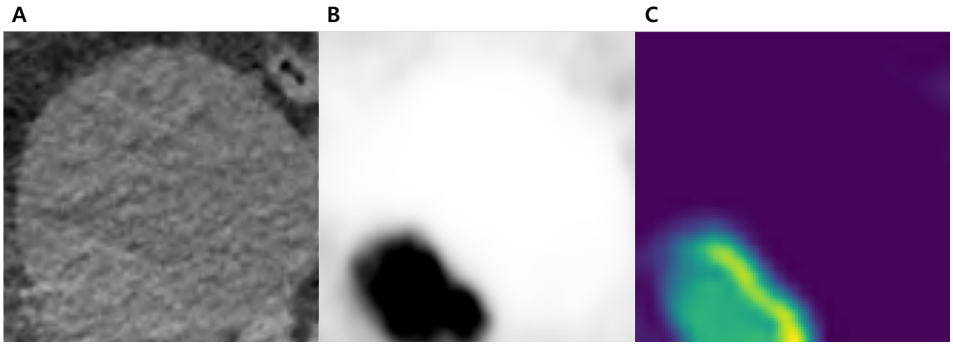

Figure 3. Grad-CAM-based Highlighting of Key Regions (True Positive Example). A, CT image showing ovarian cancer. B, PET image shows high glucose metabolism (SUVmax: 14.9) in the solid component of ovarian cancer. C, heatmap highlighting key regions for cancer detection.

such as SUV, in distinguishing malignant from benign ovarian tumors [15, 16]. Prior research focusing on texture analysis has also reported promising results; however, simpler quantitative parameters, such as SUV, have frequently shown comparable performance [17, 18]. The present study further corroborates this by demonstrating that a CNN model targeting hypermetabolic regions in PET scans can yield high diagnostic accuracy, reinforcing the importance of SUV as a critical biomarker for ovarian tumor differentiation.

Although recent studies have applied CNN-based deep learning models to ovarian cancer patients, few investigations have focused on FDG PET/CT imaging. Moreover, the most recent PET/CT-based studies included relatively small cohorts and primarily targeted recurrent cases, highlighting a clear distinction from the present study [19, 20]. In contrast, our work evaluates a larger and more diverse patient population, encompassing both benign and malignant primary ovarian tumors, thereby providing novel insights into the utility of CNN-based approaches in PET/CT imaging for ovarian tumor classification.

In addition to these findings, an examination of false positive predictions revealed that certain benign ovarian tumors, such as endometriotic cysts, exhibited mildly increased FDG uptake in their solid components. Figure 4 illustrates one such case, in which the model incorrectly predicted malignancy, which may have been associated with hypermetabolic activity in a benign lesion. This observation indicates that the CNN model's predictions may be influenced by mild FDG uptake in non-malignant solid regions, which could contribute to false positive results and a limitation in specificity. It should be noted that the occurrence of false positive predictions in this study may, in part, be attributable to the limited size of the dataset. Certain benign ovarian lesions with mildly

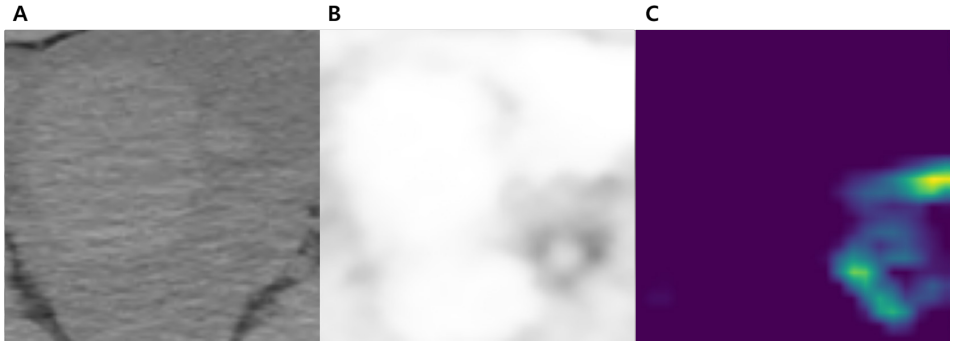

Figure 4. Grad-CAM-based Highlighting of Key Regions (False Positive Example). A, CT image showing benign ovarian tumor (endometriotic cyst). B, PET image shows mildly increased glucose metabolism (SUVmax: 4.6) in the solid portion of benign ovarian tumor. C, heatmap highlighting key regions for cancer detection.

increased FDG uptake may have been underrepresented, potentially contributing to misclassification. Future studies incorporating larger and more diverse patient cohorts are warranted to more comprehensively evaluate the model's specificity and to reduce the likelihood of false positive predictions. Expanding the dataset would allow the CNN model to better learn the distinction between benign hypermetabolic activity and true malignancy, ultimately enhancing its clinical applicability.

From a clinical perspective, integrating a CNN-based approach into radiology workflows may enhance the accuracy and efficiency of FDG PET/CT interpretation, potentially leading to faster and more reliable diagnoses. This, in turn, could optimize clinical decision-making in ovarian cancer management. Additionally, the implementation of visualization techniques such as Grad-CAM may improve model interpretability by highlighting regions of interest that contribute most to classification decisions [21, 22]. Such transparency is essential in fostering clinician confidence in AI-assisted diagnostic tools and facilitating validation of AI-derived conclusions.

A notable challenge encountered in this study was the difficulty in accurately ROI due to physiological FDG uptake in adjacent tissues, such as bowel activity, which occasionally led to inaccuracies in ROI selection. Future investigations should explore more advanced segmentation strategies, including contour-based ROI delineation, to more precisely define tumor boundaries while excluding non-tumor activity. The development of automated segmentation techniques may further enhance diagnostic efficiency by obviating the need for manual ROI selection and streamlining the analysis process.

While this study focused on tumor classification using predefined ROIs, future research should investigate a fully automated AI pipeline incorporating both tumor segmentation and classification. Such an approach would not only localize ovarian tumors but also predict their malignancy status, ultimately providing a comprehensive diagnostic framework. The development of an end-to-end AI solution represents a significant advancement in AI-assisted medical imaging and could facilitate standardized and reproducible diagnostic assessments in clinical settings.

Despite the promising findings, several limitations should be acknowledged. First, this was a single-institution study, and the model was trained and tested on a limited patient cohort. As a result, generalizability remains a concern, as performance may vary across different institutions and populations. Future studies should incorporate multi-institutional datasets to enhance model robustness and ensure broader clinical applicability. Furthermore, although data augmentation techniques were employed, the overall sample size was relatively small. Expanding the dataset with more diverse cases may improve model performance and yield more reliable diagnostic outcomes.

Another limitation is the relatively low representation of benign conditions that exhibit hypermetabolic activity on FDG PET/CT, such as inflammatory diseases and abscesses, which can mimic malignancy and lead to false-positive results. Future research should aim to include a broader spectrum of benign hypermetabolic conditions to more comprehensively evaluate the model's ability to distinguish malignant from benign lesions in real-world clinical scenarios.

Additionally, this study did not incorporate clinical parameters, such as patient history and serum biomarkers (e.g., CA125, HE4), which may provide valuable complementary information. The integration of multimodal data, combining imaging features with clinical and serological markers, could enhance diagnostic performance and improve the clinical utility of AI-based predictive models for ovarian tumors. Future studies should investigate the potential of such an approach to develop a more comprehensive and clinically relevant AI-driven diagnostic framework.

In conclusion, this study highlights the potential of deep learning-based FDG PET/CT analysis in differentiating benign and malignant ovarian tumors. A simple CNN model demonstrated high diagnostic accuracy, underscoring the utility of AI in PET-based tumor classification. Future research should focus on refining ROI delineation, exploring advanced segmentation methodologies, and validating the model on external datasets. Ultimately, the goal is to establish a fully automated AI diagnostic system capable of accurately identifying and classifying ovarian tumors, thereby providing clinicians with a robust tool to enhance ovarian cancer diagnosis and optimize patient management.

Author Contribution Statement

JS, and SS contributed to the study design, data collection, data analysis, writing of the manuscripts, and interpretation of this work.

Acknowledgements

This work was supported by a 2024 Inje University Busan Paik Hospital research grant.

Ethics Statement

The study was approved by the Institutional Review Board of Inje University Busan Paik Hospital (IRB No. 2024-10-043) and was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

Conflict of Interest

The authors declare no other potential conflicts of interest relevant to this article

References

1. Stewart C, Ralyea C, Lockwood S. Ovarian Cancer: An Integrated Review. Semin Oncol Nurs. 2019;35(2):151-6. https://doi.org/10.1016/j.soncn.2019.02.001.
2. Matulonis UA, Sood AK, Fallowfield L, Howitt BE, Sehouli J, Karlan BY. Ovarian cancer. Nat Rev Dis Primers. 2016;2(1):16061. https://doi.org/10.1038/nrdp.2016.61.
3. Ghisoni E, Morotti M, Sarivalasis A, Grimm AJ, Kandalaft L, Laniti DD, et al. Immunotherapy for ovarian cancer: Towards

a tailored immunophenotype-based approach. Nat Rev Clin Oncol. 2024;21(11):801-17. https://doi.org/10.1038/s41571-024-00937-4.
4. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, et al. Ovarian cancer statistics, 2018. CA Cancer J Clin. 2018;68(4):284-96. https://doi.org/10.3322/caac.21456.
5. Yao X, Wang X, Wang SH, Zhang YD. A comprehensive survey on convolutional neural network in medical image analysis. Multimed Tools Appl. 2022;81(29):41361-41405. https://doi.org/10.1007/s11042-020-09634-7.
6. Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. Evol Intell. 2022;15(1):1-22. https://doi.org/10.1007/s12065-020-00540-3.
7. Yu H, Yang LT, Zhang Q, Armstrong D, Deen MJ. Convolutional neural networks for medical image analysis: State-of-the-art, comparisons, improvement and perspectives. Neurocomputing. 2021;444:92-110. https://doi.org/10.1016/j.neucom.2020.04.157.
8. Haggenmüller S, Maron RC, Hekler A, Utikal JS, Barata C, Barnhill RL, et al. Skin cancer classification via convolutional neural networks: Systematic review of studies involving human experts. Eur J Cancer. 2021;156:202-16. https://doi.org/10.1016/j.ejca.2021.06.049.
9. Hughes JW, Olgin JE, Avram R, Abreau SA, Sittler T, Radia K, et al. Performance of a Convolutional Neural Network and Explainability Technique for 12-Lead Electrocardiogram Interpretation. JAMA Cardiol. 2021;6(11):1285-1295. https://doi.org/10.1001/jamacardio.2021.2746.
10. Abdou MA. Literature review: efficient deep neural networks techniques for medical image analysis. Neural Comput Appl. 2022;34(8):5791-5812. https://doi.org/10.1007/s00521-022-06960-9.
11. Jung Y, Kim T, Han MR, Kim S, Kim G, Lee S, et al. Ovarian tumor diagnosis using deep convolutional neural networks and a denoising convolutional autoencoder. Sci Rep. 2022;12(1):17024. https://doi.org/10.1038/s41598-022-20653-2.
12. Saida T, Mori K, Hoshiai S, Sakai M, Uruchibara A, Ishiguro T, et al. Diagnosing Ovarian Cancer on MRI: A Preliminary Study Comparing Deep Learning and Radiologist Assessments. Cancers. 2022;14(4):987. https://doi.org/10.3390/cancers14040987
13. Castilla T, Martínez MS, Leguía M, Larrabide I, Orlando JI. A ResNet is all you need: modeling a strong baseline for detecting referable diabetic retinopathy in fundus images. In:18th International Symposium on Medical Information Processing and Analysis 2023 Mar 6 (Vol. 12567, pp. 212-221). SPIE.
14. Khan MA, Auvee RB. Comparative analysis of resource-efficient CNN architectures for brain tumor classification. In: 2024 27th International Conference on Computer and Information Technology (ICCIT) 2024 Dec 20 (pp. 639-644). IEEE.
15. Tanizaki Y, Kobayashi A, Shiro M, Ota N, Takano R, Mabuchi Y, et al. Diagnostic Value of Preoperative SUVmax on FDG-PET/CT for the Detection of Ovarian Cancer. Int J Gynecol Cancer. 2014;24(3):454-60. https://doi.org/10.1097/IGC.0000000000000074.
16. Kitajima K, Suzuki K, Senda M, Kita M, Nakamoto Y, Onishi Y, et al. FDG-PET/CT for diagnosis of primary ovarian cancer. Nucl Med Commun. 2011;32(7):549-553. https://doi.org/10.1097/MNM.0b013e328345b339.
17. Wang T, Sun H, Guo Y, Zou L. 18F-FDG PET/CT Quantitative Parameters and Texture Analysis Effectively Differentiate Endometrial Precancerous Lesion and Early-Stage Carcinoma. Mol Imaging. 2019;18:1536012119856965. https://doi.org/10.1177/1536012119856965.
18. Shin SY, Hong IK, Jo YS. Quantitative computed tomography texture analysis: can it improve diagnostic accuracy to differentiate malignant lymph nodes? Cancer Imaging. 2019;19(1):25. https://doi.org/10.1186/s40644-019-0214-8.
19. Sadeghi MH, Sina S, Alavi M, Giammarile F. The OCDA-Net: a 3D convolutional neural network-based system for classification and staging of ovarian cancer patients using F-18 FDG PET/CT examinations. Ann Nucl Med. 2023;37(12):645-54. https://doi.org/10.1007/s12149-023-01867-4.
20. Sadeghi MH, Sina S, Alavi M, Giammarile F, Feshani N, Farshchitabrizi AH, et al. Longitudinal deep learning models for tracking disease progression in ovarian cancer using PET/CT imaging and clinical reports. Phys Eng Sci Med. 2025. https://doi.org/10.1007/s13246-025-01669-0.
21. Chien JC, Lee JD, Hu CS, Wu CT. The Usefulness of Gradient-Weighted CAM in Assisting Medical Diagnoses. Appl Sci. 2022;12(15):7748. https://doi.org/10.3390/app12157748
22. Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Bhardwaj P, Singh V. A deep learning and Grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. Chaos Solitons Fract. 2020;140:110190. https://doi.org/10.1016/j.chaos.2020.110190.