Graph-based multimodal multi-lesion DLBCL treatment response prediction from PET images

Oriane Thiery[1], Mira Rizkallah[1], Clément Bailly[2,3], Caroline Bodet-Milin[2,3], Emmanuel Itti[4], René-Olivier Casasnovas[5], Steven Le Gouill[3], Thomas Carlier[2,3], and Diana Mateus[1]

[1] Nantes Université, Centrale Nantes, CNRS, LS2N, UMR 6004, France
[2] Nuclear Medicine Department, University Hospital, Nantes, France
[3] Nantes Université, Inserm, CNRS, Université d'Angers, CRCI2NA, Nantes, France
[4] Nuclear Medicine, CHU Henri Mondor, Paris-Est University, Créteil, France
[5] Hematology, CHU Dijon Bourgogne, Dijon, France

Abstract. Diffuse Large B-cell Lymphoma (DLBCL) is a lymphatic cancer involving one or more lymph nodes and extranodal sites. Its diagnostic and follow-up rely on Positron Emission Tomography (PET) and Computed Tomography (CT). After diagnosis, the number of non-responding patients to standard front-line therapy remains significant (30-40%). This work aims to develop a computer-aided approach to identify high-risk patients requiring adapted treatment by efficiently exploiting all the information available for each patient, including both clinical and image data. We propose a method based on recent graph neural networks that combine imaging information from multiple lesions, and a cross-attention module to integrate different data modalities efficiently. The model is trained and evaluated on a private prospective multicentric dataset of 583 patients. Experimental results show that our proposed method outperforms classical supervised methods based on either clinical, imaging or both clinical and imaging data for the 2-year progression-free survival (PFS) classification accuracy.

Keywords: Multimodal data fusion · Graph Neural Networks · Cross-attention · DLBCL · Treatment Response · PET.

1 Introduction

Diffuse Large B-cell Lymphoma (DLBCL) is a cancer of the lymphatic system and the most common type of Non-Hodgkin Lymphoma (NHL). Its incidence is regularly growing, accounting for 30-40% of the 77240 new NHL cases in the US in 2020 [15]. The diagnosis and follow-up include analysing clinical biomarkers and the semi-quantitative interpretation of 18F-Fluorodeoxyglucose (FDG)-PET/CT images. To assist such analysis, existing methods in clinical studies focus on clinical data with classical but interpretable methods [9]. In the image analysis domain, the trend is either to use deep learning methods [18] or to focus on automatically extracting quantitative information (radiomics features)

from PET images and combining them with machine learning methods [7]. In this context, we aim to develop a computer-aided method to identify high-risk patients at diagnosis, relying on both clinical and imaging information.

We face multiple challenges when designing a risk classification approach from heterogeneous multimodal data. First, the quantity of available data on this disease is often limited. Also, the information in the PET volumes is spread over multiple typically small lesions, making feature extraction difficult. In addition, both image resolution and the number of lesions can vary significantly across patients, hindering generalizability. Finally, the integration of the different modalities is still an open question in the field [3].

In this paper, we rely on recent advances in Graph Attention Networks (GATs) to combine the information from the multiple lesions while handling the variable number of lesions. We further couple the GAT with a cross-attention fusion module to efficiently integrate data from clinical and imaging modalities. The model is trained and evaluated using a private prospective multicentric dataset with 583 patients suffering from DLBCL. Experimental validation results show that our proposed method yields a good 2-year progression-free survival (PFS) classification accuracy while outperforming classical supervised methods based on either clinical, imaging or both clinical and imaging data.

2 Related work

Recently, there has been a growing interest in developing computer-assisted methods analysing full-body PET images to support diagnosis and treatment decisions of oncological patients. Different approaches have been considered, relying either on a region of interest (ROI) surrounding a single lesion, or on the full image. For example, methods in [1,10] make outcome or prognosis predictions from lesions ROIs. However, images are only part of the patient's information that physicians rely on to determine the best treatment options. Other approaches [14] rely on both clinical data and image features from the most intense focal lesion to predict the PFS of multiple myeloma patients. However, for all these methods, resuming a full-body image to a single ROI may not fully represent the patient's state as it overlooks the information from other lesions and their potentially structured spatial distribution.

Few papers tackle the problem of incorporating both the imaging descriptors and the underlying structure of all the patient lesions [11,8,2]. They rely on graph representations to model this structural information and build a graph neural network (GNN) on top to provide different types of predictions, e.g. of the probability of distant metastasis over time [8], or the PFS [11,2]. Aswathi et al. [2] exploit only imaging descriptors taken from multiple lesions, while [11] and [8] consider a naive late fusion to incorporate clinical information, i.e. the clinical features are concatenated with imaging descriptors just before the prediction computation at the last fully connected layer. However, given the naive fusion's simplicity, alternative approaches are needed to study the fusion of multiple lesions and heterogenous data modalities.

Beyond PET imaging and cancer risk prediction, there has been an increasing interest in fusing the information from multiple modalities to perform better-informed predictions. As discussed by Baltrušaitis et al. [3], there are multiple ways of fusing multimodal data, e.g. the classical: early, late and hybrid fusion approaches, kernel-based methods, graphical models and some neural networks. However, none is today consensual for dealing with heterogeneous medical data.

Recently, cross-attention modules have been explored to fuse multiple modalities in bio-medical applications. For instance, Mo et al. [12] implemented a cross-attention strategy to fuse the information from two MRI imaging modalities for a segmentation task. Chen et al. [6] computed a cross-attention based on transformers [16] to register two imaging modalities, by considering different modalities for query than for the keys/values. Finally, targeting heterogenous data, Bhalodia et al. [4] used cross-attention for pneumonia localization by computing cosine similarities between images and text embeddings. Beyond the medical domain but relying on graphs, Xie et al. [17] proposed to fuse vectorial and graph data with cross-attention modules for open relation extraction in text analysis.

In this work, we build a multi-lesion graph to capture image and structural properties [11,8,2]. In addition, we take inspiration from [6] and [17] to propose a cross-attention method between the image lesion graph and clinical data. The proposed model addresses the identification of high-risk patients in DLBCL.

3 Method

Problem statement Let a DLBCL clinical exam before treatment be composed of a full-body PET image acquired on a patient, and a set of tabular clinical indicators. Our goal is to perform a PFS 2-year classification, intended to predict whether the disease of a patient will progress within two years after the beginning of the treatment. This indicator helps to identify high-risk patients (more likely to progress). In this context, we propose a learning framework (c.f. Fig. 1), taking as input clinical tabular data and a full-body 3D PET image with 3D segmentation of the lesions, trained to predict a probability of 2-year PFS.

First, we design a *lesion graph* to simultaneously represent the image features of individual lesions and their spatial distribution. Then, a GNN is built on the top of the constructed graph, composed of i) *graph attention* modules that learns a latent representation from multiple lesions; and ii) a *cross-modal* fusion blocks integrating clinical data. A final *prediction* module aggregates the fused information into a classification score.

Lesion Graph construction The first step of our framework is the creation of a fully connected graph $\mathcal{G}^{(n)} = \{\mathcal{V}^{(n)}, \mathcal{E}^{(n)}\}$ to group the information from the $L^{(n)}$ lesions present on the PET scan of the n^{th} patient. We construct this graph as in [2]: each node $v_i^{(n)} \in \mathcal{V}^{(n)}$ corresponds to a single lesion, and is associated with a feature vector $\mathbf{z}_i^{(n)} \in \mathbb{R}^{D_{\text{features}}}$. This vector contains both classical

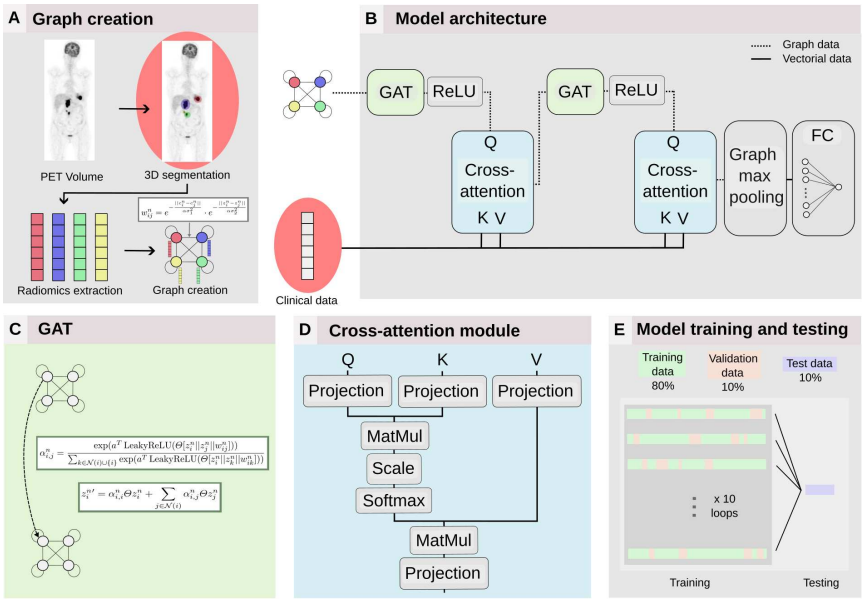

Fig. 1. Method overview: (A) patient-level graph with imaging information from every lesion, (B) model architecture, propagating the information from the multiple nodes (with the GATv2) and fusing it with the clinical data by the cross-attention block, (C) explanation of the GATv2, (D) the cross-attention mechanism, (E) training and testing schemes. The red circles indicate the patient's information provided in the dataset.

intensity-based and radiomics features[6](c.f Table 1 in the Supp. material). In the following, we denote by $\mathbf{Z}^{(n)} \in \mathbb{R}^{L^{(n)} \times D_{\text{features}}}$ the matrix concatenating all nodes' features $\mathbf{z}_i^{(n)}$, with D_{features} the dimension of the vector including classical and radiomics features.

Edges $e_{ij}^{(n)}$ are drawn between every pair of nodes $v_i^{(n)}$ and $v_j^{(n)}$, including self-loops. Weights $w_{ij}^{(n)}$ are assigned to each edge to favor message passing between closer and more similar lesions. The values of $w_{ij}^{(n)}$ are defined based on the distances between both the feature vectors $\mathbf{z}_i^{(n)}$ and the lesions centroids $\mathbf{p}_i^{(n)}$:

$$w_{ij}^{(n)} = \exp\left(-\frac{||\mathbf{p}_i^{(n)} - \mathbf{p}_j^{(n)}||_2}{\gamma\sigma_1^2}\right) \cdot \exp\left(-\frac{||\mathbf{z}_i^{(n)} - \mathbf{z}_j^{(n)}||_2}{\gamma\sigma_2^2}\right), \tag{1}$$

where $||.||_2$ stands for the L2 norm; σ_1, σ_2 denote the population-level standard deviations of the centroid and the feature distances, respectively; and γ is a hyper-parameter tuned to find the best edge weight distribution for our task.

[6] Here, classical features are quantitative measurements on the segmented lesion describing the intensity distribution of the voxels. Radiomics features instead describe the 3D structure of the lesion, such as shape, or second-order features that reveal the inter-relationship among voxels.

Multi-Lesion Representation Learning To study the relations between the lesions and to pool their information, we define a GNN over our lesion graph. We rely on the GATv2 convolution layer [5] for its capacity to adapt the neighbors' attention weights independently for each node. In our context, the attention scheme of GATv2 implies that the feature vector of each lesion is updated based on information propagated from the most relevant neighboring lesions only. We implement the `torch_geometric` version of this operator, which takes into account edge weights by computing the attention coefficients $\alpha_{i,j}$ as follows:

$$\alpha_{i,j}^{(n)} = \frac{\exp(\mathbf{a}^T \,\text{LeakyReLU}(\mathbf{\Theta}[\mathbf{z}_i^{(n)}||\mathbf{z}_j^{(n)}||w_{ij}^{(n)}]))}{\sum_{k\in\mathcal{N}(i)\cup\{i\}} \exp(\mathbf{a}^T \,\text{LeakyReLU}(\mathbf{\Theta}[\mathbf{z}_i^{(n)}||\mathbf{z}_k^{(n)}||w_{ik}^{(n)}]))}, \tag{2}$$

with $\mathbf{a}$ and $\mathbf{\Theta}$ learned parameter matrices, $\cdot||\cdot$ the concatenation operation and $\mathcal{N}(i)$ the neighboring nodes of $v_i^{(n)}$. The features assigned to each lesion (i.e. node in the graph) are updated as:

$$\mathbf{z}_{i_{\text{GAT}}}^{(n)} = \alpha_{i,i}^{(n)}\mathbf{\Theta}\mathbf{z}_i^{(n)} + \sum_{j\in\mathcal{N}(i)} \alpha_{i,j}^{(n)}\mathbf{\Theta}\mathbf{z}_j^{(n)}. \tag{3}$$

The D_{GAT} dimension of the node's representation $\mathbf{z}_{i_{\text{GAT}}}^{(n)}$ at the output of a GATv2 block is determined by a grid search, as is also the dropout probability applied to this module. Finally, the updated lesion representations are passed through a ReLU activation. The resultant $L^{(n)} \times D_{\text{GAT}}$ feature matrix is a concatenation of the lesions feature vectors: $\mathbf{Z}_{\text{GAT}}^{(n)} = \text{ReLU}([\mathbf{z}_{1_{\text{GAT}}}^{(n)}|| \ldots ||\mathbf{z}_{L^{(n)}{}_{\text{GAT}}}^{(n)}]^\top)$.

Multimodal Multi-lesion Cross-Attention We aim now at projecting the updated node features $\mathbf{Z}_{\text{GAT}}^{(n)}$, into a more representative space by integrating the clinical knowledge of the patient n represented by a vector $\mathbf{c}^{(n)} \in \mathbb{R}^{D_{\text{clin}}}$; note that there is a vector $\mathbf{c}^{(n)}$ per patient (and not per lesion). For this purpose, we take advantage of the self-attention module proposed in [16] adapted to the cross-modal case. The module takes as input a query vector $\mathbf{Q}$ and a key/value pair of vectors $\mathbf{K}$ and $\mathbf{V}$ and outputs a weighted sum of the values, where the weight assigned to each value is computed from a compatibility function (*i.e.* a scalar product) of the query with the corresponding key (normalized by the key dimension d_k). By defining $\mathbf{Q} = \mathbf{Z}_{\text{GAT}}^{(n)}$ and $\mathbf{K} = \mathbf{V} = \mathbf{c}^{(n)}$, the signals assigned to each lesion are updated with the information procured by the clinical data:

$$\begin{aligned}\mathbf{Z}_{\text{CrossAtt}}^{(n)} &= \text{CrossAtt}(\mathbf{Z}_{\text{GAT}}^{(n)}, \mathbf{c}^{(n)}, \mathbf{c}^{(n)}) \\ &= \text{softmax}\left(\frac{\mathbf{Z}_{\text{GAT}}^{(n)}\mathbf{W}^Q(\mathbf{c}^{(n)}\mathbf{W}^K)^T}{\sqrt{d_k}}\right)\mathbf{c}^{(n)}\mathbf{W}^V.\end{aligned} \tag{4}$$

We optimize during training the latent representations of $\mathbf{Q}$, $\mathbf{K}$ and $\mathbf{V}$ and the cross attention output via three learnable matrices $\mathbf{W}^Q \in \mathbb{R}^{D_{\text{GAT}}\times D_{\text{clin}}}$, $\mathbf{W}^K \in \mathbb{R}^{1\times D_{\text{clin}}}$ and $\mathbf{W}^V \in \mathbb{R}^{1\times D_{\text{GAT}}}$. The result of the cross-attention operation is a matrix $\mathbf{Z}_{\text{CrossAtt}}^{(n)}$ of same size as $\mathbf{Q}$ ($L^{(n)} \times D_{\text{CrossAtt}} = L^{(n)} \times D_{\text{GAT}}$).

Intuitively, matrices $\mathbf{W}^Q$ and $\mathbf{W}^K$ project the multi-lesion image data and the clinical vector on to a common space, before computing their compatibility. The softmax output, of size $(L^{(n)} \times D_{\text{clin}})$, provides the attention values that each lesion should give to the entries of the clinical vector. Finally, the attention scores are multiplied with the clinical data vector, lifted to the D_{GAT} dimension by $\mathbf{W}^V$. The updated individual node features correspond to the rows of $\mathbf{Z}^{(n)}_{\text{CrossAtt}}$. The multi-lesion and cross attention modules are repeated for two layers. After the second layer, we end up with $\mathbf{Z}^{(n)}_{\text{CrossAtt}}{}' \in \mathbb{R}^{L^{(n)} \times D_{\text{GAT}}}$.

Prediction A max pooling on $\mathbf{Z}^{(n)}_{\text{CAtt}}{}'$, across the node dimension, resumes the graph features to a D_{GAT}-dimensional vector, allowing us to handle patients with different numbers of lesions. The pooled vector is given to a linear layer with a sigmoid activation function to make a prediction of the 2-year PFS for a given patient. The learning is controlled by a weighted binary cross-entropy loss function, where weights compensate for the class imbalance (ratio of positive to negative samples $\sim 1:5$).

4 Experiments

Dataset The proposed method was evaluated on the prospective GAINED study (NCT 01659099) [9] which enrolled 670 newly diagnosed and untreated DLBCL patients. In order to perform our binary prediction of the 2-year PFS, we removed the patients who were censored before this time, which left us with 583 samples. Among these patients, 101 were deemed as positive for the PFS because of a progression or a relapse of the disease within two years, while 12 were positive because of death without progression of the disease. In this dataset, are assigned to each patient a PET image at the beginning of the protocol as well as clinical indicators such as age, ECOG scale, Ann Arbor stage or number of extranodal sites (full list is presented in Supp. material). The lesion detection on the PET images is done manually by a clinician and the segmentation is performed using a majority vote between three usual lesion segmentation methods: i) a K-means clustering ($K = 2$), ii) a thresholding that retains only voxels with intensity values larger than 41% of the maximum intensity, and iii) a second thresholding to keep voxels whose normalized SUV (Standard Uptake Value) is more than 2.5. The imaging and clinical features are both standardised by removing the median of the training data and scaling the whole dataset according to the quartile range: $Scaled\ value = \frac{Original\ value - training\ median}{training\ interquartile\ range}$. The distance between the centroid of the lesions (Eq. 1) is standardized in a similar way, but considering the mean and quartiles of the lesions' centroids individually for each patient.

Comparison to baseline models Our model was compared to six other baseline models performing the same task:

- **MLP clinical**: An MLP whose only input is the vector of clinical data of a patient. The model comprises two linear layers with ReLU activations and

a 1-dim linear output layer with sigmoid activation. The two intermediate layers have the same dimension, in practice chosen via a grid search.
- **MLP image**: An MLP with the same configuration but taking as input the imaging data. We compute the input image vector as the average of the feature vectors from individual lesions to handle the variable lesion number across patients. For each lesion we extract features as in Sec. 3.
- **MLP clinical+image**: An MLP, with the same configuration as the previous ones, but taking as input the concatenation of both the clinical and imaging data (*i.e.* the input image vector as for the MLP image).
- **MIL image**: A MIL approach taking as input the imaging features from the $L^{(n)}$ lesions of a patient, applies a one-layer MLP followed by a ReLU on each lesion's feature vector, aggregates the results by a maximum operation and projects it linearly (with a sigmoid activation) to get the prediction.
- **GraphConv image**: A GraphConv model [13], taking as input a lesion graph as in Sec. 3, but using a graph convolution aggregation function, see Eq. 5. The model is composed of two GraphConv layers, the first having an output dimension determined by grid search, and the second with an output size of 1. The first layer has a ReLU activation, and the second is followed by a max pooling operation and a sigmoid activation to predict the PFS.

$$\mathbf{z}^{(n)}_{i_{\text{GraphConv}}} = \mathbf{W}_1 \mathbf{z}^{(n)}_i + \mathbf{W}_2 \Big(\sum_{j \in \mathcal{N}(i)} w^{(n)}_{ij} \mathbf{z}^{(n)}_j \Big). \tag{5}$$

Ablation study In order to prove the interest of each module in our framework we also do two ablation studies. First, we implement our model with GraphConv layers replacing the GATv2 layers to study the impact of the learned attention weights between the lesions. Then, we replace the cross-attention layers by a simple concatenation $[\mathbf{z}^{(n)}_{i_{GAT}} || \mathbf{c}^{(n)}]$ to verify if the proposed learnable fusion between the two modalities improves the performance of the model.

Experimental setup We strictly divide the 583 patients in three distinct sets of training (80%), validation (10%) and test (10%). Test results are reported for the model with the best validation ROC AUC. To evaluate our model, the split is repeated ten times as follows: a single test set is left out from all the loops, and at each loop the remaining data is randomly split into training and validation sets, while ensuring that the ratio of positive patients is the same in all the sets. Furthermore, to ensure the scores are computed on balanced sets, we repeat the validation and test phases five times: for each run we build a balanced set with all the available positive data and $1/5$ of the negative data, randomly sampled from the validation and test sets respectively. The resulting metrics are then averaged to get the final validation or test results. A grid search (c.f. Table 3 in Supp. material) is performed on the learning rate, the hidden channel size and, for the GNN, the parameter γ (used in the lesion graphs construction) to find the model configuration that grants the best validation ROC AUC. Furthermore, in order to validate the statistical significance of our results, we use a t-test to

compare the results of our model against the baselines. The whole framework has been coded in Python with `PyTorch` and `torch_geometric` modules.

5 Results

Quantitative results We report in Table 1 the results of our comparative study. Our experiments reveal that models based on clinical data perform better than models using imaging data only. Furthermore, for models based on imaging data, considering the lesions individually (as the nodes of a graph or a bag of nodes in the MIL) seems to improve the predictions compared to averaging the feature vectors. Also, using a graph improves over the bag of lesions/MIL approach. Finally, the proposed framework performs significantly better than all the other models (p-value < 0.005), showing it efficiently fuses the information from multiple lesions and from the two considered modalities.

For the **ablation** studies, replacing the cross-attention layers by a simple concatenation results in a big performance drop (test ROC AUC of 0.59 ± 0.06 against 0.72 ± 0.03 initially), proving the benefit of our multimodal data fusion method. However, replacing the GATv2 layers with GraphConv layers does not significantly affect the performances (test ROC AUC of 0.71 ± 0.04).

The better performance of clinical-based models compared to those based on imaging can be partially explained by the selection of a subset of clinical variables known for being predictive [9]. Another aspect influencing the image-based models is the high complexity of the lesion segmentation task for DLBCL patients given that lesions tend to superpose and have diffuse contours. Nonetheless, we argue that an efficient integration of both kinds of data, and all the lesions, as proposed here, should allow for a better assessment of the patient's state.

Table 1. Test ROC AUC of the considered models (best performance in bold), with the p-value comparing the results to those of the cross-attention model.

Model	Clinical data	Image data	AUC	p-value
MLP	x	-	0.66 ± 0.04	0.002
MLP	x	x (average)	0.61 ± 0.04	< 0.001
MLP	-	x (average)	0.47 ± 0.04	< 0.001
MIL	-	x (per lesion)	0.56 ± 0.06	< 0.001
GraphConv	-	x (per lesion)	0.58 ± 0.06	< 0.001
Cross-attention	x	x (per lesion)	$\mathbf{0.72 \pm 0.03}$	—

Qualitative results We also studied the learned attention weights in the cross-attention modules (c.f. figures in Supp. material) in order to better understand where the model focuses when learning to predict the patients' 2-year PFS. Firstly, we observe that the cross-attention weights across patients can behave differently, with either overall constant weights across rows (lesions) and columns (clinical variables), or approximately constant rows, or variations across rows and columns. However, the two cross-attention modules for a patient tend to be similar. Secondly, the contribution of the different clinical features is mostly

equilibrated: each clinical feature is given approximately the same amount of attention, which is expected since, as we mentioned before, we rely on known biomarkers. For some patients, few clinical features stand out. For example, for one patient (Fig. 2 in Supp. material), the model puts a strong attention on his LDH value, which is quite low, and on his aaIPI (age-adjusted International Prognostic Index, which is equal to 1). The prediction for this patient is negative, i.e., no relapse within two years. This seems coherent with the physician's thinking process when trying to asses the condition of a patient, confirming the relevance of the multimodal fusion by the cross-attention module.

6 Conclusion

We address treatment response prediction of DLBCL patients two years after diagnosis. To this end, we propose a new cross-attention graph learning method integrating image information from multiple lesions and clinical tabular data. Experimental validation on a prospective clinical dataset shows that our model can efficiently exploit the complementary information, performing significantly better than all compared baselines. As perspectives, we will consider cost functions adapted to survival analysis for a more fine-grained treatment response estimation in time and a better modelling of censored patients. In addition, studying graphs defined on lesions sub-regions rather than whole lesions [11] could help mitigate the impact of intra/inter-operator segmentation variability, especially for lesions whose delimitation is unclear. Finally, we plan to investigate the generalisation ability of our model to other pathologies.

Acknowledgements This work has been funded by the AIby4 project (Centrale Nantes-Project ANR-20-THIA-0011), INCa-DGOS-INSERM-ITMO Cancer 18011 (SIRIC ILIAD) with the support from the Pays de la Loire region (GCS IRECAN 220729), the European Regional Development Fund (FEDER), the Pays de la Loire region on the Connect Talent MILCOM programme and Nantes Métropole (Conv. 2017-10470).

References

1. Amyar, A., Ruan, S., Gardin, I., Chatelain, C., Decazes, P., Modzelewski, R.: 3-D RPET-NET: Development of a 3-D PET imaging convolutional neural network for radiomics analysis and outcome prediction. IEEE TPRMS **3**(2) (2019)
2. Aswathi, A., Rizkallah, M., Frecon, G., Bailly, C., Bodet-Milin, C.M., Casasnovas, O., Gouill, S.L., Kraeber-Bodéré, F., Carlier, T., Mateus, D.: Lesion graph neural networks for 2-year progression free survival classification of diffuse large B-cell lymphoma patients. In: IEEE Int. Symp. on Biomedical Imaging (ISBI) (2023)
3. Baltrušaitis, T., Ahuja, C., Morency, L.P.: Multimodal machine learning: A survey and taxonomy. IEEE TPAMI **41**(2), 423–443 (Feb 2019)
4. Bhalodia, R., Hatamizadeh, A., Tam, L., Xu, Z., Wang, X., Turkbey, E., Xu, D.: Improving pneumonia localization via cross-attention on medical images and reports. In: Med Image Comp and Comp Assisted Interventions (MICCAI) (2021)

5. Brody, S., Alon, U., Yahav, E.: How attentive are graph attention networks? In: Int. Conf. on Learning Representations (2022)
6. Chen, J., Liu, Y., He, Y., Du, Y.: Deformable cross-attention transformer for medical image registration. ArXiv:2303.06179 (2023)
7. Jiang, C., Li, A., Teng, Y., Huang, X., Ding, C., Chen, J., Xu, J., Zhou, Z.: Optimal PET-based radiomic signature construction based on the cross-combination method for predicting the survival of patients with diffuse large B-cell lymphoma. Eur J Nucl Med Mol Imaging **49**(8), 2902–2916 (Jul 2022)
8. Kazmierski, M., Haibe-Kains, B.: Lymph node graph neural networks for cancer metastasis prediction (Jun 2021), http://arxiv.org/abs/2106.01711
9. Le Gouill, S., Ghesquières, H., Oberic, L., Morschhauser, F., Tilly, H., Ribrag, V., Lamy, T., Thieblemont, C., Maisonneuve, H., Gressin, R., Bouhabdallah, K., Haioun, C., Damaj, G., Fornecker, L., Bouhabdallah, R., Feugier, P., Sibon, D., Cartron, G., Bonnet, C., André, M., Chartier, L., Ruminy, P., Kraeber-Bodéré, F., Bodet-Milin, C., Berriolo-Riedinger, A., Brière, J., Jais, J.P., Molina, T.J., Itti, E., Casasnovas, R.O.: Obinutuzumab vs rituximab for advanced DLBCL: a PET-guided and randomized phase 3 study by LYSA. Blood **137**(17), 2307–2320 (2021)
10. Li, H., Boimel, P., Janopaul-Naylor, J., Zhong, H., Xiao, Y., Ben-Josef, E., Fan, Y.: Deep convolutional neural networks for imaging data based survival analysis of rectal cancer. In: IEEE Int. Symp. on Biomedical Imaging (ISBI) (2019)
11. Lv, W., Zhou, Z., Peng, J., Peng, L., Lin, G., Wu, H., Xu, H., Lu, L.: Functional-structural sub-region graph convolutional network (FSGCN): Application to the prognosis of head and neck cancer with PET/CT imaging. Computer Methods and Programs in Biomedicine **230**, 107341 (Mar 2023)
12. Mo, S., Cai, M., Lin, L., Tong, R., Chen, Q., Wang, F., Hu, H., Iwamoto, Y., Han, X.H., Chen, Y.W.: Mutual information-based graph co-attention networks for multimodal prior-guided magnetic resonance imaging segmentation. IEEE Trans. on Circuits and Systems for Video Technology **32**(5), 2512–2526 (May 2022)
13. Morris, C., Ritzert, M., Fey, M., Hamilton, W., Lenssen, J., Rattan, G., Grohe, M.: Weisfeiler and Leman go neural: Higher-order graph neural networks. Proc. of the AAAI Conf. on Artificial Intelligence **33**, 4602–4609 (Jul 2019)
14. Morvan, L., Carlier, T., Jamet, B., Bailly, C., Bodet-Milin, C., Moreau, P., Kraeber Bodéré, F., Mateus, D.: Leveraging RSF and PET images for prognosis of multiple myeloma at diagnosis. Int J Comp Assisted Radiology and Surgery (2019)
15. Susanibar-Adaniya, S., Barta, S.K.: 2021 update on Diffuse large B cell lymphoma: A review of current data and potential applications on risk stratification and management. American journal of hematology **96**(5), 617–629 (May 2021)
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Neural Information Processing Systems (NeurIPS) (2017)
17. Xie, B., Li, Y., Zhao, H., Pan, L., Wang, E.: A cross-attention fusion based graph convolution auto-encoder for open relation extraction. IEEE/ACM TASLP **31**, 476–485 (2023)
18. Yuan, C., Shi, Q., Huang, X., Wang, L., He, Y., Li, B., Zhao, W.L., Qian, D.: Multimodal deep learning model on interim 18F-FDG PET/CT for predicting primary treatment failure in diffuse large B-cell lymphoma. European Radiology **33**, 77–88 (2022)