European Journal of Radiology Open 14 (2025) 100624

Contents lists available at ScienceDirect

European Journal of Radiology Open

journal homepage: www.elsevier.com/locate/ejro

Preliminary study on detection and diagnosis of focal liver lesions based on a deep learning model using multimodal PET/CT images

Yingqi Luo [a,1], Qingqi Yang [b,1], Jinglang Hu [c], Xiaowen Qin [a], Shengnan Jiang [a], Ying Liu [a,*]

[a] Department of Nuclear medicine, The Second Affiliated Hospital, Guangzhou Medical University, Guangzhou, China
[b] Department of Vascular Surgery, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, China
[c] School of Medicine, Sun Yat-Sen University, Guangzhou, China

ARTICLE INFO

Keywords:
Deep learning
Focal liver lesion
PET/CT

ABSTRACT

Objectives: To develop and validate a deep learning model using multimodal PET/CT imaging for detecting and classifying focal liver lesions (FLL).
Methods: This study included 185 patients who underwent ^{18}F-FDG PET/CT imaging at our institution from March 2022 to February 2023. We analyzed serological data and imaging. Liver lesions were segmented on PET and CT, serving as the "reference standard". Deep learning models were trained using PET and CT images to generate predicted segmentations and classify lesion nature. Model performance was evaluated by comparing the predicted segmentations with the reference segmentations, using metrics such as Dice, Precision, Recall, F1-score, ROC, and AUC, and compared it with physician diagnoses.
Results: This study finally included 150 patients, comprising 46 patients with benign liver nodules, 51 patients with malignant liver nodules, and 53 patients with no FLLs. Significant differences were observed among groups for age, AST, ALP, GGT, AFP, CA19–9and CEA. On the validation set, the Dice coefficient of the model was 0.740. For the normal group, the recall was 0.918, precision was 0.904, F1-score was 0.909, and AUC was 0.976. For the benign group, the recall was 0.869, precision was 0.862, F1-score was 0.863, and AUC was 0.928. For the malignant group, the recall was 0.858, precision was 0.914, F1-score was 0.883, and AUC was 0.979. The model's overall diagnostic performance was between that of junior and senior physician.
Conclusion: This deep learning model demonstrated high sensitivity in detecting FLLs and effectively differentiated between benign and malignant lesions.

1. Introduction

Liver cancer is the third leading cause of cancer-related deaths globally. According to reports, in 2020, the global estimated incidence of liver cancer was approximately 905,700 cases. It is projected that the annual number of new liver cancer cases will increase by 55.0 % between 2020 and 2040 [1,2]. By 2025, the global incidence of liver cancer is expected to exceed one million [3], indicating that liver cancer remains a global health challenge.In terms of diagnosis, it is crucial to differentiate liver cancer from other focal liver lesions (FLL) such as hemangiomas, focal nodular hyperplasia, and liver cysts. Early detection and accurate diagnosis of FLL are essential for treatment and prognosis. Liver biopsy is considered the gold standard for determining the nature of FLL. However, this method has drawbacks including invasiveness, risks of procedure-related complications, and sampling errors [4]. In recent years, imaging techniques such as ultrasound and CT have been widely used in the diagnosis of FLL due to their non-invasive nature and ease of use [5].

Among various imaging modalities, PET/CT has unique value in tumor detection. 18F-fluorodeoxyglucose (^{18}F-FDG) PET/CT is a molecular imaging technique that exhibits high sensitivity and specificity in tumor detection. CT provides anatomical localization information, while ^{18}F-FDG PET provides information on lesion glucose metabolism, reflecting functional and metabolic characteristics [6,7]. However, the normal liver parenchyma exhibits a certain degree of ^{18}F-FDG uptake, which can make it challenging to distinguish between physiological uptake and pathological lesions, particularly when the lesions are not well-defined. Additionally, although most malignant tumors show high

* Correspondence to: No. 250 Changgang East Road, Haizhu District, Guangzhou City, Guangdong Province, China.
E-mail address: liuying999901@126.com (Y. Liu).
[1] These authors contributed equally to this study, and shared co-first authorship.

https://doi.org/10.1016/j.ejro.2024.100624
Received 17 June 2024; Received in revised form 23 October 2024; Accepted 11 December 2024

^{18}F-FDG uptake, some tumors may exhibit negative uptake, while benign lesions may exhibit positive uptake [8]. The diagnosis based solely on visual analysis is prone to missed or misdiagnosis, as it can be subjectively influenced by the clinical experience and expertise of the physician [9].

In recent years, there have been significant advancements in the application of artificial intelligence (AI) in the field of medicine, with applications in disease prediction, diagnosis, prognosis, and treatment response assessment [10]. Deep learning (DL), a subfield of AI, can automatically extract and learn data features from complex nonlinear processes based on neural network structures. It has been widely applied in the field of clinical medical imaging. DL can automatically extract features from images for analysis, enabling localization, segmentation, and classification of image targets, and demonstrating accuracy that is on par with or even exceeds human performance [11,12].

There have been studies on the application of DL in the liver [13–15]. Wang et al. [16] utilized ResNet18 to construct a multimodal model incorporating both CT and MRI images, demonstrating strong capabilities in preoperatively predicting microvascular invasion of hepatocellular carcinoma. The validated multimodal model achieved an AUC of 0.819, outperforming the single-modality models using CT or MRI alone. Lai et al. [17] employed DL to extract and integrate ^{18}F-FDG PET/CT images of patients to predict overall survival in liver cancer patients before liver transplantation. The study found that the model based on PET/CT images achieved a sensitivity of 0.571, exhibiting better performance than the model based solely on CT images.

Currently, there is a lack of relevant reports on the application of DL in the detection and classification of liver lesions through the fusion of multimodal PET/CT images. Therefore, this study proposes a DL model based on multimodal PET/CT images to evaluate its performance in the detection and diagnosis of FLL.

2. Methods

2.1. Patients

This retrospective study was approved by the Institutional Review Board, and the requirement for informed consent was waived. All procedures were performed in accordance with the relevant guidelines and regulations. Collect the patient's medical history and serological data for analysis, including whether there is a history of hepatitis, smoking, alcohol consumption, and diabetes mellitus. Measure the levels of tumor markers such as alpha-fetoprotein (AFP), Carbohydrate antigen 19–9 (CA19–9), and Carcinoembryonic Antigen (CEA). As well as liver function parameters including serum albumin (ALB), alanine aminotransferase (ALT), aspartate transaminase (AST), alkaline phosphatase (ALP), gamma-glutamyl transferase (GGT), and total bilirubin (TB). The study utilized a PET/CT scanner (Discovery 710, GE) to acquire whole-body images. Patients fasted for at least 6 hours prior to the scan and underwent ^{18}F-FDG administration, with imaging obtained approximately 60 minutes after tracer injection.

As shown in Fig. 1, this study included 185 patients who underwent whole-body ^{18}F-FDG PET/CT imaging at our hospital from March 2022 to February 2023. The data were anonymized. Patients who met the following criteria were included in the benign lesion group: the presence of at least one FLL in the images, diagnosed as a benign nodule based on pathological analysis or a combination of typical imaging features and clinical data. Patients who met the following criteria were included in the malignant lesion group: the presence of at least one FLL in the images, diagnosed as a malignant nodule based on pathological analysis or a combination of typical imaging features and clinical data. Patients who met the following criteria were included in the normal group: no FLL observed in the images. Exclusion criteria were as follows: patients who received relevant treatments for liver lesions prior to the examination, including transcatheter arterial chemoembolization, surgery, radiofrequency ablation, systemic chemotherapy, etc.; patients with a history of liver resection or liver transplantation; poor image quality. Based on these criteria, a total of 46 patients with benign liver nodules, 51 patients with malignant liver nodules, and 53 patients without FLLs were included in the study. The study framework is shown in Fig. 2.

2.2. Lesion segmentation

The CT and PET images were processed in 3D Slicer (https://www.slicer.org) in DICOM format, with the imaging range from the diaphragm level to the pubic symphysis level. Each PET slice was 256 × 256 pixels in size, and each CT slice was 512 × 512 pixels in size. The region of the liver lesions in each CT and PET image layer was

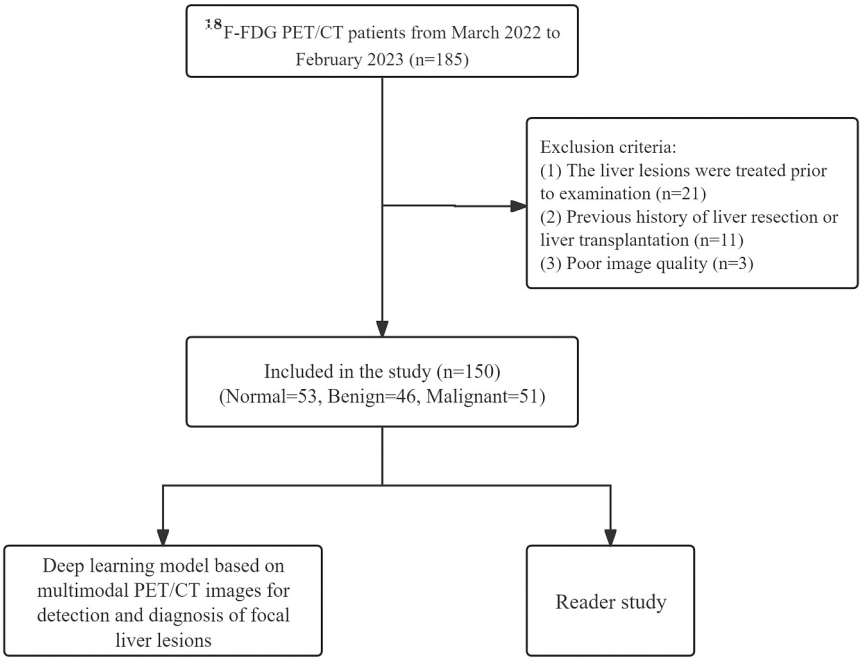

Fig. 1. A total of 185 patients who underwent ^{18}F-FDG PET/CT were studied, 150 of whom were eventually included in the analysis.

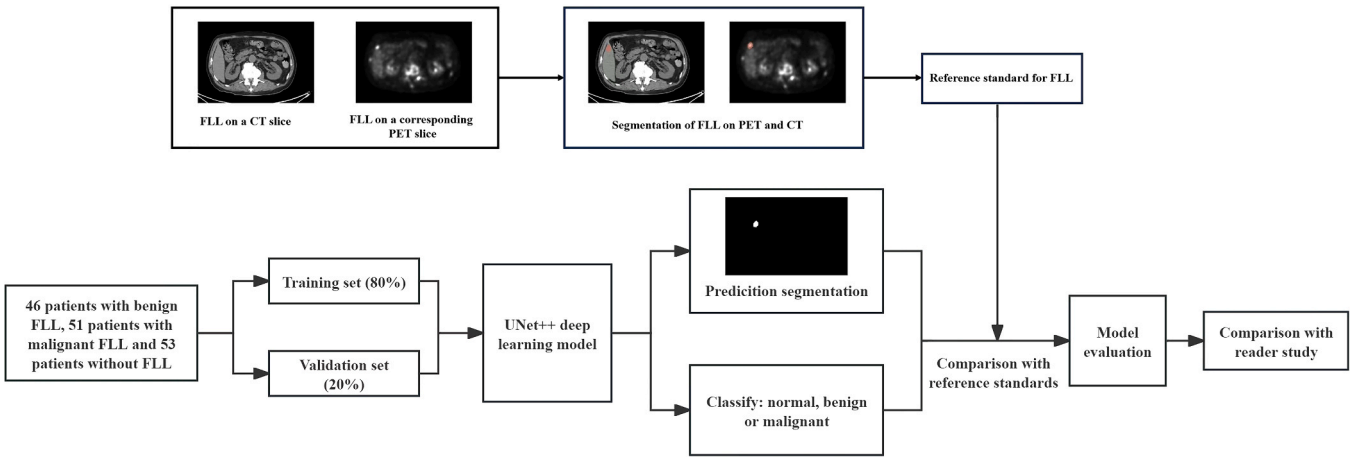

Fig. 2. The framework of the study. Abbreviation: FLL, Focal Liver Lesion.

segmented, excluding large blood vessels and bile ducts, and defined as the "reference standard" for the lesions. This process was completed by a resident physician and reviewed by two nuclear medicine physicians with over 20 years of experience.

2.3. Deep learning model training

The architecture of the DL model is shown in Fig. 3. When using multimodal data for segmentation or detection tasks, a common approach is to use multiple encoder branches to extract features separately from each modality. In this method, PET and CT images cannot benefit from each other during the feature extraction process. Therefore, this study propose the Shared Down Block (SDB) module to help the encoder branches improve the effectiveness of feature extraction. The model utilized the SDB to map the PET-CT feature maps to the same spatial Region of Interest (ROI), which can maintain spatial and structural consistency. The SDB module performs segmentation on the input feature maps and then applies separate downsampling to each branch, which not only preserves more spatial information but also enhances the model's ability to learn features from different modalities. This approach allows the model to perform feature extraction and fusion for both CT and PET images simultaneously when processing these multimodal images. Through this module, the CT features can help remove some false positives in the PET feature maps, while the PET features can help detect lesions that are difficult to identify in the CT.

The encoder branches have independent encoder blocks and the same SDB for PET-CT feature extraction. The model first reduce the size of the feature maps through 2×2 convolution, and then reduce the number of filters through 1×1 convolution, which also promotes cross PET-CT channel information interaction and improves the model's nonlinear expression. During the training process, the weights of the SDB are updated twice per epoch. When the input to the SDB is from PET, the output of the SDB can contain some corresponding CT information, and vice versa. This also means that the output of each SDB can

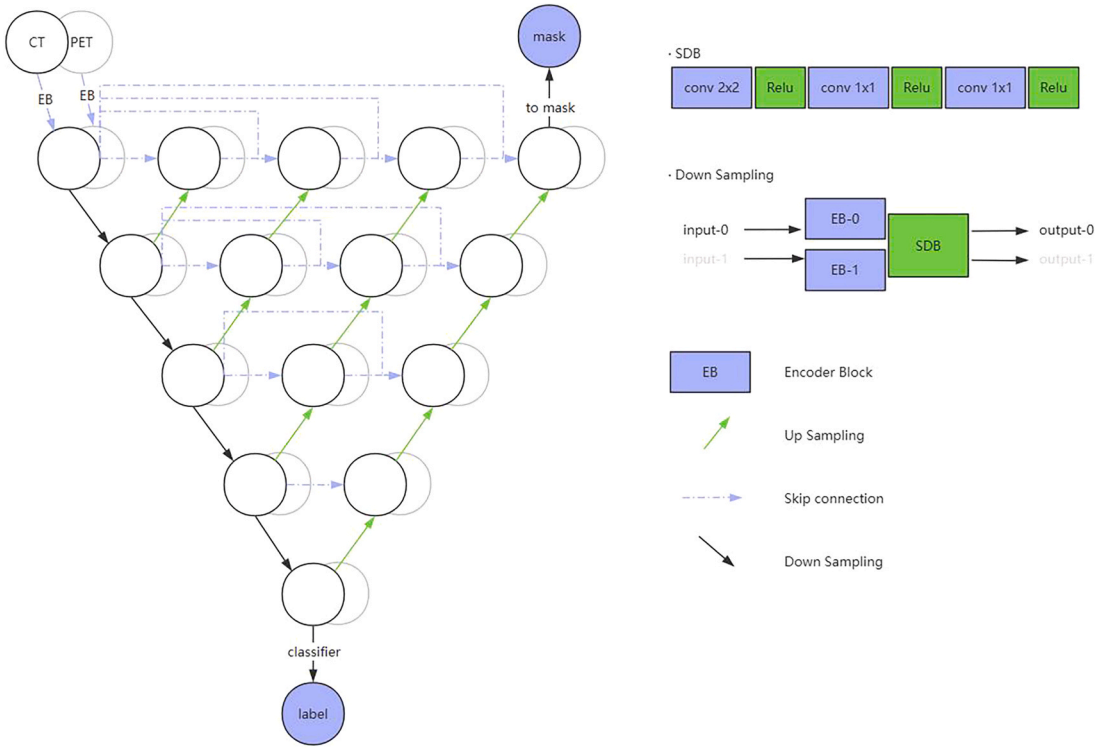

Fig. 3. The architecture of the deep learning model. Abbreviation: SDB, Shared Down Block.

represent information from both modalities (PET and CT).

In the expansive path of UNet++, the nested skip connections further promote the combination of multi-scale features from the contracting path with the upsampled features. This multilevel feature fusion strategy enables the model to not only capture the detailed information in the images but also understand the contextual information, thereby providing more accurate predictions during the segmentation process. In this way, UNet++ can effectively handle noise and complex structures in medical images, particularly when dealing with highly heterogeneous lesion regions. Through end-to-end training, UNet++ can learn to extract fine-grained segmentation maps from the input multimodal medical images. During the training process, the model is optimized by minimizing a combination of segmentation loss and classification loss, which allows the model to not only perform accurate segmentation but also identify different pathological states in the multimodal data.

To add classification prediction capabilities on top of the UNet++ architecture, the model embedded classification heads at one or more selected levels within the expansive pathway. These levels are typically located in the middle part of the network in order to capture rich image features. For each selected level, we extract global information from the feature maps using global average pooling or global max pooling, forming a fixed-length feature vector. These feature vectors are then concatenated and processed through a series of fully connected layers to learn feature representations suitable for the classification task. Finally, an output layer with size corresponding to the number of classes in the classification task is used, and the results are output using a Softmax activation function. To reduce overfitting, this study employed K-Fold cross validation, where all the data were divided into training and validation sets at a 4:1 ratio to conduct the training.

2.4. Reader study

Two nuclear medicine physicians (a junior physician with 3 years of experience and a senior physician with 20 years of experience) did not participate in the model development or data processing. They were provided with imaging data consistent with the model, without knowledge of any patient information, to assess the nature of the findings. This allowed for a comparison of the performance between the model and the physicians in classifying FLLs.

2.5. Evaluation

In this study, the Dice similarity coefficient (Dice) was used to evaluate the performance of the model in detecting lesions. It measures the similarity between the pre-segmented ground truth mask and the predicted mask. The performance of the model in lesion classification was assessed using precision, recall, F1-score, receiver operating characteristic curve (ROC) and the area under the receiver operating characteristic curve (AUC). Based on the reference values obtained from expert annotations, the predictions were classified as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The formulas for calculating precision, recall, and F1-score are as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN}, \text{F1} = \frac{2 * Precision * Recall}{Precision + Recall}$$

The diagnostic performance of physicians and the model was evaluated using sensitivity, specificity, and accuracy. Sensitivity is equal to recall as defined above, and specificity and accuracy are calculated as follows:

$$\text{Specificity} = \frac{TN}{TN + FP}, \text{Accuracy} = \frac{TP + TN}{TP + \text{FP} + \text{TN} + FN}$$

2.6. Statistical analysis

The statistical analysis was performed using IBM SPSS Statistics 25.0 software (IBM Corp., Armonk, NY, USA). The Shapiro-Wilk test was used to assess the normality of the distribution. For data with a normal distribution, the Student's *t*-test was used for analysis. Otherwise, the Mann-Whitney *U* test was applied. p-value less than 0.05 is considered to indicate a significant difference.

3. Results

3.1. Patient clinical characteristics

This study ultimately included 150 patients (mean [SD] age, 56.7 [13.2] years; 68 males [45.3 %]). Patients were categorized into three groups based on pathology and typical imaging features: benign, malignancy, and normal. Table 1 presents the clinical characteristics of the patients. Significant differences were observed among the three groups in terms of age, AST, ALP, GGT, AFP, CA19–9, and CEA.

3.2. Evaluation of deep learning model performance

Fig. 4 displays the ROC for the validation set. Table 2 displays the performance of the model in the five-fold cross-validation. The Dice coefficient for the training set was 0.760 (95 % CI, 0.717–0.803), and for the validation set, it was 0.740 (95 % CI, 0.706–0.774). In the validation set, the recall for the normal group was 0.918 (0.865–0.972), precision was 0.904 (0.826–0.981), F1 score was 0.909 (0.885–0.933), and AUC was 0.976 (0.969–0.983). For the benign group in the validation set, the recall was 0.869 (0.798–0.940), precision was 0.862 (0.801–0.922), F1 score was 0.863 (0.837–0.889), and AUC was 0.928 (0.894–0.963). For the malignant group in the validation set, the recall was 0.858 (0.789–0.927), precision was 0.914 (0.870–0.958), F1 score was 0.883 (0.863–0.904), and AUC was 0.979 (0.972–0.986).

Fig. 5 shows the case of lesions detected by the model, where A-E, F-J and K-O are images of normal, benign and malignant patients respectively. It includes CT images (A, F, K), PET images (B, G, L),

Table 1
Patient clinical characteristics.

Characteristics	Benign (n = 46)	Malignant (n = 51)	Normal (n = 53)	p-value
Male(n[%])	26 (56.5)	25 (49.0)	17 (32.1)	
Hepatitis(n[%])	3 (6.5)	8 (15.7)	1 (1.9)	
Smoking history (n[%])	9 (19.6)	5 (9.8)	3 (5.7)	
Alcohol history(n [%])	2 (4.3)	5 (9.8)	1 (1.9)	
Diabetes	2 (27.2)	12 (23.5)	2 (3.8)	
Age,years	57.0 ± 15.5	61.8 ± 11.5	47.0 ± 8.3	0.001*
ALB, g/L	38.3 ± 4.9	37.7 ± 4.8	37.2 ± 5.6	0.307
ALT, U/L	14.0 (10.0–25.0)	20.0 (11.0–32.0)	19.9 (13.3–37.3)	0.239
AST, U/L	29.0 (19.0–51.0)	22.5 (15.0–32.5)	22.5 (15–32.5)	0.001*
ALP, U/L	63.0 (52.0–89.0)	110.0 (73.0–159.0)	68 (51.5–84.5)	<0.001*
GGT, U/L	23.0 (15.0–32.7)	51.0 (26.0–146.0)	28 (20.5–60.3)	0.001*
TB, umol/L	8.1 (5.0–14.2)	11.3 (8.0–16.8)	7.7 (5.4–12.0)	0.010
AFP, ug/L	1.9 (1.4–3.5)	2.9 (2.2–4.8)	2.2 (1.5–3.3)	0.005*
CA19–9, U/ml	8.2 (2.6–19.6)	18.5 (3.9–477.7)	9.4 (3.6–17.8)	0.017*
CEA, ug/L	2.3 (1.6–4.8)	4.2 (2.1–14.6)	2.3 (0.9–4.1)	0.001*

ALB, Serum Albumin; ALT, Alanine Aminotransferase; AST, Aspartate Transaminase; ALP, Alkaline Phosphatase; GGT, γ-Glutamyl Transpeptadase; TB, Total Bilirubin; AFP, Alpha-Fetoprotein; CA19–9, Carbohydrate antigen 19–9; CEA, Carcinoembryonic Antigen; *p < 0.05 indicates a statistical significance.

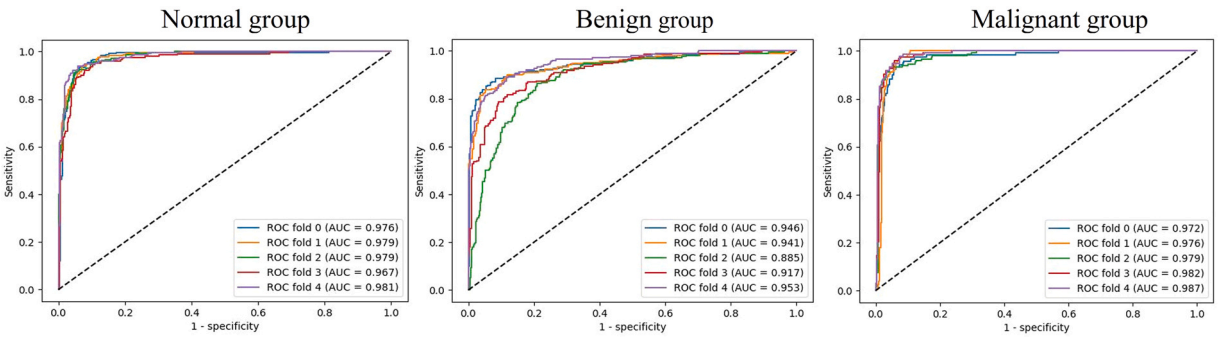

Fig. 4. The ROC of the model in the validation set for the normal, benign, and malignant groups are shown. The five curves, labeled as Fold 0–4, represent the results of the five rounds of validation.

Table 2
Results of model performance.

	Recall(95 % CI)	Precision (95 %CI)	F1(95 %CI)	AUC(95 %CI)
Normal training group	0.944 (0.905–0.983)	0.934 (0.894–0.975)	0.938 (0.926–0.951)	0.987 (0.984–0.990)
Normal validation group	0.918 (0.865–0.972)	0.904 (0.826–0.981)	0.909 (0.885–0.933)	0.976 (0.969–0.983)
Benign training group	0.910 (0.880–0.941)	0.902 (0.859–0.945)	0.905 (0.886–0.925)	0.930 (0.891–0.970)
Benign validation group	0.869 (0.798–0.940)	0.862 (0.801–0.922)	0.863 (0.837–0.889)	0.928 (0.894–0.963)
Malignant training group	0.897 (0.847–0.947)	0.933 (0.897–0.970)	0.914 (0.887–0.941)	0.985 (0.979–0.991)
Malignant validation group	0.858 (0.789–0.927)	0.914 (0.870–0.958)	0.883 (0.863–0.904)	0.979 (0.972–0.986)

segmentation of the lesion on CT images (C, H, M), segmentation of the lesion on PET images (D, I, N), and lesions predicted by the deep learning model (E, J, O). The model demonstrated excellent performance in the identification and segmentation of the target regions, exhibiting both superior segmentation quality and a high degree of accuracy.

3.3. Reader study

We compared the detection performance of the model and physicians. The model had an average accuracy of 0.936, with a sensitivity of 0.918 and specificity of 0.671 for normal group; a sensitivity of 0.869 and specificity of 0.989 for benign lesions; and a sensitivity of 0.858 and specificity of 0.984 for malignant lesions. The junior physician had an average accuracy of 0.909, with sensitivities and specificities of 0.906 and 0.887 for normal group, 0.848 and 0.962 for benign lesions, and 0.843 and 0.939 for malignant lesions. The senior physician achieved an average accuracy of 0.944, with sensitivities and specificities of 0.925 and 0.918 for normal group, 0.891 and 0.981 for benign lesions, and 0.922 and 0.980 for malignant lesions (Table S1).

4. Discussion

FLLs are a common clinical issue, and the benignity or malignancy of the lesions can impact the treatment strategy [5]. Among various imaging modalities, PET/CT has unique advantages in the diagnosis of FLLs. ^{18}F-FDG PET/CT is an advanced molecular imaging technique that can detect cellular metabolic activity, which helps to accurately delineate the lesion boundaries and better differentiate the benign or malignant nature of the lesions [8,18]. Deep learning models have already demonstrated excellent performance in the diagnosis and prognosis of various diseases [19]. Therefore, to detect FLLs and distinguish their benign or malignant nature, this study proposes a multimodal model based on PET/CT images.

The model demonstrated excellent performance in the detection of liver lesions. The high Dice coefficient indicates that the predicted segmentation mask from the model trained on PET/CT images closely matches the pre-annotated ground truth mask, suggesting the model can accurately identify liver lesions in PET/CT images. The results show the model achieved the highest classification performance for malignant lesions, followed by normal images, and lastly benign lesions. This is likely because malignant tumors have increased metabolic activity and glucose uptake, resulting in ^{18}F-FDG accumulation on PET, as well as

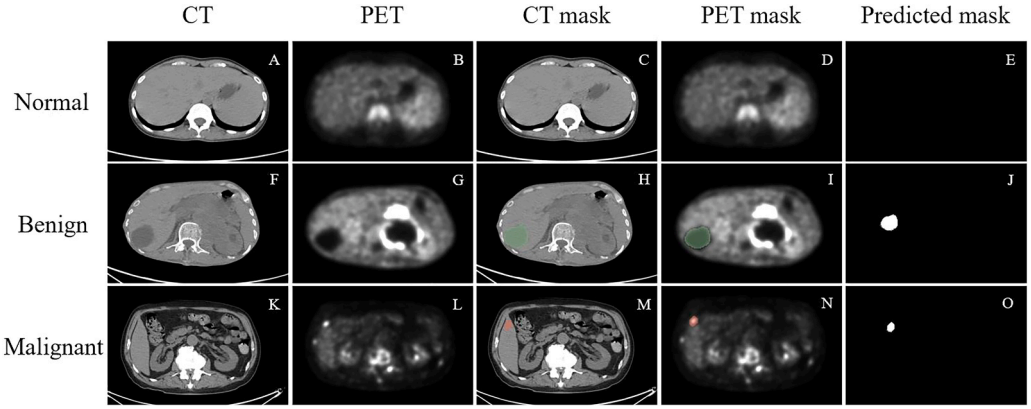

Fig. 5. Examples of the model detecting normal, benign, and malignant lesions.

structural changes in the tissue during tumor growth, such as lesion shape and margins, which allows the model to recognize the characteristic features of malignant lesions on CT [20]. While the detection rate for benign lesions was slightly lower than normal, the model still exhibited excellent performance. This may because when the liver has infectious or other lesions, the liver parenchymal density may change, making it difficult to differentiate benign lesions on non-contrast CT [21]; some benign liver lesions can also have varying degrees of elevated glucose metabolism [8], which could be misinterpreted by the model as malignant.

To clinically evaluate the model, we compared its diagnostic performance with two physicians. The model's overall diagnostic performance fell between the junior and senior physicians, and it even demonstrated slightly higher specificity for benign and malignant lesions than the senior physician. This indicates the model's significant potential for classifying FLLs in actual clinical practice. In future clinical applications, the model will provide diagnostic support for junior physicians in their learning phase. Additionally, its use is expected to reduce the workload of physicians, enabling them to concentrate more on complex cases, thereby enhancing the overall quality of healthcare services.

The UNet+ + model used in this study is a deep learning architecture for image segmentation tasks. By introducing nested skip connections and SDB modules, it significantly enhances the model's capability to process complex images. UNet+ + is an improvement and extension of the classical UNet model, fusing features at multiple levels to strengthen the model's ability to capture multi-scale contextual information [22, 23]. Through this approach, we were able to create a deep learning model that not only performs precise segmentation on medical images but also provides accurate classification predictions. In addition to outputting the segmentation masks, the model also predicted a mask type after multiple downsampling steps of the CT and PET images. This multi-task learning approach not only improves the model's efficiency but also enhances its overall performance through cross task synergies, making it well-suited for complex medical image analysis tasks that require simultaneous segmentation and classification.

Overfitting is a significant issue faced by deep learning models, particularly when training data is sourced from a single center, which limits the model's generalization ability. To enhance the model's applicability, we employed data augmentation techniques. In this study, we applied random noise to some training images to simulate potential noise introduced during the image acquisition process. The noise levels were randomly generated to prevent the model from becoming overly reliant on specific noise patterns, thereby improving its robustness against noise interference. Additionally, we applied random scaling to each image in the training set, with scaling factors ranging from 0.8 to 1.2. This operation simulates the varying sizes of targets that may occur in the real world, enhancing the model's adaptability to changes in image dimensions. To improve the model's ability to recognize horizontally flipped images, we performed horizontal flipping on the training images. This approach enables the model to learn consistent features even after flipping, thereby increasing its invariance to horizontal transformations. We also employed random cropping techniques to augment the dataset. By randomly selecting a portion of the image and using it as a new image, we simulated local occlusions or cropping scenarios that may be encountered in practical scenarios. The size and position of the cropping areas were randomly determined to ensure that the model does not become overly reliant on specific image regions, thereby enhancing its ability to recognize local features. To further enhance the model's generalization ability, we applied perspective transformation to the training images, simulating different viewpoints and depth effects by altering the image perspective. By combining these data augmentation techniques, we effectively expanded the training set and improved the model's adaptability to various changes, resulting in better performance.

Male gender, middle-age, hepatitis, diabetes, obesity, and alcohol consumption are important risk factors for the development of liver cancer. Tumor markers and other liver function serum markers can provide adjunct diagnostic information for liver cancer [3,24]. Consistent with the results of this study, age, AST, ALP, GGT, AFP, CA19–9 and CEA showed significant differences among liver lesions. However, ALB, ALT, and TB as liver function indicators did not show significant differences in the study results. This may because some patients were in the early stage of liver cancer, where the tumor had not yet significantly impacted liver function.

This study has the following limitations. First, this is a single-center retrospective study, lacking diversity in the study information, which may constrain the broad applicability of the model. Second, due to the limited sample size, the model was only capable of distinguishing between benign and malignant FLLs, without further subclassification of specific pathological types. Most benign lesions do not require subsequent surgical treatment, making it challenging to conduct histopathological evaluations for all lesions. However, benignity or malignancy can often be assessed through typical imaging features or continuous follow-up. To ensure the accuracy of the results, our data were labeled based on the analysis of all available imaging and clinical information by two physicians. As this preliminary study focused on ensuring the accuracy of the model's outputs, specific pathological types were not differentiated. In the future, we aim to expand the sample size to gather more data, enabling a more detailed classification of pathological subtypes for better application in clinical practice. Third, the segmentation process of the lesions currently relies on manual annotations and physician reviews. Despite verification and adjustments, human errors cannot be completely eliminated, which affects the training and performance of the model. Future research should focus on developing deep learning models capable of automatic image segmentation, incorporating unsupervised or self-supervised learning techniques to reduce reliance on manual input. Fourth, this model is currently in the preliminary research stage and has not yet been applied in clinical practice. Future studies should further evaluate the diagnostic performance of physicians in conjunction with the model to clarify its value in assisting with clinical diagnoses.,Fifth, deep learning models have inherent limitations, such as sample size, overfitting, selection bias, and lack of interpretability. In future studies, it will be crucial to increase the sample size, improve data utilization efficiency, and continuously refine the learning process.

5. Conclusion

In summary, this study proposes a DL model based on multimodal PET/CT images for the detection and diagnosis of FLL, which demonstrates excellent performance in identifying FLL and determining its benign or malignant nature.

Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

Yingqi Luo: study design, drafting and revising the manuscript
Qingqi Yang: training model, image interpretation and processing
Jinglang Hu: data analysis and training model
Xiaowen Qin: acquisition of data and image processing
Shengnan Jiang: data and statistical analysis
Ying Liu: study design and revising the manuscript

All the authors approved the submission and agreed to be accountable for all factual aspects of the work and for ensuring questions related to the accuracy or integrity of any part of the article.

Sources of financial support

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical Statement

The study was conducted according to the guidelines of the Declaration of Helsinki and was approved by the Institutional Review Board. All procedures were performed in accordance with the relevant guidelines and regulations.

We treat all individuals, communities, and the environment with the utmost dignity, empathy, and care. Prioritizing the well-being and safety of all stakeholders, we actively avoid any actions that could potentially cause harm. Our commitment to fair and equitable treatment is unwavering, and we do not discriminate based on any personal characteristics or group affiliations. We respect individual privacy and handle all personal information with the greatest care, in full compliance with applicable laws and regulations.

Throughout the research process, we have remained steadfast in upholding ethical principles and safeguarding the rights and dignity of our research participants.

CRediT authorship contribution statement

Xiaowen Qin: Investigation, Data curation. **Shengnan Jiang:** Data curation. **yingqi Luo:** Writing – original draft, Conceptualization. **Qingqi Yang:** Methodology, Data curation. **Ying Liu:** Writing – review & editing, Conceptualization. **Jinglang Hu:** Software, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejro.2024.100624.

References

[1] H. Rumgay, M. Arnold, J. Ferlay, et al., Global burden of primary liver cancer in 2020 and predictions to 2040, J. Hepatol. 77 (6) (2022) 1598–1606, https://doi.org/10.1016/j.jhep.2022.08.021.
[2] D.Q. Huang, A.G. Singal, Y. Kono, et al., Changing global epidemiology of liver cancer from 2010 to 2019: NASH is the fastest growing cause of liver cancer, Cell Metab. 34 (7) (2022) 969–977.e2, https://doi.org/10.1016/j.cmet.2022.05.003.
[3] B. Alawyia, C. Constantinou, Hepatocellular carcinoma: a narrative review on current knowledge and future prospects, Curr. Treat. Options Oncol. 24 (7) (2023) 711–724, https://doi.org/10.1007/s11864-023-01098-9.
[4] L. Di Tommaso, M. Spadaccini, M. Donadon, et al., Role of liver biopsy in hepatocellular carcinoma, World J. Gastroenterol. 25 (40) (2019) 6041–6052, https://doi.org/10.3748/wjg.v25.i40.6041.
[5] J.A. Marrero, J. Ahn, K. Rajender Reddy, Americal College of Gastroenterology. ACG clinical guideline: the diagnosis and management of focal liver lesions, Am. J. Gastroenterol. 109 (9) (2014) 1328–1348, https://doi.org/10.1038/ajg.2014.213.
[6] T.D. Poeppel, B.J. Krause, T.A. Heusner, et al., PET/CT for the staging and follow-up of patients with malignancies, Eur. J. Radio. 70 (3) (2009) 382–392, https://doi.org/10.1016/j.ejrad.2009.03.051.
[7] N. Tamaki, K. Hirata, T. Kotani, et al., Four-dimensional quantitative analysis using FDG-PET in clinical oncology, Jpn J. Radio. 41 (8) (2023) 831–842, https://doi.org/10.1007/s11604-023-01411-4.
[8] K. Ozaki, K. Harada, N. Terayama, et al., FDG-PET/CT imaging findings of hepatic tumors and tumor-like lesions based on molecular background, Jpn J. Radio. 38 (8) (2020) 697–718, https://doi.org/10.1007/s11604-020-00961-1.
[9] L.Q. Zhou, J.Y. Wang, S.Y. Yu, et al., Artificial intelligence in medical imaging of the liver, World J. Gastroenterol. 25 (6) (2019) 672–682, https://doi.org/10.3748/wjg.v25.i6.672.
[10] A. Hosny, C. Parmar, J. Quackenbush, et al., Artificial intelligence in radiology, Nat. Rev. Cancer 18 (8) (2018) 500–510, https://doi.org/10.1038/s41568-018-0016-5.
[11] M.A. Mazurowski, M. Buda, A. Saha, et al., Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI, J. Magn. Reson Imaging 49 (4) (2019) 939–954, https://doi.org/10.1002/jmri.26534.
[12] S.P. Singh, L. Wang, S. Gupta, et al., 3D deep learning on medical images: a review, Published 2020 Sep 7. Sens. (Basel) 20 (18) (2020) 5097, https://doi.org/10.3390/s20185097.
[13] R. Gao, S. Zhao, K. Aishanjiang, et al., Deep learning for differential diagnosis of malignant hepatic tumors based on multi-phase contrast-enhanced CT and clinical data, Published 2021 Sep 26. J. Hematol. Oncol. 14 (1) (2021) 154, https://doi.org/10.1186/s13045-021-01167-2.
[14] K. Sridhar, K. C, W.C. Lai, et al., Detection of liver tumour using deep learning based segmentation with coot extreme learning model, Published 2023 Mar 6. Biomedicines 11 (3) (2023) 800, https://doi.org/10.3390/biomedicines11030800.
[15] C.A. Hamm, C.J. Wang, L.J. Savic, et al., Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI, Eur. Radio. 29 (7) (2019) 3338–3347, https://doi.org/10.1007/s00330-019-06205-9.
[16] F. Wang, Q. Chen, Y. Chen, et al., A novel multimodal deep learning model for preoperative prediction of microvascular invasion and outcome in hepatocellular carcinoma, Eur. J. Surg. Oncol. 49 (1) (2023) 156–164, https://doi.org/10.1016/j.ejso.2022.08.036.
[17] Y.C. Lai, K.C. Wu, C.J. Chang, et al., Predicting overall survival with deep learning from 18F-FDG PET-CT images in patients with hepatocellular carcinoma before liver transplantation, Diagnostics (Basel) 13 (5) (2023) 981, https://doi.org/10.3390/diagnostics13050981.
[18] L. Hartmann, L. Bundschuh, N. Zsótér, et al., Tumor heterogeneity for differentiation between liver tumors and normal liver tissue in 18F-FDG PET/CT, Nuklearmedizin 60 (1) (2021) 25–32, https://doi.org/10.1055/a-1270-5568.
[19] D. Visvikis, C. Cheze Le Rest, V. Jaouen, et al., Artificial intelligence, machine (deep) learning and radio(geno)mics: definitions and nuclear medicine imaging applications, Eur. J. Nucl. Med. Mol. Imaging 46 (13) (2019) 2630–2637, https://doi.org/10.1007/s00259-019-04373-w.
[20] Z. Xue, P. Li, L. Zhang, et al., Multi-modal co-learning for liver lesion segmentation on PET-CT images, IEEE Trans. Med. Imaging 40 (12) (2021) 3531–3542, https://doi.org/10.1109/TMI.2021.3089702.
[21] M. Renzulli, N. Brandi, G. Argalia, et al., Morphological, dynamic and functional characteristics of liver pseudolesions and benign lesions, Radio. Med. 127 (2) (2022) 129–144, https://doi.org/10.1007/s11547-022-01449-w.
[22] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, et al., UNet+ +: redesigning skip connections to exploit multiscale features in image segmentation, IEEE Trans. Med Imaging 39 (6) (2020) 1856–1867, https://doi.org/10.1109/TMI.2019.2959609.
[23] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, et al., UNet++: a nested u-net architecture for medical image segmentation, Deep Learn Med Image Anal. Multimodal Learn Clin. Decis. Support (2018) 11045 (2018) 3–11, https://doi.org/10.1007/978-3-030-00889-5_1.
[24] European Association for the Study of the Liver, EASl clinical practice guidelines: management of hepatocellular carcinoma, J. Hepatol. 69 (1) (2018) 182–236, https://doi.org/10.1016/j.jhep.2018.03.019.