Multimodal deep learning model on interim $^{[18]F}$FDG PET/CT for predicting primary treatment failure in diffuse large B-cell lymphoma

Cheng Yuan, Qing Shi, Xinyun Huang, Li Wang, Yang He, Biao Li, Weili Zhao, Dahong Qian

Received: 14 March 2022 / Revised: 30 May 2022 / Accepted: 13 July 2022
© The Author(s), under exclusive licence to European Society of Radiology 2022

Abstract

Objectives The prediction of primary treatment failure (PTF) is necessary for patients with diffuse large B-cell lymphoma (DLBCL) since it serves as a prominent means for improving front-line outcomes. Using interim $^{[18]F}$-fluoro-2-deoxyglucose ($^{[18]F}$FDG) positron emission tomography/computed tomography (PET/CT) imaging data, we aimed to construct multimodal deep learning (MDL) models to predict possible PTF in low-risk DLBCL.

Methods Initially, 205 DLBCL patients undergoing interim $^{[18]F}$FDG PET/CT scans and the front-line standard of care were included in the primary dataset for model development. Then, 44 other patients were included in the external dataset for generalization evaluation. Based on the powerful backbone of the Conv-LSTM network, we incorporated five different multimodal fusion strategies (pixel intermixing, separate channel, separate branch, quantitative weighting, and hybrid learning) to make full use of PET/CT features and built five corresponding MDL models. Moreover, we found the best model, that is, the hybrid learning model, and optimized it by integrating the contrastive training objective to further improve its prediction performance.

Results The final model with contrastive objective optimization, named the contrastive hybrid learning model, performed best, with an accuracy of 91.22% and an area under the receiver operating characteristic curve (AUC) of 0.926, in the primary dataset. In the external dataset, its accuracy and AUC remained at 88.64% and 0.925, respectively, indicating its good generalization ability.

Conclusions The proposed model achieved good performance, validated the predictive value of interim PET/CT, and holds promise for directing individualized clinical treatment.

Key Points
- The proposed multimodal models achieved accurate prediction of primary treatment failure in DLBCL patients.
- Using an appropriate feature-level fusion strategy can make the same class close to each other regardless of the modal heterogeneity of the data source domain and positively impact the prediction performance.
- Deep learning validated the predictive value of interim PET/CT in a way that exceeded human capabilities.

Keywords Lymphoma · Positron emission tomography/computed tomography · Treatment failure · Deep learning

Cheng Yuan, Qing Shi, and Xinyun Huang contributed equally to this work.

✉ Biao Li
lb10363@rjh.com.cn
✉ Weili Zhao
zhao.weili@yahoo.com
✉ Dahong Qian
dahong.qian@sjtu.edu.cn

1 School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200040, China
2 Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China
3 Department of Nuclear Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

Published online: 27 August 2022
Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most frequently observed histologic subtype of lymphoma and is particularly prevalent in Asia [1]. In clinical DLBCL treatment, R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone) chemotherapy is used as the front-line standard of care [2]. However, as many as 15% of patients undergoing this chemotherapy experience primary treatment failure (PTF), which limits their median survival time to 1 year at most [3]. In addition to high-dose chemotherapy and autologous stem-cell transplantation, novel therapies, including chimeric antigen receptor T cell therapy, have shown promising efficacy and have provided increasing alternatives for patients at high risk for PTF to R-CHOP [4–6]. Thus, it is crucial to identify these patients at high risk of PTF, so they can promptly receive a more potent therapy to improve their survival time and quality of life.

Recently, some studies have identified refractory DLBCL patients and predicted their survival after treatment with R-CHOP chemotherapy [7, 8]. The revised international prognostic index (R-IPI) redistributed IPI factors and predicted the outcome of DLBCL patients based on the number of negative prognostic factors present at the time of diagnosis [9]. Many genetic changes, such as the presence of TP53 and KMT2D mutations, have been found to be independent markers of poor prognosis in DLBCL patients [10]. Nevertheless, R-IPI has limited utility in predicting PTF, especially for low-risk DLBCL, as most patients by definition have a favourable risk profile [7]. The confirmation of TP53 and KMT2D mutations relies on large-scale high-throughput gene sequencing, which requires extra time and high labour intensity [8]. With the development of medical imaging, $^{18}$F-fluoro-2-deoxyglucose ($^{18}$F-FDG) positron emission tomography/computed tomography (PET/CT) serves as an effective tool in the diagnosis and staging of DLBCL [11–13]. Some initial studies based on Kaplan–Meier survival analyses suggested that PET/CT scans performed early during treatment (that is, interim PET/CT) after 4 cycles of R-CHOP chemotherapy could identify patients likely to relapse [14, 15]. However, no research to determine the predictive value of interim PET/CT in identifying PTF in DLBCL patients is currently available, and it remains an unmet need.

With the rapid development of artificial intelligence, deep learning approaches have developed as popular mathematical models to analyse PET/CT data [16], and in the field of lymphoma, they have been applied for tasks related to lymph node detection [17, 18], normal physiological FDG uptake identification [19], lesion segmentation [20, 21], and, only recently, survival prediction [22]. Capobianco et al [23] estimated total metabolic tumour volume (TMTV) from $^{18}$F-FDG PET and suggested that TMTV using a deep learning–based method displayed significant prognostic value for overall survival in DLBCL patients. In addition, the identification of CT features provides a noninvasive and accurate method that indicates possible PTF early during treatment so more potent therapies may be selected.

Abbreviations

| Acronym | Full Form | Description |
|---------|-----------|-------------|
| AUC     | Area under the receiver operating characteristic curve | |
| DLBCL   | Diffuse large B-cell lymphoma | |
| ECOG    | Eastern Cooperative Oncology Group | |
| IQR     | Interquartile range | |
| LDH     | Lactate dehydrogenase | |
| LSTM    | Long short-term memory | |
| MDL     | Multimodal deep learning | |
| PET/CT  | Positron emission tomography/computed tomography | |
| PTF     | Primary treatment failure | |
| R-CHOP  | Rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone | |
| R-IPI   | Revised international prognostic index | |
| TMTV    | Total metabolic tumour volume | |
Materials and methods

Patients and dataset

All patients were collected from Ruijin Hospital and comprised two subsets of a consecutively collected observational DLBCL cohort, which was in accordance with the Declaration of Helsinki. A complete flowchart of the data collection process is shown in Fig. 1. Prior to analysis, the patients were divided into two groups for comparison: PTF and non-PTF DLBCL. Specific rules of data inclusion, exclusion, and grouping are described in Sec. 1 of the supplementary materials. In addition, detailed clinical characteristics of all patients were collected, including age (median with interquartile range [IQR]), age range (≤60 years versus >60 years), sex, IPI (0 versus 1), stage (I–II versus III–IV), Eastern Cooperative Oncology Group (ECOG) performance status (0 versus 1), serum lactate dehydrogenase (LDH) level (normal versus elevated), extralymphatic involvement, and B symptoms.

Image acquisition and preprocessing

Image data were acquired from a PET/CT scanner (GE Healthcare) with the reconstruction method of ordered subset expectation maximization. Each sample contained one CT volume with a resolution of 512 × 512 pixels at 0.98 mm × 0.98 mm and one PET volume with a resolution of 128 × 128 pixels at 5.47 mm × 5.47 mm. Both volumes were reconstructed with the same number of slices, and the interdistance was 3.27 mm. The standard data preprocessing routine is shown in Sec. 2 of the supplementary materials [33, 34]. Of note, the number of positive samples was obviously smaller than the
number of negative samples, and the discrepancy was due to the favourable prognosis of low-risk DLBCL in the clinic. To reduce the overfitting impact of data imbalance, we employed data augmentation strategies only in the network training of the each-fold cross-validation experiment, wherein we included random horizontal and vertical flipping of the input images. Instead, we kept the initial positive ratio of the test cohort in the primary dataset.

**MDL model development**

An overview of our prediction framework is shown in Fig. 2. The starting point of our model is Conv-LSTM [31], a classic deep learning architecture for natural image recognition that has been recently applied to medical image analysis and achieved satisfactory performance [35]. Our network backbone (Fig. 2(a)) was built based on Conv-LSTM with a simple adjustment of constructing two identical encoders of PET and CT data. To extract the hidden image features of the input data, four blocks of convolution and pooling operations were conducted. Then, a recursive learning framework was introduced, with a structure called “long short-term memory” (LSTM) [36, 37], which performed simple learned gating functions to allow the learning parameters to be updated or reset. The above extracted features were concatenated into a sequence, which was then transformed by the LSTM into a composite feature vector for the sample. Thus, the heterogeneous information of the input data was derived to high-level semantic features reflecting intraslice spatial structures and interslice contextual correlations. The output of the model was a set of two continuous variables representing the prediction probability (on the scale of 0.0 to 1.0) for each category and was treated as a discrete probability distribution. The final prediction was calculated as the probability-weighted average of the categories rounded to the nearest integer.

In addition to the above backbone, multiple MDL models using different feature fusion strategies were developed and compared (Fig. 2(b)), including the pixel intermixing model (I), separate channel model (II), separate branch model (III), quantitative weighting model (IV), and hybrid learning model (V). Their main differences are exhibited in terms of feature encoding. The first model (I) is based on input-level fusion distinguishing from other feature-level fusion approaches. Here, a PET slice and its corresponding CT slice were integrated as one input image via pixel intermixing for single-branch encoding. Second (II), PET and CT data were read into one encoder by separate channels and were simply concatenated after the first group of convolution and pooling operations. For the third model (III), the output feature maps from separate PET and CT encoding branches were concatenated before being fed into the following LSTM predictor. Fourth (IV), the model learned the spatial contribution of feature maps from PET and CT encoders by a quantitative weighting strategy that calculated the convolutional result as a weighted matrix. In the last model (V), PET and CT features extracted from two identical encoders were combined by the hybrid learning approach, a modal fusion method we published before, which generated spatial fusion maps and quantified the contribution of the complementary information. Then, these fusion maps were concatenated with specific-modality (i.e. PET and CT) feature maps to obtain a representation of the final fused feature maps at different scales.

To achieve better performance, we further aimed to promote the intraclass cohesion and interclass separation of the semantic embeddings of PTF and non-PTF cases. Thus, we adopted contrastive learning [32] in the hybrid learning model to achieve that goal. Traditionally, we trained the models (I–V) with a cross-entropy of the prediction and ground truth (Fig. 2(c)). Here, the cross-entropy loss was integrated with a contrastive training objective (Fig. 2(d)) derived from the similarity between a pair of samples to generate the overall loss function. Trained in this way, we built the contrastive hybrid learning model (VI) with an enhanced prediction ability because the same class lies close to each other regardless of the modal heterogeneity of the data source domain and away from those in different classes. Details of constructing the overall training objective are described in Sec. 3 of the supplementary materials.

**Model implementation and visualization**

We implemented the MDL models using TensorFlow 1.14 [38] on a machine running Windows 10 with CUDA 10.0 and cuDNN 7.6 [39]. Model training was performed on an 11 GB NVIDIA GeForce RTX2080 Ti. To establish appropriate training parameters, we employed values of 0.1, 0.01, and 4 for the parameters of the regularization factor, the learning rate, and the batch size, respectively. We used Sec. 4 of the supplementary materials to present the detailed architecture parameters used in the MDL models.

For each sample, model attention can be visualized for physician comprehension and validation. Here, focused regions of hidden-layer feature maps were instructed in the form of a rough location heatmap [40], which highlighted the entry area of prediction targets and interpreted the explanatory nature of such models about what kinds of features contributed to outputs. In this way, complex features that passed deep convolutional and pooling layers were projected onto the original input image.

**Statistical analysis**

To reveal the difference between the clinical characteristics of the PTF group and the non-PTF group, we utilized the statistical package SPSS version 22.0 for various aspects of univariate analysis, including the Mann–Whitney U test for
numerical variables and the Pearson’s chi-square test and, if necessary, Fisher’s exact tests for categorical variables. Here, \( p \) values of < 0.05 were considered statistically significant. The prediction results were drawn from a fivefold cross-validation, where samples of the primary dataset were divided into three cohorts for model training, validation, and testing at a ratio of 3:1:1. Samples of the external dataset were only tested for generalization validation. The main metrics used to evaluate the model performance were accuracy as well as sensitivity, specificity, and F1 score. Sec. 5 of the supplementary materials lists the specific formulas for these evaluation metrics.

Fig. 2 Workflow of multimodal deep learning (MDL) models based on PET/CT for predicting primary treatment failure (PTF) in patients with diffuse large B-cell lymphoma (DLBCL). The legend is shown in the bottom right corner.
Results

Baseline clinical characteristics

Table 1 displays the baseline characteristics of the patients in the primary dataset (median age: 55.00 years; 95 females, 110 males) and external dataset (median age: 54.50 years; 23 females, 21 males). The PTF rates were 9.76% (20/205) and 9.10% (4/44), respectively, showing a similarity between the two datasets. The percentages of PTF patients and non-PTF patients who possessed one IPI risk factor were 95.00% and 54.05% (100/185), respectively, which demonstrates the existence of a significant difference between the two groups in the primary dataset ($p < 0.001$).

Performance comparison of MDL models

The prediction performance metrics of the MDL models in the primary (that is, the test cohort) and external datasets are listed in Table 2. Due to the number imbalance between the PTF and non-PTF groups, the F1 scores of all MDL models were relatively low, which we still considered a meaningful result. Overall, the hybrid learning model achieved the best performance among all evaluation metrics in the test cohort of the primary dataset for predicting PTF. This indicated that the hybrid learning feature fusion strategy increased both the ratio of true positives to false positives and that of true negatives to false negatives. Figure 3(a) shows that the areas under the receiver operating characteristic curves (AUCs) were 0.837 in the test cohort of the primary dataset and 0.869 in the external dataset. Although the quantitative weighting model achieved a better AUC of 0.844 in the primary dataset, we took the AUC of the external dataset as a more important indicator to ensure its better generalization ability. Therefore, we further optimized the hybrid learning model by integrating it with a contrastive training objective and finally established the contrastive hybrid learning model.

The contrastive hybrid learning model achieved AUCs of 0.926 and 0.925 in the primary (that is, the test cohort) and external datasets, respectively. In addition, the DeLong test demonstrated that the AUC of this model was significantly better than

Table 1  Baseline clinical characteristics of the primary and external datasets

| Variable | Primary dataset ($n=205$) | P value | External dataset ($n=44$) | P value |
|----------|-----------------------------|---------|-----------------------------|---------|
|          | Non-PTF ($n=185$) | PTF ($n=20$) |                      | Non-PTF ($n=40$) | PTF ($n=4$) |
| Age (y)  | 55.00 (45.50–63.00) | 57.50 (42.00–62.75) | 0.671<sup>a</sup> | 54.50 (44.25–62.25) | 49.50 (40.75–59.00) | 0.653<sup>a</sup> |
| Age range| 0.162<sup>b</sup> |                         |                      | 0.559<sup>b</sup> |                         |
| ≤ 60 years | 130 (70.27%) | 11 (55.00%) |                         | 30 (75.00%) | 4 (100.00%) |                      |
| > 60 years | 55 (29.73%) | 9 (45.00%) |                         | 10 (25.00%) | 0 (0.00%) |                      |
| Sex      | 0.284<sup>b</sup> |                         |                      | 0.335<sup>b</sup> |                         |
| Female   | 88 (47.57%) | 7 (35.00%) |                          | 22 (55.00%) | 1 (25.00%) |                      |
| Male     | 97 (52.43%) | 13 (65.00%) |                         | 18 (45.00%) | 3 (75.00%) |                      |
| IPI      | < 0.001<sup>b</sup> |                      |                      | 1.000<sup>b</sup> |                      |
| 0        | 85 (45.95%) | 1 (5.00%) |                          | 19 (47.50%) | 2 (50.00%) |                      |
| 1        | 100 (54.05%) | 19 (95.00%) |                         | 21 (52.50%) | 2 (50.00%) |                      |
| Stage    | 1.000<sup>b</sup> |                      |                      | 1.000<sup>b</sup> |                      |
| I–II     | 174 (94.05%) | 19 (95.00%) |                         | 33 (82.50%) | 2 (50.00%) |                      |
| III–IV   | 11 (5.95%) | 1 (5.00%) |                          | 7 (17.50%) | 2 (50.00%) |                      |
| ECOG status | 0.293<sup>b</sup> |                      |                      | 1.000<sup>b</sup> |                      |
| 0        | 163 (88.11%) | 16 (80.00%) |                         | 35 (87.50%) | 4 (100.00%) |                      |
| 1        | 22 (11.89%) | 4 (20.00%) |                           | 5 (12.50%) | 0 (0.00%) |                      |
| LDH level | 0.543<sup>b</sup> |                      |                      | 1.000<sup>b</sup> |                      |
| Normal   | 152 (82.16%) | 15 (75.00%) |                         | 36 (90.00%) | 4 (100.00%) |                      |
| Elevated | 33 (17.84%) | 5 (25.00%) |                           | 4 (10.00%) | 0 (0.00%) |                      |
| Extralymphatic involvement | 0.066<sup>b</sup> | 0.725<sup>b</sup> | 2 (5.00%) | 0 (0.00%) |                      |
| B symptoms | 23 (12.43%) | 3 (15.00%) |                         | 2 (5.00%) | 0 (0.00%) |                      |

Note that data are number of patients

ECOG Eastern Cooperative Oncology Group, LDH lactate dehydrogenase

<sup>a</sup>Mann–Whitney U test

<sup>b</sup>Pearson’s chi-squared test; if necessary, Fisher’s exact test

European Radiology
Table 2  Performance comparison of multimodal deep learning (MDL) models

| Models                      | Datasets | Sensitivity (%) | Specificity (%) | Accuracy (%) | F1 score (%) | AUC |
|-----------------------------|----------|----------------|-----------------|--------------|--------------|-----|
| Model I                     | Primary  | 40.00          | 86.49           | 81.95        | 30.19        | 0.774|
| Pixel intermixing           | External | (21.88–61.34)  | (80.81–90.68)   | (76.11–86.61)| (16.18–49.17)| 0.688|
|                             | External | 50.00          | 77.50           | 75.00        | 26.67        | 0.794|
| Model II                    | Primary  | 45.00          | 87.03           | 82.93        | 33.96        | 0.810|
| Separate channel            | External | (25.82–65.79)  | (81.42–91.13)   | (77.87–87.46)| (19.03–52.89)| 0.706|
|                             | External | 50.00          | 85.00           | 81.82        | 33.33        | 0.725|
| Model III                   | Primary  | 50.00          | 90.27           | 86.34        | 41.67        | 0.844|
| Separate branch             | External | (29.93–70.07)  | (85.15–93.76)   | (80.97–90.38)| (24.48–61.10)| 0.844|
|                             | External | 50.00          | 85.00           | 81.82        | 33.33        | 0.725|
| Model IV                    | Primary  | 45.00          | 91.35           | 86.83        | 40.00        | 0.844|
| Quantitative weighting      | External | (25.82–65.79)  | (86.41–94.61)   | (81.52–90.79)| (22.70–60.20)| 0.766|
|                             | External | 50.00          | 87.50           | 84.09        | 36.36        | 0.837|
| Model V                     | Primary  | 65.00          | 92.43           | 89.76        | 55.32        | 0.869|
| Hybrid learning             | External | (43.29–81.88)  | (87.70–95.44)   | (84.85–93.20)| (35.95–73.10)| 0.844|
|                             | External | 75.00          | 87.50           | 86.36        | 50.00        | 0.869|
| Model VI                    | Primary  | 70.00          | 93.51           | 91.22        | 60.87        | 0.926|
| Contrastive hybrid learning | External | (48.10–85.45)  | (89.01–96.25)   | (86.55–94.37)| (40.82–77.70)| 0.925|
|                             | External | 75.00          | 90.00           | 88.64        | 54.55        | (20.73–83.97)|

AUC comparison between the contrastive hybrid learning model and other models using Delong test

| Compared method | Model I | Model II | Model III | Model IV | Model V |
|-----------------|---------|----------|-----------|----------|---------|
| Primary dataset | 0.001   | 0.006    | 0.020     | 0.028    | 0.051   |
| External dataset| 0.030   | 0.021    | 0.054     | 0.018    | 0.088   |

The best metrics are shown in the bold numbers. Numbers in parentheses are 95% confidence intervals (95% CIs)

that of the pixel intermixing model (p < 0.001), separate channel model (p < 0.05), separate branch model (p < 0.05), and quantitative weighting model (p < 0.05) in the primary dataset. Quantitative comparisons of the other four evaluation metrics are exhibited in Table 2. In addition, the normalized confusion matrices of all MDL models for distinguishing PTF from non-PTF in the test cohort of the primary dataset are shown in Fig. 3(b). Notably, the contrastive hybrid learning model achieved continuous improvements in the overall sensitivity of PTF patients, as indicated through models I to VI.

Interpretability of MDL models

For each sample, model attention can be visualized for clinical comprehension and validation. Here, we aimed to understand which areas of input images and what kinds of features contributed to the prediction. Figure 4 shows input PET/CT images and the heatmaps of corresponding locations for three patients randomly chosen from the test cohort, which demonstrates the existence of a common pattern that is consistently shared among all samples. As shown in PET images, physiological tracer uptake and lesions are delineated in green and red contours, respectively. Despite the interference of physiological uptake, the contrastive hybrid learning model paid great attention to the structure of lesions. We thought the learning mechanism inside it could help rectify the spillover by iteratively recognizing the differences between the lesions and the regions with physiological tracer uptake and then reduce the spill-in effect. This suggested that this model actively sought tumour lesion distribution areas to classify PTF and non-PTF. In addition, the compared models viewed different lesion-adjacent areas for the same patient, a discrepancy that explains why these models differed in their prediction performance. Specifically, the heatmap of the contrastive hybrid learning model contained more gradient attention on specific regions related to the tumour itself, except necrosis and peripheral inflammation [41].

Discussion

The prediction of PTF in DLBCL patients has been a prominent challenge that clinicians have faced for a long time. In this work, we developed and validated a group of deep learning–based multimodal models that learned complementary high-level semantic features from interim [18F]FDG PET/
CT images and achieved an individualized and noninvasive prediction of PTF in patients with low-risk DLBCL at the end of treatment. The major findings of our experiments covered the following issues: (1) As far as we know, our work is a seminal pioneering one, apparently the first that applies the MDL approach to interim PET/CT images acquired from DLBCL patients to validate its predictive value for PTF. (2) The prediction performance of the present MDL model, based on both the incorporation of a hybrid learning feature fusion strategy and the enhancement of the contrastive training objective, significantly outperformed (almost all AUCs, $p < 0.05$) other models using different feature fusion strategies in ablation comparisons. (3) Our work provides solid evidence that the contrastive hybrid learning approach on PET/CT images provides an effective method for the prediction of PTF and the risk stratification of patients with DLBCL.
From the clinical point of view, the tumour involvement changes calculated from pathological FDG uptake in PET imaging can indicate possible PTF, but its use remains limited because of its dependence on data both prior to and after treatment. Moreover, the identification of CT parameters in the tumour region has been demonstrated to be an effective discrimination technique for predicting PTF in different carcinomas [24, 25]. However, the present models for extracting and combining invisible imaging features of PET and CT are still not adequate for predicting PTF in DLBCL. In the current work, we proposed several MDL models based on [18F]FDG PET/CT to predict PTF, and we investigated ways of fully utilizing PET/CT data to achieve the best performance. The proposed contrastive hybrid learning model demonstrated particularly outstanding predictive performance for PTF, as it effectively integrated PET metabolic features with corresponding CT anatomic features [42]. In contrast, the quantitative weighting model implemented element multiplication to encompass the level of importance given to information from each modality, although it considerably weakened the natural
characteristics of each modality [30]. As shown in Table 2, the separate branch model included a layer-level fusion strategy based on simple concatenation and then applied prediction layers [43]. Thus, some useful information associated with complementary features may be lost. In addition, the separate channel model combined the PET and CT images resulting after the first convolutional layer to derive fused feature maps [44], which led to lopsided attention to a modality with dominant pixel strength. The pixel intermixing strategy was used to construct a type of early fusion model [45], which shared a similar weakness with the separate channel model, reducing the prediction accuracy.

To interpret the MDL model, we visualized regions of interest of the network by generating rough location heatmaps. The activated areas in the heatmap were found to be primarily located in the tumour lesion and its surrounding areas according to several MDL models. All these areas were consistent with the predictive region observed by experienced radiologists. Notably, the contrastive hybrid learning model paid more precise attention while avoiding interference from physiological uptake. These common patterns served as a clue for the working principle of MDL models for analysing PET/CT data.

Although our results are definitely promising, our present work has a few limitations, thereby leaving room for several future improvements. First, this work was a retrospective study based on a relatively small sample size, especially in terms of the number of positive samples. Although large numbers of PET/CT scans from different institutions are difficult to obtain, the addition of in-house data would definitely be of paramount importance to the current work. We are actively working on this task. Second, to indicate accurate PTF for planning personalized treatments, the sensitivity and accuracy should be very high. Thus, the MDL model provided a reference result but not a direct decision for clinical practice, given the need for prospective validation. In the future, we aim to make it convenient for clinical decision-making by encapsulating the well-pretrained model as a ready-to-use tool. Then, physicians could feed the original data sample into the model and directly obtain its PTF probability by simple operations. Finally, we did not consider other possible prognostic factors. We suggest that integrating biologic markers, including blood biomarkers and pathological and genetic features, may improve the accuracy and robustness of our model.

In conclusion, our work proposed MDL models and developed the best model named the contrastive hybrid learning model. It utilized the hybrid learning strategy and was optimized by contrastive training objective to generate complementary features for multimodal learning. Adequate ablation experiments proved the superiority of the proposed model. In addition, it is end-to-end trainable and avoids the need for time-consuming manual delineations. Therefore, it not only quantitatively validates the predictive value of interim PET/CT but also provides proof of concept for multimodal data analysis in clinical decision-making.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s00330-022-09031-8.

**Code availability** The code of our study is publicly accessible at https://github.com/cyuan-sjtu/MDL-model.

**Funding** This study has received funding by the National Natural Science Foundation of China (81974276, 81830007, 81520108003, 81670176, and 82070204), Chang Jiang Scholars Program, Shanghai Municipal Education Commission Gaofeng Clinical Medicine Grant Support (20152206, and 20152208), Clinical Research Plan of Shanghai Hospital Development Center (SHDC2020CR1032B), Multicenter Clinical Research Project by Shanghai Jiao Tong University School of Medicine (DLY201601), Collaborative Innovation Center of Systems Biomedicine, and the Samuel Waxman Cancer Research Foundation.

**Data availability** The data that support the findings of this study are available on reasonable request from the corresponding author.

**Declarations**

**Guarantor** The scientific guarantor of this publication is Dr. Dahong Qian.

**Conflict of interest** The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

**Statistics and biometry** No complex statistical methods were necessary for this paper.

**Informed consent** Written informed consent was not required for this study because of its retrospective nature.

**Ethics approval** Institutional Review Board approval was obtained at Shanghai Ruijin Hospital.

**Methodology**
- **Retrospective**
- **Diagnostic or prognostic study**
- **Performed at one institution**

**References**

1. Sung H, Ferlay J, Siegel RL et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71(3): 209–249

2. Feugier P, Van Hoof A, Sebban C et al (2005) Long-term results of the R-CHOP study in the treatment of elderly patients with diffuse
large B-cell lymphoma: a study by the Groupe d’Etude des Lymphomes de l’Adulte. J Clin Oncol 23:4117–4126

3. Crump M, Neelapu SS, Farooq U et al (2017) Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study. Blood 130(16):1800–1808

4. Gisselbrecht C, Nestle EVD (2018) How I manage patients with relapsed/refractory diffuse large B-cell lymphoma. Br J Haematol 182:632–643

5. Locke FL, Ghobadi A, Jacobson CA et al (2019) Long-term safety and activity of axi-cabtagene ciloleucel in refractory large B-cell lymphoma (ZUMA-1): a single-arm, multicentre, phase 1-2 trial. Lancet Oncol 20(1):31–42

6. Kalakonda N, Maerevoet M, Cavallo F et al (2020) Selinexor in patients with relapsed or refractory diffuse large B-cell lymphoma (SADAL): a single-arm, multinational, multicentre, open-label, phase 2 trial. Lancet Haematol 7(7):e511–e522

7. Hawkes EA, Barracough A, Sehn LH (2022) Limited-stage diffuse large B-cell lymphoma. Blood 139(6):822–834

8. Lv X, Wang Q, Ge X, Xue C, Liu X (2021) Application of high-temperature third-order 18F-FDG uptake after a few cycles of polychemotherapy is predictive for malignant lymphoma. J Clin Oncol 25(5):579–583

9. Sehn LH, Berry B, Chhanabhai M et al (2007) The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. Blood 190(6):1857–1861

10. Xu-Monette ZY, Wu L, Visco C et al (2012) Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma. Blood 119(24):5866–5870

11. Juweid ME, Stroobants S, Hoekstra OS et al (2007) Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. J Clin Oncol 25(5):571–578

12. Cheson BD, Pfistner B, Juweid ME et al (2007) Revised response criteria for malignant lymphoma. J Clin Oncol 25(5):579–586

13. Cheson BD (2011) Role of functional imaging in the management of lymphoma. J Clin Oncol 29(14):1844–1854

14. Jerusalem G, Beguin Y, Fassotte MF et al (2000) Persistent tumor [18F]FDG uptake after a few cycles of polychemotherapy is predictive of treatment failure in non-Hodgkin’s lymphoma. Haematologica 85(6):613–618

15. Spaepen K, Stroobants S, Dupont P et al (2002) Early restaging positron emission tomography with 18F-fluorodeoxyglucose predicts outcome in patients with aggressive non-Hodgkin’s lymphoma. Ann Oncol 13(9):1356–1363

16. Kirenko MI, Birol M, Gelardi F, Seregini E, Chiti A, Sollini M (2021) Deep learning in nuclear medicine—focus on CNN-based approaches for PET/CT and PET/MR: where do we stand? Clin Transl Imaging 9:37–55

17. Roth HR, Lu L, Seff A et al (2014) A new 2.5D representation for automated detection of pathological mediastinal lymph nodes in lung cancer patients. Eur J Nucl Med Mol Imaging 49:881–888

18. Bii L, Kim J, Kumar A, Wen L, Feng D, Fulham M (2017) Automatic detection and classification of regions of FDG uptake in whole-body PET-CT lymphoma studies. Comput Med Imaging Graph 60:3–10

19. Hu H, Shen L, Zhou T, Decazes P, Vera P, Ruan S (2020) Lymphoma segmentation in PET images based on multi-view and Conv3D fusion strategy. IEEE 17th International Symposium on Biomedical Imaging (ISBI):1197–1200

20. Revailier W, Cottereau AS, Rossi C et al (2022) Deep learning approach to automatize TMTV calculations regardless of segmentation methodology for major FDG-avid lymphomas. Diagnostics (Basel) 12(2):417

21. Sadik M, Lind E, Polymeri E, Enqvist O, Ulen J, Trägårdh E (2019) Automated quantification of reference levels in liver and mediastinal blood pool for the Deauville therapy response classification using FDG-PET/CT in Hodgkin and non-Hodgkin lymphomas. Clin Physiol Funct Imaging 39(1):78–84

22. Capobianco N, Meignan M, Cottereau AS et al (2021) Deep-learning [18F]FDG uptake classification enables total metabolic tumor volume estimation in diffuse large B-cell lymphoma. J Nucl Med 62(1):30–36

23. Seidler M, Forghani B, Reindolf C et al (2019) Dual-energy CT texture analysis with machine learning for the evaluation and characterization of cervical lymphadenopathy. Comput Struct Biotechnol J 17:1009–1015

24. Ganesan B, Miles KA, Babikir S et al (2017) CT-based texture analysis potentially provides prognostic information complementary to interim FDG-PET for patients with Hodgkin’s and aggressive non-Hodgkin’s lymphomas. Eur Radiol 27:1012–1020

25. Santiago R, Jimenez JO, Forghani R et al (2021) CT-based radiomics model with machine learning for predicting primary treatment failure in diffuse large B-cell Lymphoma. Transl Oncol 14(10):101188

26. Zhou T, Ruan S, Canu S et al (2019) A review: Deep learning for medical image segmentation using multi-modality fusion. Array 3:10000

27. Li K, Zhang R, Cai W (2021) Deep learning convolutional neural network (DLCNN): unleashed the potential of 18 F-FDG PET/CT in lymphoma. Am J Nucl Med Mol Imaging 11(4):327–331

28. Jin C, Yu H, Ke J et al (2021) Predicting treatment response from longitudinal images using multi-task deep learning. Nat Commun 12:1851

29. Kumar A, Fulham M, Feng D, Kim J (2020) Co-learning feature fusion maps from PET-CT images of lung cancer. IEEE Trans Med Imaging 39(1):204–217

30. Donahue J, Hendricks LA, Rohrbach M et al (2017) Long-term recurrent convolutional networks for visual recognition and description. IEEE Trans Pattern Anal Mach Intell 39(4):677–691

31. Chen T, Kornblith S, Norouzi M, Hinton G (2020) A simple framework for contrastive learning of visual representations. Preprint arXiv:2002.05709

32. Zhong Z, Kim Y, Plichta K et al (2019) Simultaneous co-segmentation of tumors in PET-CT images using deep fully convolutional networks. Med Phys 46(2):619–633

33. Zhao X, Li L, Lu W, Tan S (2018) Tumor co-segmentation in PET/CT using multi-modality fully convolutional neural network. Phys Med Biol 64(1):015011

34. Humphries SM, Notary AM, Centeno JP et al (2019) Deep learning enables automatic classification of emphysema pattern at CT. Radiology 294(2):426–444

35. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

36. Lu N, Wu Y, Feng L, Song J (2019) Deep learning for fall detection: three-dimensional CNN combined with LSTM on video kinematic data. IEEE J Biomed Health Inform 23(1):314–323

37. Abadi M, Barham P, Chen J et al (2016) TensorFlow: a system for large-scale machine learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation: 265–283

38. Chetlur A, Woolley C, Vandermeersch P et al (2014) cuDNN: efficient primitives for deep learning. Preprint arXiv:1410.0759

39. Chetlur A, Woolley C, Vandermersch P, et al (2014) cuDNN: efficient primitives for deep learning. Preprint arXiv:1410.0759

40. Zhou B, Khosla A, Lapedriza A, Oliva A, Torrralba A (2016) Network in Network. Preprint arXiv:1410.0759
41. Du D, Feng H, Lv W et al (2020) Machine learning methods for optimal radiomics-based differentiation between recurrence and inflammation: application to nasopharyngeal carcinoma post-therapy PET/CT images. Mol Imaging Biol 22:730–738

42. Yuan C, Zhang M, Huang X et al (2021) Diffuse large B-cell lymphoma segmentation in PET-CT images via hybrid learning for feature fusion. Med Phys 48(7):3665–3678

43. Peng Y, Bi L, Guo Y, Feng D, Fulham M, Kim J (2019) Deep multi-modality collaborative learning for distant metastases predication in PET-CT soft-tissue sarcoma studies. Annu Int Conf IEEE Eng Med Biol Soc:3658-3688

44. Zhang W, Li R, Deng H et al (2015) Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. Neuroimage 108:214–224

45. Zhong Z, Kim Y, Zhou L, et al (2018) 3D fully convolutional networks for co-segmentation of tumors on PET-CT images. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI):228-231

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.